



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

MODELAMIENTO DE VOLÚMENES GEOLÓGICOS BASADO EN LA ESTIMACIÓN  
Y SIMULACIÓN ESTOCÁSTICA DE DISTANCIAS AL CONTACTO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS MENCIÓN  
COMPUTACIÓN

FELIPE ANDRÉS GARRIDO RODRÍGUEZ

PROFESOR GUÍA:  
NANCY HITSCHFELD KAHLER  
PROFESOR GUÍA 2:  
JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:  
MARÍA CECILIA RIVARA ZUÑIGA  
PATRICIO INOSTROZA FAJARDIN  
LUIS SALINAS CARRASCO

SANTIAGO DE CHILE  
ABRIL 2016

# Resumen

El modelamiento geológico es uno de los primeros pasos en la estimación de recursos en un yacimiento. Este modelamiento se realiza a partir de una interpolación basada en un número limitado de muestras obtenidas en sondajes y tiene por objetivo reproducir de la manera más exacta lo que ocurre en el subsuelo. Esta representación es de suma importancia debido a que influye directamente sobre la toma de decisiones y, por lo tanto, en la metodología para explotar el yacimiento, la rentabilidad y otros asuntos operacionales.

El objetivo de este trabajo es implementar una nueva metodología para el modelamiento de cuerpos geológicos utilizando geoestadística para estimar y simular la distancia de las muestras al contacto entre el cuerpo mineralizado y la roca caja que lo contiene. Esta nueva metodología permite incorporar información interpretativa o blanda, en los algoritmos de estimación mediante kriging ordinario y simulación secuencial gaussiana.

El trabajo consistió en modificaciones a GSLIB, ésta es una librería que contiene los algoritmos más usados en geoestadística. Las rutinas modificadas fueron dos: *kt3d*, que corresponde al algoritmo de kriging, y *sgsim*, que corresponde a la simulación secuencial gaussiana. Estos algoritmos se modificaron para incluir la información blanda en dos modalidades. La primera consiste en sondajes artificiales, o pseudosondajes, creados por el usuario y que explicitan lo que el experto infiere con respecto a la continuidad y conectividad del cuerpo y/o de la roca caja. La segunda es usar direcciones variables, las cuales definen las direcciones de continuidad en el algoritmo. Estas direcciones se utilizan de dos maneras: orientando la búsqueda y el variograma según el campo de direcciones o bien, utilizándolas en una nueva definición de distancia no euclidiana que las incorpora.

Los resultados obtenidos, para una prueba en un yacimiento real, muestran una mejora de entre 1% y 5% para el acierto total. Si bien, para dicho indicador es una leve mejora, sigue siendo una mejora, mas aún considerando que el yacimiento no posee plegamientos importantes que permitan explotar todo el potencial de estos algoritmos. De los algoritmos de estudiados, cabe destacar el comportamiento de la simulación utilizando anisotropía variable, pues mostró una mejora importante en la precisión y en el calce con el volumen real del cuerpo, aunque acompañada de una baja considerable en la cobertura. Esta característica hace de este método una buena alternativa para evaluar yacimientos de una manera conservadora, cuando la inversión se ve restringida y el riesgo debe ser el menor posible, es decir, no tanto alcance, pero con gran precisión.

El uso de estos nuevos algoritmos constituyen una nueva alternativa para el modelamiento, donde el usuario puede asistir con su conocimiento experto, la creación del modelo, sobre todo en yacimientos donde exista estructuras complejas que dificulten el proceso normal. La ventaja de éstos sobre otras técnicas que persiguen el mismo objetivo es lo intuitivo de las ideas detrás de las modificaciones y de la información adicional que se genera. Estos algoritmos constituyen parte del backend de una aplicación que permitirá al usuario interactuar con los datos para guiar el modelamiento de cuerpos geológicos.

# Tabla de contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.2.1. Objetivo General . . . . .	2
1.2.2. Objetivos Específicos . . . . .	3
1.3. Alcances . . . . .	3
1.4. Metodología . . . . .	3
1.5. Descripción por Capítulos . . . . .	5
<b>2. Antecedentes</b>	<b>7</b>
2.1. Geoestadística . . . . .	7
2.1.1. Introducción . . . . .	7
2.1.2. Estimación geoestadística . . . . .	13
2.1.3. Simulación geoestadística de variables continuas . . . . .	20
2.1.4. Simulación geoestadística de variables categóricas . . . . .	23
2.1.5. Simulación Multipunto . . . . .	26
2.2. Modelamiento Geológico . . . . .	27
2.2.1. Plantas y secciones . . . . .	27
2.2.2. Simulación basada en procesos . . . . .	29
2.2.3. Simulación basada en distancia entre objetos (ODSIM) . . . . .	30
2.2.4. Estimación directa de la probabilidad multivariada . . . . .	30

2.2.5.	Estimación y simulación basada en la distancia al contorno . . . . .	30
2.2.6.	Simulación en presencia de direcciones variables . . . . .	31
2.3.	Manejo de rugosidad del contacto entre el cuerpo y el exterior . . . . .	33
<b>3.</b>	<b>Modificación de algoritmos y generación de información blanda</b>	<b>34</b>
3.1.	<i>GSLIB</i> . . . . .	34
3.2.	Glosario . . . . .	35
3.3.	Modificación de la rutina KT3D . . . . .	37
3.3.1.	Manejo de Datos . . . . .	38
3.3.2.	Nueva rutina KT3D_SDH . . . . .	39
3.3.3.	Nueva rutina KT3D_LVO . . . . .	40
3.4.	Modificación de la rutina SGSIM . . . . .	42
3.4.1.	Manejo de Datos . . . . .	44
3.4.2.	Nueva rutina SGSIM_SDH . . . . .	45
3.4.3.	Nueva rutina SGSIM_LVO . . . . .	45
3.4.4.	Manejo de la rugosidad . . . . .	45
3.5.	Generación de datos desde la interpretación del usuario . . . . .	46
3.5.1.	Pseudosondajes . . . . .	46
3.5.2.	Campo de direcciones . . . . .	50
<b>4.</b>	<b>Caso de estudio</b>	<b>53</b>
4.1.	Cálculo de la distancia . . . . .	54
4.2.	Datos disponibles . . . . .	55
4.2.1.	Información Real . . . . .	56
4.2.2.	Información Blanda . . . . .	56
4.3.	Análisis exploratorio de datos . . . . .	57
4.4.	Kriging . . . . .	63
4.4.1.	Variografía . . . . .	63

4.4.2.	Parámetros para la ejecución . . . . .	66
4.5.	Resultados de Kriging . . . . .	67
4.5.1.	Original . . . . .	67
4.5.2.	Con pseudosondajes: caso 1 . . . . .	68
4.5.3.	Con pseudosondajes: caso 2 . . . . .	70
4.5.4.	Con direcciones variables: Orientación de la búsqueda y variograma .	72
4.5.5.	Con direcciones variables: Anisotropías locales . . . . .	74
4.5.6.	Discusión de los resultados . . . . .	76
4.6.	Simulación . . . . .	77
4.6.1.	Desagrupamiento y Transformación a Gaussiana . . . . .	77
4.6.2.	Variografía . . . . .	79
4.6.3.	Parámetros para la ejecución . . . . .	80
4.7.	Resultados de Simulación . . . . .	81
4.7.1.	Original . . . . .	81
4.7.2.	Con pseudosondajes: caso 1 . . . . .	83
4.7.3.	Con pseudosondajes: caso 2 . . . . .	84
4.7.4.	Con direcciones variables: Orientación de la búsqueda y variograma .	86
4.7.5.	Con direcciones variables: Anisotropías locales . . . . .	87
4.7.6.	Discusión de los resultados . . . . .	89
4.8.	Control de la rugosidad . . . . .	90
<b>5.</b>	<b>Conclusiones</b>	<b>92</b>
	<b>Bibliografía</b>	<b>96</b>
<b>A.</b>	<b>Validación hipótesis Multigaussiana</b>	<b>98</b>
<b>B.</b>	<b>Control de rugosidad</b>	<b>100</b>

# Índice de tablas

1.1. Tabla de contingencia. . . . .	5
1.2. Tabla de Indicadores. . . . .	6
3.1. Cantidad de líneas de código para KT3D y SGSIM, para algoritmos originales (ORIG), con pseudosondajes (SDH) y con orientaciones variables (LVO). . .	35
4.1. Coordenadas de la zona de estudio. . . . .	58
4.2. Sondajes: cantidad de muestras. . . . .	59
4.3. Pozos de tronadura: cantidad de muestras. . . . .	59
4.4. Estadísticas de las muestras de sondajes. . . . .	61
4.5. Parámetros del variograma experimental para kriging. . . . .	64
4.6. Grilla para estimación, simulación y del campo de direcciones variables. . . .	66
4.7. Parámetros para kriging. . . . .	66
4.8. Estadísticas de la distancia estimada al contacto para kriging original. . . . .	68
4.9. Tablas de contingencia para validación de kriging de distancias original. . . .	68
4.10. Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias original. . . . .	68
4.11. Estadísticas de la distancia estimada al contacto para kriging con pseudosondajes (caso 1). . . . .	69
4.12. Tablas de contingencia para validación de para kriging de distancias con pseudosondajes (caso 1). . . . .	70
4.13. Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con pseudosondajes (caso 1). . .	70
4.14. Estadísticas de la distancia estimada al contacto para kriging con pseudosondajes (caso 2). . . . .	72

4.15. Tabla de contingencia para validación de kriging de distancias con pseudosondajes (caso 1). . . . .	72
4.16. Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con pseudosondajes (caso 2). . .	72
4.17. Estadísticas de la distancia estimada al contacto para kriging con orientaciones variables. . . . .	73
4.18. Tabla de contingencia para validación de kriging de distancias con orientaciones variables. . . . .	74
4.19. Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con orientaciones variables. . .	74
4.20. Estadísticas de la distancia estimada al contacto para kriging con anisotropías variables. . . . .	75
4.21. Tabla de contingencia para validación de kriging de distancias con anisotropías variables. . . . .	76
4.22. Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con anisotropías variables. . . .	76
4.23. Resumen de indicadores obtenidos para los distintos kriging. . . . .	77
4.24. Estadísticas de la variable distancia para los datos desagrupados. . . . .	79
4.25. Parámetros del variograma experimental para simulación. . . . .	79
4.26. Parámetros para simulación. . . . .	81
4.27. Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana original. . . . .	82
4.28. Indicadores para simulación secuencial gaussiana original. . . . .	83
4.29. Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 1). . . . .	83
4.30. Indicadores para simulación secuencial gaussiana con pseudosondajes (caso 1). . . . .	84
4.31. Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 2). . . . .	85
4.32. Indicadores para simulación secuencial gaussiana con pseudosondajes (caso 2). . . . .	86

4.33. Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con orientación variable. . . . .	86
4.34. Indicadores para simulación secuencial gaussiana con orientación variable. . .	87
4.35. Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con anisotropías variables. . . . .	88
4.36. Indicadores para simulación secuencial gaussiana con anisotropías variables. .	89
4.37. Resumen de resultados para las distintas simulaciones con junto al kriging respectivo. . . . .	90

# Índice de ilustraciones

2.1. Esquema del variograma . . . . .	10
2.2. Gráfico del modelo esférico . . . . .	11
2.3. Gráfico del modelo exponencial . . . . .	12
2.4. Gráfico del modelo gaussiano . . . . .	13
2.5. Gráfico del modelo potencia . . . . .	13
2.6. Idea tras la anamorfosis para histograma acumulado . . . . .	21
2.7. Ejemplo de capas sedimentarias con sus umbrales para gaussiana truncada. . . . .	25
2.8. Ejemplos de banderas y una realización asociada a ella[19]. . . . .	26
2.9. Ejemplo de dos perfiles (x=80000 e y=20000). . . . .	28
2.10. Ejemplo de secciones para algoritmo semi-automático. . . . .	28
2.11. Ejemplo de la restricción sobre los triángulos para la triangulación. . . . .	29
2.12. Ejemplo de triangulación entre 2 polígonos. . . . .	29
2.13. Ejemplos de formas de utilizar las direcciones variables en la búsqueda de muestras . . . . .	31
2.14. Transformación que generan las direcciones variables[2] . . . . .	32
3.1. Esquema de los componentes de <i>gslib</i> . Con borde rojo aparecen las subrutinas utilizadas; con relleno amarillo, las subrutinas creadas y con relleno verde, los ejecutables modificados. . . . .	35
3.2. Archivo de parámetros de KT3D . . . . .	38
3.3. Diagramas de funcionamiento para KT3D original, KT3D_SDH y KT3D_LVO . . . . .	39
3.4. Ejemplo de búsqueda para rutina KT3D_LVO en 2D . . . . .	41
3.5. Diagramas de funcionamiento para SGSIM original, SGSIM_SDH y SGSIM_LVO . . . . .	43

3.6. Archivo de parámetros de SGSIM . . . . .	44
3.7. Esquema en 2D que explica la idea tras los pseudosondajes. . . . .	47
3.8. Ejemplo de selección de sondajes. . . . .	48
3.9. Ejemplo de generación de triangulación. . . . .	49
3.10. Ejemplo de pseudosondaje creado. . . . .	50
3.11. Ejemplo de plano de corte y planos que delimitan la zona donde las muestras se consideran. . . . .	51
3.12. Ejemplo de proyección de las muestras y spline generada. . . . .	52
4.1. Ejemplo de cálculo de distancia. . . . .	55
4.2. Ejemplo de selección de muestras modificado. . . . .	57
4.3. Ejemplo del cálculo de la distancia al contacto antes y después de filtrar las muestras. . . . .	58
4.4. Vista isométrica de los datos para modelamiento. . . . .	60
4.5. Campo de direcciones variables. . . . .	60
4.6. Histogramas de las variables utilizadas para datos de sondajes y pseudosondaje. . . . .	62
4.7. Vista isométrica de los datos de validación. . . . .	62
4.8. Visualización en perfil NORTE-ELEVACIÓN para datos reales . . . . .	63
4.9. Visualización en perfil ESTE-ELEVACIÓN para datos reales. . . . .	63
4.10. Visualización en perfil ESTE-NORTE para datos reales. . . . .	63
4.11. Variogramas para kriging. . . . .	65
4.12. Vista del modelo e histograma de la distancia estimada al contacto para kriging original. . . . .	67
4.13. Vista del modelo e histograma de la distancia estimada al contacto para kriging con pseudosondajes (caso 1). . . . .	69
4.14. Vista de los nuevos pseudosondajes. . . . .	71
4.15. Vista del modelo e histograma de la distancia estimada al contacto para kriging con pseudosondajes (caso 2). . . . .	71
4.16. Vista del modelo e histograma de la distancia estimada al contacto para kriging con orientaciones variables. . . . .	73

4.17. Vista del modelo e histograma de la distancia estimada al contacto para kriging con anisotropías variables. . . . .	75
4.18. Histograma de la variable distancia al contacto para datos desagrupados. . .	78
4.19. Variogramas para simulación. . . . .	80
4.20. Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana original. . . . .	82
4.21. Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 1). . . . .	84
4.22. Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 2). . . . .	85
4.23. Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con orientación variable. . . . .	87
4.24. Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con anisotropías variables. . . . .	88
A.1. Nubes de correlación diferida. . . . .	98
A.2. Razón entre raíz del variograma y madograma. . . . .	99
B.1. Cortes para simulaciones con diferentes valores de efecto pepa, desde 0.05 (a) hasta 0.50 (j) en intervalos incrementales de 0.05. . . . .	101
B.2. Cortes para simulaciones con diferentes aportes de las estructuras, desde 100 % exponencial(a) a 100 % gaussiano(k) en intervalos de 10 %. . . . .	102
B.3. Mapa de aciertos en función del efecto pepa y de la estructura. . . . .	103

# Capítulo 1

## Introducción

El modelamiento de los cuerpos geológicos que componen un yacimiento es una de las primeras etapas dentro del ciclo de vida de una mina. Este modelamiento consiste en generar una representación numérica, generalmente modelos de bloques, de los minerales que hay en la corteza terrestre, cerca de la superficie. Estos modelos son de suma importancia para las siguientes etapas, pues a partir de éstos se toman una serie de decisiones, como el diseño de la mina, planificar la extracción, la cantidad de maquinaria, entre otras decisiones importantes.

De entre las técnicas que se utilizan actualmente se pueden identificar tres líneas. La primera de ellas utiliza herramientas de la geometría computacional para construir mallas geométricas que delimitan el cuerpo de manera determinística[27]. La segunda línea corresponde al uso de técnicas geoestadísticas bivariadas[9][7], por medio de las cuales, se interpola y/o simula alguna característica como la litología, tipo de mineral o la alteración; utilizando la continuidad espacial de dichas características. La tercera y última clasificación corresponde a los métodos que utilizan estadísticas multipunto[10][22][26][28][13][18]. Este tipo de técnicas utilizan imágenes de entrenamiento para definir la estructura que debiese tener el cuerpo. Estos métodos obtienen patrones de imágenes de entrenamiento y simulan valores en los sitios no muestreados, condicionados por la frecuencia de los patrones que se forman a partir de los datos reales en la vecindad del punto a simular.

### 1.1. Motivación

El problema de modelar un sólido a partir de datos de sondajes de manera determinística puede ser visto de distintas maneras, sin embargo, todas decantan en problemas conocidos en ciencias de la computación. Algunos de estos problemas son: encontrar la envolvente convexa a partir de un conjunto de puntos, encontrar las superficies de nivel u obtener una malla geométrica a partir de puntos de una superficie. Los algoritmos que resuelven estos problemas generan mallas donde la única información utilizada es la posición de los puntos, mas no utilizan tendencias que se puedan desprender de un análisis estadístico sobre las demás variables disponibles. Adicionalmente, los problemas en los que se suelen utilizar estos algoritmos son aquellos en que la cantidad de datos es muy grande, como ocurre en el caso

de las máquinas utilizadas para exámenes médicos o escaneadores 3D, los cuales generan un muestreo bastante denso y regular, lo que permiten tener mallas sumamente finas. Estas dos características hacen que generar polígonos planos, es decir utilizar una interpolación lineal entre los puntos, sea suficientemente bueno. Sin embargo, en la minería la cantidad de datos con los que se cuenta inicialmente para modelar los cuerpos en el yacimiento es escasa, por ende los modelos que se pueden obtener utilizando estas técnicas no son muy buenos, pues no reproduce la rugosidad<sup>1</sup> del contacto entre el cuerpo y el exterior.

El enfoque estadístico para resolver el problema del modelamiento de cuerpos geológicos tampoco está libre de dificultades. Como su nombre lo indica, estos algoritmos utilizan estadísticos calculados sobre los datos, luego el resultado puede, eventualmente, no acercarse a la realidad, obteniendo modelos con ruido o cuerpos altamente desagregados. Debe tenerse en cuenta también que, en etapas tempranas de la prospección, la cantidad de sondajes puede ser lo suficientemente pequeña como para que las técnicas aplicadas dejen de ser robustas y la estimación y/o simulación resulten defectuosas.

Uno de los problemas de todos los métodos es el manejo que se da a la información interpretativa. En el caso de los métodos que utilizan estadística multipunto, el geólogo puede incorporar su conocimiento del terreno al algoritmo a través de las imágenes de entrenamiento. Sin embargo, en estos algoritmos las imágenes de entrenamiento son de gran importancia, por lo cual el algoritmo es altamente susceptible a interpretaciones erróneas. El modelamiento por medio de secciones es, en gran medida, un método en que la interpretación del geólogo es el elemento fundamental, siendo los datos reales información que el geólogo interpreta y no una entrada del método. En el caso de la geoestadística bivariada la forma de incorporar la interpretación del geólogo es a través de la variografía, limitando bastante lo que el usuario puede hacer.

El presente trabajo busca crear un conjunto de herramientas geoestadísticas que resuelva las dificultades anteriormente expuestas, permitiendo al usuario guiar el modelamiento por medio de su conocimiento y experiencia. Estas herramientas deben ser intuitivas, es decir, que la generación de los datos blandos a partir de la intuición del geólogo sea clara, al igual que cómo los datos blandos afectan al modelo final. El primer paso, pasar de la interpretación del geólogo a los datos blandos, es un tema que no se aborda en este trabajo. Los algoritmos desarrollados resuelven el segundo paso, que corresponde a tomar la información blanda e incorporarla en los algoritmos de estimación y simulación geoestadística.

## 1.2. Objetivos

### 1.2.1. Objetivo General

El objetivo de este trabajo es diseñar e implementar un algoritmo para realizar el modelamiento geológico mediante estimación y simulación geoestadística, utilizando información proveniente de sondajes y basándose en la distancia de las muestras a los contactos

---

<sup>1</sup>La rugosidad del contacto corresponde a la regularidad de éste. Un contacto es más rugoso a medida que la superficie es más irregular, por el contrario, mientras más plano o suave se dice menos rugoso.

entre unidades geológicas. El algoritmo debe ser capaz además de incorporar la información que el geólogo crea importante para mejorar los resultados del modelamiento.

### 1.2.2. Objetivos Específicos

1. Incorporar información blanda en los algoritmos de kriging ordinario y en la simulación secuencial gaussiana.
2. Proponer método para incorporar la interpretación del geólogo acerca del yacimiento al algoritmo.
3. Incorporar el ajuste de la rugosidad entre contactos a partir de los parámetros del variograma.
4. Proponer e implementar un método para calibrar las rugosidad a partir de los datos disponibles.
5. Validar los resultados obtenidos con validación cruzada para datos reales de un yacimiento.

## 1.3. Alcances

El presente trabajo corresponde a los algoritmos que soportan la integración de la información blanda. Luego, es necesario explicitar que en este trabajo se asume como datos de entrada los pseudosondajes y el campo de direcciones variable, y no se incluye el desarrollo de algoritmos para la creación de éstos. Es pertinente mencionar también que no se considera la creación de una interfaz gráfica, pues existen bastantes programas que permiten la visualización de datos de sondajes, pozos de tronadura y de modelos de bloques con los que se trabaja.

## 1.4. Metodología

El presente trabajo corresponde a una implementación de modelamiento de cuerpos basado en distancias al contacto. Este modelamiento consiste en la codificación de las muestras en función a su distancia al contacto entre el cuerpo y el exterior, para posteriormente estimar y/o simular esta variable distancia, la cual es continua, con kriging ordinario y simulación secuencial gaussiana. Posterior a la simulación o estimación se realiza un postproceso donde se asigna 1 a las distancias que corresponden al interior del cuerpo y 0 al exterior. Adicionalmente se incorpora en la etapa de búsqueda y/o del kriging, la información blanda que ayuda al modelamiento.

El trabajo se divide de manera natural en tres grandes temas. El primero de ellos es la modificación de las rutinas KT3D y SGSIM de *gslib*[1][5], las cuales corresponden a kriging

y la simulación secuencial gaussiana, respectivamente. Las modificaciones a estas rutinas corresponden al manejo de los pseudosondajes y direcciones variables, para su posterior uso en los respectivos algoritmos. El segundo tema corresponde a un caso de estudio de los algoritmos modificados sobre un yacimiento real, en el cual se contrastan con el algoritmo original y donde se espera explicitar ventajas y desventajas de cada método. El tercer tema corresponde al análisis de los resultados, de los cuales se puede extraer una serie de métricas que permiten evaluar la calidad de las respuestas en diferentes aspectos como la certeza o recuperabilidad.

Las modificaciones hechas a las rutinas KT3D y SGSIM consisten básicamente en cambios en la etapa de kriging de ambos algoritmos. En el caso del uso de pseudosondajes, se usan ideas del kriging factorial para suavizar el efecto de los datos blandos sobre el modelo. En términos prácticos, el cambio consiste en utilizar los datos como si fueran muestras normales, mas al momento de crear el sistema se debe quitar, para los datos provenientes de pseudosondajes, el efecto pepa a la covarianza entre el punto a estimar y la muestra. Esto se traduce en una modificación al sistema de kriging, específicamente en el lado derecho del sistema. En el caso de las direcciones variables, éstas se usan sobre la búsqueda y orientación del variograma. Para simplificar este proceso se utilizó la misma estructura de búsqueda original y reducir el conjunto de candidatos a uno mucho más pequeño. Este conjunto de muestras no descartadas se ordena y se filtra, obteniendo sólo las muestras en la elipsoide definida por el campo de direcciones. La idea de este procedimiento es utilizar de la mejor manera posible lo que ya existe. También se utilizó en la comparación las rutinas KT3D\_LVA y SISIM\_LVA, las cuales implementan otro algoritmo, que utiliza las direcciones variables en el cálculo de las distancias.

La segunda parte corresponde a un caso de estudio, donde se aplicaron los algoritmos estudiados a una parte del yacimiento *Cerro Colorado* perteneciente a la compañía *BHP Billiton Pampa Norte*. Para realizar el caso de estudio se cuenta con los datos de sondajes los cuales contienen una gran cantidad de variables, de las cuales sólo se utilizó la variable tipo de mineral (*minty*). Los datos de sondajes se dividieron en dos sets: el primero, de un 10% del total, se utilizó para obtener los modelos mediante kriging y simulaciones; y el segundo, con el resto de los datos, para validar la calidad de la respuesta. Adicionalmente se utilizaron los datos de pozos de tronadura para la validación.

La última etapa corresponde a la discusión de los resultados obtenidos. La discusión se enfoca en esta nueva metodología como una herramienta más entre las ya existentes, por lo que una conclusión única general no es posible sino conclusiones respecto a sus ventajas, criterios de comparación, desventajas y escenarios. La comparación se realiza sobre la pertenencia o no al cuerpo mineralizado. La pertenencia al cuerpo mineralizado se obtiene aplicando un umbral de corte igual a 0 sobre la distancia estimada o simulada al contacto: los bloques mayores a 0 pertenecen al exterior y, los menores o iguales a 0, se consideran como parte del cuerpo. El cálculo de los indicadores utilizados para evaluar la calidad de los modelos obtenidos, se realiza sobre esta variable binaria que indica la pertenencia al cuerpo mineralizado y no sobre la distancia estimada o simulada al contacto directamente. Los indicadores utilizados se muestran en la tabla de contingencia de la Tabla 1.1. En ésta se muestra la cantidad de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN). En este trabajo, el caso *positivo* se refiere a “informar

como cuerpo”, luego los verdaderos positivos son los datos de validación con valor real igual a *cuerpo* y valor informado también *cuerpo*. Siguiendo la lógica de lo anterior, los verdaderos negativos corresponden a los datos con valor real *exterior* y valor informado *exterior*, los falsos positivos corresponden a los datos con valor real *exterior* y valor informado *cuerpo* y, finalmente, los falsos negativos corresponden a los datos cuyo valor real es *cuerpo* y su valor informado *exterior*. Los números indicados en la Tabla 1.1 corresponden a *precision* para positivos y negativos(1 y 2), *recall* para positivos y negativos (3 y 4), y finalmente el acierto total (5). El indicador *precision* indica qué fracción de los bloques informados como cuerpo son realmente parte del cuerpo. A su vez, *Recall* indica la fracción de los bloques que son parte del cuerpo que son informadas como tal en el modelo. Estos dos indicadores son los clásicos por su complementariedad: mientras uno se preocupa de la precisión el otro se preocupa de la cobertura. Ambos indicadores deben ser buenos simultáneamente. Si se tiene alta precisión y baja cobertura sólo se informa una pequeña parte de los valores de interés, por otro lado si la cobertura es alta y la precisión es baja no se está siendo selectivo.

		Muestras		
		Cuerpo	Exterior	
Modelo	Cuerpo	VP	FP	1
	Exterior	FN	VN	2
		3	4	5

Tabla 1.1: Tabla de contingencia.

Además de estos indicadores se calcularon también el acierto en cuanto a volumen del cuerpo. Este indicador sólo hace referencia al volumen y no considera la forma ni la posición. Estos indicadores, a diferencia de los demás, debe mantenerse cercano a 1, pues eso indica que el volumen del cuerpo modelado y el volumen del cuerpo real son similares. Un valor menor a 1 indica que el volumen del cuerpo modelado es menor que el real, por lo tanto se está subestimando. Análogamente, en el caso que el indicador es mayor que 1 se está sobrestimando. En la Tabla 1.2 se muestran los indicadores utilizados y sus respectivas fórmulas. Tanto el conteo de VP, FP, FN y VN como las fórmulas, se calcularon usando los puntos estimados, es decir, se descartó los puntos que no pudieron ser estimados. Como consecuencia de esto, la suma de VP, FP, FN y VN puede no ser igual para todos los algoritmos.

## 1.5. Descripción por Capítulos

En el Capítulo 2 se detallan los antecedentes que preceden a este trabajo, entre ellos se encuentran los algoritmos en los que se soporta este trabajo, como también algoritmos que intentan resolver el mismo problema de una manera diferente. En el siguiente capítulo, el Capítulo 3 se explica las modificaciones realizadas, mostrando, con diagramas de flujo, los cambios hechos en los algoritmos para incorporar el manejo de información blanda. Pos-

precision positivos	$PP = \frac{VP}{VP+FP}$
precision negativos	$PN = \frac{VN}{VN+FN}$
recall positivos	$RP = \frac{VP}{VP+FN}$
recall negativos	$RN = \frac{VN}{VN+FP}$
acierto total	$AT = \frac{VP+VN}{VP+VN+FP+FN}$
razón volumen cuerpo modelado/real	$RVC = \frac{VP+FP}{VP+FN}$
razón volumen exterior modelado/real	$RVE = \frac{VN+FN}{VN+FP}$

Tabla 1.2: Tabla de Indicadores.

teriormente, en el Capítulo 4, se muestra el caso de estudio: los parámetros utilizados, las decisiones tomadas y, en general, todo lo necesario para ejecutar el kriging y la simulación incorporando información experta, y su posterior análisis. Para finalizar, en el Capítulo 5 se concluye acerca de los resultados obtenidos en el caso de estudio.

# Capítulo 2

## Antecedentes

En este capítulo se describen las técnicas más conocidas del área de modelamiento geológico, tanto de leyes como de *facies* geológicas (variables categóricas). El énfasis está puesto en el área de la geoestadística, pues es la base del trabajo realizado. El primer tema a describir es la geoestadística y todo lo que ésta conlleva, a saber, la teoría, supuestos y herramientas. También se muestran algunas técnicas convencionales de estimación y simulación. El segundo tema comprende las nuevas técnicas que se han desarrollado para el modelamiento geológico y que escapan de la geoestadística tradicional o la utilizan de manera no convencional.

### 2.1. Geoestadística

#### 2.1.1. Introducción

La geoestadística es una rama de la matemática desarrollada en la segunda mitad del siglo XX, principalmente, por la necesidad de una metodología para la evaluación de depósitos minerales. Esta metodología se soporta en una teoría desarrollada en torno a la correlación de datos. Estos datos poseen, además de un valor para cierta característica, una ubicación en el espacio y/o tiempo. La disposición de estos datos en el espacio o en el tiempo liga de alguna manera sus valores y es utilizada por las diversas herramientas de esta disciplina. Actualmente, la geoestadística es utilizada en diversas áreas de la ciencia y del sector productivo, como la minería, hidrogeología, minería del petróleo, medición de biomasa, aplicaciones forestales, contaminación, entre otras[9].

La definición de geoestadística hace referencia al estudio de fenómenos “regionalizados”. Este tipo de fenómenos ocurren generalmente en el espacio en tres dimensiones o en el temporal, pues la mayoría de los problemas que resuelve la geoestadística, son problemas de las ciencias naturales. Sin embargo, la teoría es suficientemente robusta para lidiar con espacios más complejos. Para describir el fenómeno en estudio se trabaja con “variables regionalizadas”, las cuales son funciones que representan numéricamente las propiedades y/o

atributos que caracterizan el fenómeno[7].

Las variables regionalizadas son, en términos matemáticos, una función determinística, la cual cumple con dos propiedades que son complementarias. La primera de ellas, habla de la continuidad espacial de ésta, pues cada valor en una posición determinada, está relacionada con los valores de su vecindad. La segunda propiedad tiene relación con que ésta varía de formas irregulares, lo cual hace sumamente difícil intentar representarla de una manera simple.

Para trabajar con las variables regionalizadas se debe definir un dominio, pues si bien la variable regionalizada puede abarcar un espacio muy grande, la solución que se busca está acotada a un espacio particular. Este dominio puede estar determinado por una serie de factores, el principal de ellos, en el caso de la minería, es el tema económico: los recursos que se cuenta para la campaña de exploración o sondajes, versus lo que los estudios preliminares infieren en cuanto a recursos en el subsuelo se refiere.

Los modelos generados usando la geoestadística asumen el valor de la variable regionalizada,  $z(x)$  con  $x$  en el dominio  $D$ , como una realización de una variable aleatoria  $Z(x)$ , en otras palabras, se dice que la variable regionalizada es el resultado obtenido al realizar un experimento aleatorio. El conjunto de valores  $Z(x), x \in D$  corresponden a un proceso estocástico o, como se le suele llamar en geoestadística, una función aleatoria. La variable aleatoria, como se dijo anteriormente, es una realización de la función aleatoria, luego, es posible extraer cierta información acerca de la función aleatoria a través de esta realización particular que es la variable regionalizada estudiada. Para caracterizar la función aleatoria se utilizan los momentos de orden uno y dos. El momento de orden uno es la esperanza o también conocida como la media. En el caso de los momentos de orden dos se utilizan la varianza y el variograma. De estos tres estadísticos se hablará a continuación, dando especial énfasis al variograma, pues suele ser desconocida para muchos, a diferencia de la media y la varianza.

## Esperanza

La esperanza, media o valor esperado es el momento de primer orden para una función aleatoria. Esta se define como la suma de la probabilidad de cada evento, multiplicado por el valor de éste. La esperanza se calcula para los valores posibles en cada posición  $x$  en el dominio, luego ésta es también una función  $m(x)$  definida para el mismo dominio  $D$ , tal como se muestra en la Ecuación 2.1.

$$\mathbb{E}[Z(x)] = m(x) = \int_D f_z(z(x)) \cdot z(x) \cdot dz \quad (2.1)$$

## Varianza

La varianza corresponde, al igual que en el caso de una variable aleatoria, a una medida de dispersión de los datos. La varianza para una función aleatoria indica, cuánto se alejan

en promedio los valores del valor medio, y, en el caso de una función aleatoria, esta varianza es también una función dependiente de la posición  $x \in D$  como se muestra en la Ecuación 2.2. Suele utilizarse también la desviación estándar  $\sigma(x)$  como medida de la dispersión y de la aleatoriedad de los datos.

$$\text{Var} [Z(x)] = \sigma^2(x) = \mathbb{E} [(Z(x) - m(x))^2] = \mathbb{E} [Z(x)^2] - m(x)^2 \quad (2.2)$$

## Variograma

El semi-variograma, también conocido simplemente como variograma, es una función que indica el nivel de diferencia existente entre pares de puntos a una distancia dada. Para obtener el variograma se requiere que la función aleatoria sea estacionaria de orden dos y que no posea deriva. La fórmula del variograma se muestra en la Ecuación 2.3

$$\gamma(h) = \frac{1}{2} \text{Var} [Z(x+h) - Z(x)] \quad (2.3)$$

La estacionaridad es un supuesto que permite la inferencia de la función aleatoria, cambiando la repetición de ésta para obtener varias realizaciones, por la repetición de eventos en el espacio. En palabras simples, se asume que la función aleatoria se comporta de manera similar en todo el dominio, independiente de la posición, luego se consideran distintas zonas del dominio como si fueran distintas realizaciones. Sin este supuesto el formalismo probabilístico no sirve, pues se cuenta sólo con una realización de la función aleatoria, correspondiente a la variable regionalizada en estudio.

En términos concretos, la hipótesis de estacionaridad se traduce en que la distribución espacial multivariada de la función aleatoria es invariante a traslaciones. En otras palabras, las propiedades de los datos en su conjunto no dependen de la posición absoluta, sino de su posición relativa respecto a los vecinos. Este supuesto es bastante fuerte, sin embargo, se puede asumir, estacionaridad de segundo orden, la cual es más débil, pero suele ser suficiente para las técnicas de estimación y simulación convencionales.

La estacionaridad de segundo orden significa que la distribución univariable no depende de la posición absoluta y, que la distribución bivariable, sólo depende de la posición relativa entre las posiciones consideradas. Corolario de esta estacionaridad de segundo orden se tiene que la esperanza y varianza son constantes en el dominio  $D$ , y para el variograma, y en consecuencia la covarianza y el correlograma, sólo dependen de la distancia entre los puntos considerados, pero no de su posición en el espacio. El correlograma  $\rho(h)$  y la covarianza  $C(h)$ , corresponden a medidas de similitud entre pares de puntos. En la Ecuación 2.4 se muestra la relación entre el variograma y la covarianza y en la Ecuación 2.5 se muestra la relación entre el correlograma y la covarianza.

$$\gamma(h) = C(0) - C(h) \quad (2.4)$$

$$\rho(h) = \frac{C(h)}{C(0)} \quad (2.5)$$

En la Figura 2.1 se muestra un esquema del variograma. En ella se indican tres características importantes. La primera es el alcance, el cual indica la distancia a la cual el variograma llega a un valor estable, este valor indica también a qué distancia se pierde la correlación entre dos puntos. La segunda característica es la meseta, que es el valor donde se estabiliza el variograma, y está asociado a la varianza de la población. La tercera característica es el efecto pepita, que indica la diferencia para posiciones a distancia muy pequeña. Otro elemento que afecta el comportamiento a distancias pequeñas es el crecimiento en el primer tramo del variograma.

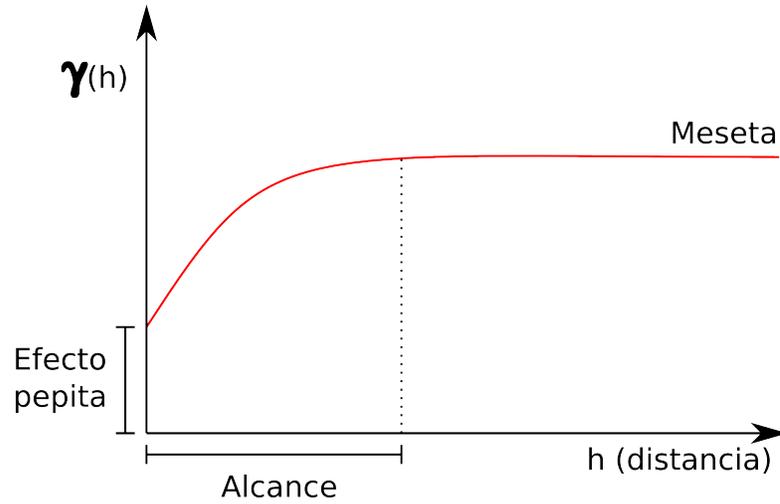


Figura 2.1: Esquema del variograma

### Variograma experimental

Al igual que en la estadística tradicional, la distribución de la variable o función aleatoria es desconocida y los momentos de ésta se deben estimar por medio de estadísticos. En el caso de la media y de la varianza los estadísticos son ampliamente conocidos. En el caso del variograma se utiliza el llamado variograma experimental cuya fórmula se muestra en la Ecuación 2.6, el cual no es más que aplicar la fórmula original sobre los datos disponibles.

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [z(x_\alpha) - z(x_\beta)]^2 \quad (2.6)$$

donde  $N(h) = \{(\alpha, \beta) \mid (x_\alpha - x_\beta = h)\}$  y  $|N(h)|$  es la cantidad de elementos en el conjunto.

Es importante mencionar que la variable  $h$  es una distancia en una dirección particular, en un plano o totalmente independiente de la dirección. El caso general corresponde a 3

direcciones ortogonales, con tres anisotropías  $a_x$ ,  $a_y$  y  $a_z$ . Así el valor de  $\gamma(h)$  corresponde a una combinación de la variabilidad en las proyecciones de  $h$  en la base que definen las direcciones de anisotropía. Cuando 2 de estas anisotropías son iguales se genera un plano de anisotropía, donde para los puntos en dicho plano no importa la dirección. En este caso el valor de  $\gamma(h)$  corresponde a la combinación de la proyección de  $h$  en el plano y en la vertical. El caso más simple es cuando las tres anisotropías son iguales, en cuyo caso el valor de  $\gamma(h)$  no depende de la dirección de  $h$ .

## Modelo variográfico

El variograma experimental no puede ser utilizado directamente en las técnicas donde se requiere. Las razones de esto son que el variograma experimental está definido sólo para algunas distancias, adicionalmente, es sólo una aproximación del comportamiento de la función aleatoria, pues la cantidad de datos es exigua y los parámetros con los que se calcula están sujetos a tolerancias de diversa índole. Para subsanar estas falencias se ajusta un modelo teórico a la curva obtenida para el variograma experimental.

Existen varios tipos de modelos variográficos, los cuales pueden combinarse para ajustarse de mejor manera al variograma experimental. Los más conocidos, y que se utilizan en *gslib*[1][5] son:

**Esférico** : En la Ecuación 2.7 se muestra la fórmula del modelo esférico para un alcance  $a$  y una meseta  $C$ . En la Figura 2.2 se puede ver la forma de este variograma para un caso particular.

$$\gamma(h) = \begin{cases} C \cdot \left( \frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \left( \frac{|h|}{a} \right)^3 \right) & \text{si } |h| \leq a \\ C & \sim \end{cases} \quad (2.7)$$

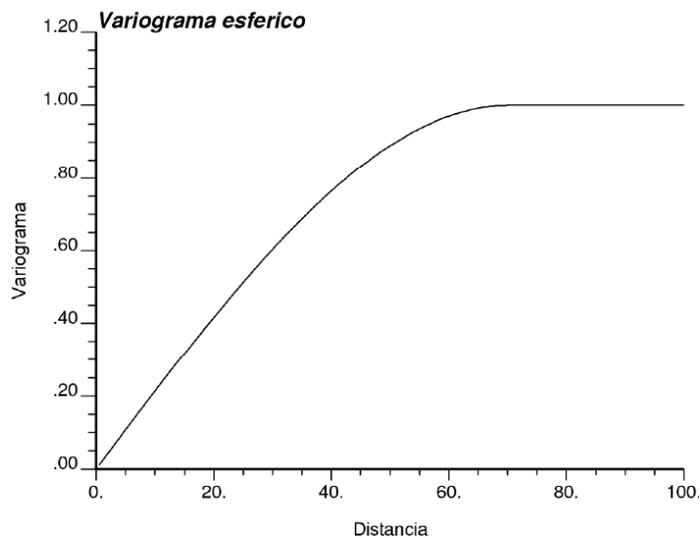


Figura 2.2: Gráfico del modelo esférico

**Exponencial** : En la Ecuación 2.8 se muestra la fórmula del modelo exponencial para un alcance práctico  $3a$  y una meseta  $C$ . En este modelo la meseta se alcanza asintóticamente. El alcance en la práctica es en  $3a$ , cuando se llega al 95 % de la meseta. En la Figura 2.3 se puede ver la forma de este variograma para un caso particular.

$$\gamma(h) = C \cdot \left(1 - e^{-\left(\frac{|h|}{a}\right)}\right) \quad (2.8)$$

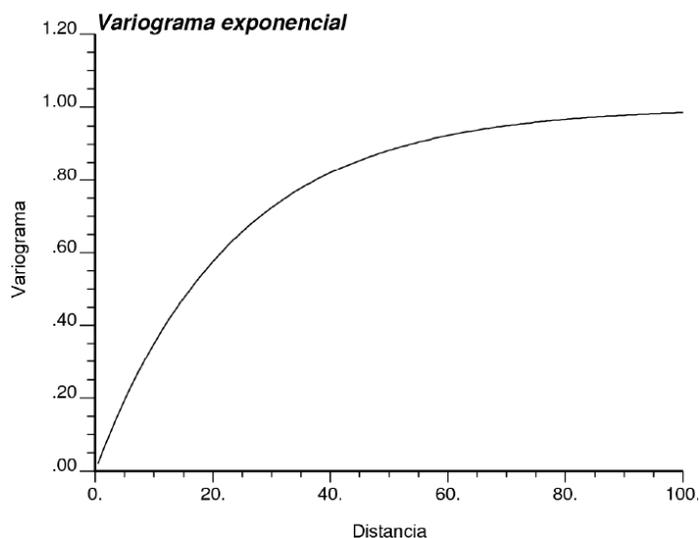


Figura 2.3: Gráfico del modelo exponencial

**Gaussiano** : En la Ecuación 2.9 se muestra la fórmula del modelo gaussiano para un alcance  $a\sqrt{3}$  y una meseta  $C$ . En este modelo la meseta se alcanza asintóticamente. El alcance en la práctica es en  $a\sqrt{3}$ , cuando supera el 95 % de la meseta. En la Figura 2.4 se puede ver la forma de este variograma para un caso particular.

$$\gamma(h) = C \cdot \left(1 - e^{-\left(\frac{|h|}{a}\right)^2}\right) \quad (2.9)$$

**Potencia** : En la Ecuación 2.10 se muestra la fórmula del modelo potencia. Este modelo no necesita un alcance ni una meseta, pues es un modelo que crece infinitamente. Los parámetros de este modelo son la pendiente  $\omega$  y el exponente  $\theta$ . En la Figura 2.5 se puede ver la forma de este variograma para distintos casos.

$$\gamma(h) = \omega \cdot |h|^\theta \quad (2.10)$$

Es importante destacar algunas cosas respecto al modelo variográfico. La primera de ellas es que modelar el variograma experimental es una tarea subjetiva donde éste se puede corregir con interpretaciones sobre información adicional, por lo cual no es buena idea utilizar sólo ajustes automáticos. Frente a un problema al intentar ajustar un modelo a un variograma experimental se debe priorizar las distancias más pequeñas sobre las más grandes, pues los valores más cercanos son los que más influyen al momento de estimar o simular un valor. Además se debe considerar la escala de trabajo, pues el definirla implica desechar cierta información, lo cual facilita el ajuste.

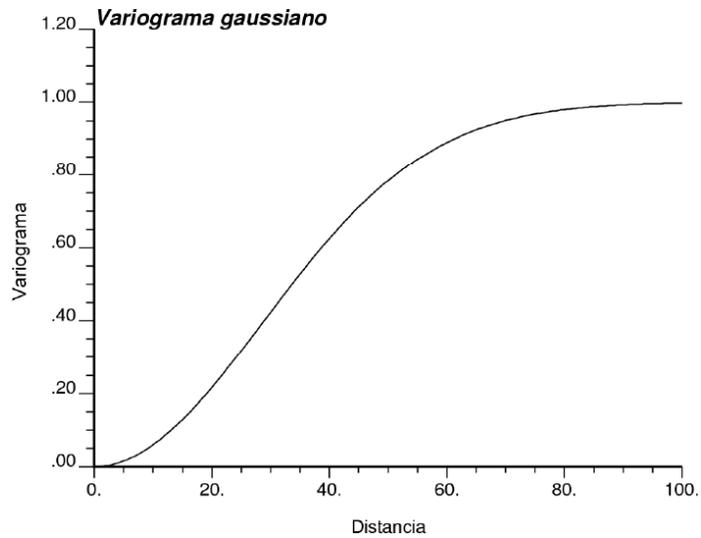


Figura 2.4: Gráfico del modelo gaussiano

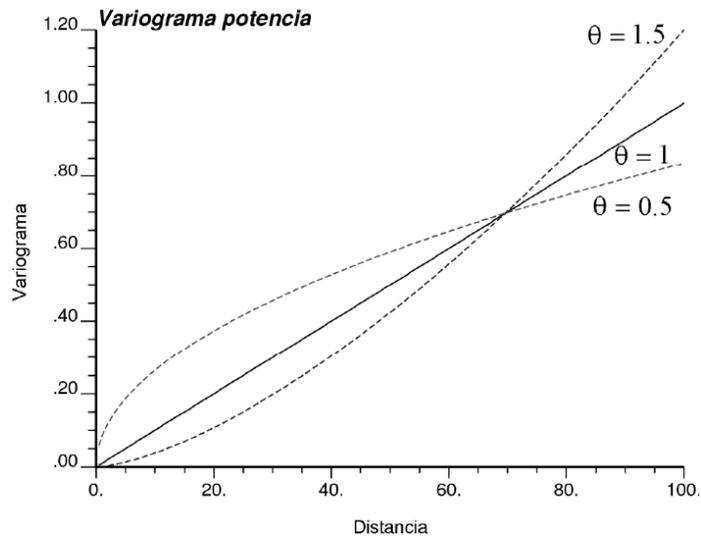


Figura 2.5: Gráfico del modelo potencia

### 2.1.2. Estimación geoestadística

Una de las aplicaciones importantes de la geoestadística en minería es la estimación de un atributo de interés a partir de muestras de sondajes. Existen muchos tipos de estimaciones, están desde las básicas, que utilizan la muestra más cercana hasta el kriging, el cual utiliza ponderadores distintos para cada una de las muestras vecinas y que asegura la menor varianza del error en la estimación.

Los métodos tradicionales se basan en la asignación de ponderadores al valor de la variable medida en las muestras vecinas al punto a estimar. Ejemplos de este tipo de métodos existen muchos, mas sólo se hablará, brevemente, del vecino más cercano y del inverso de la distancia. El kriging, por ser el mejor método en cuanto a precisión (varianza del error) se

refiere, se describirá más en detalle.

## Vecino más cercano

El método más simple para estimar un valor a partir de datos medidos es el vecino más cercano. Como su nombre bien lo indica, éste utiliza el valor del atributo para el punto más cercano y le asigna este valor al punto a estimar. Esta forma de estimar si bien es la más simple, es raramente utilizada, pues no considera ningún tipo de información geológica.

## Inverso de la distancia

La idea detrás de este método es calcular un promedio ponderado sobre los valores del atributo en los vecinos del punto a estimar, por lo cual tiene sentido sólo para variables continuas. Los pesos para los valores medidos en este método son proporcionales a  $\frac{1}{d^\alpha}$ , donde  $d$  es la distancia entre el punto a estimar y el vecino en cuestión. El valor del parámetro  $\alpha$  determina el comportamiento del método. Si el valor de  $\alpha$  es cercano a cero, el ponderador será cercano a uno, ponderando casi de manera equivalente e independiente de las distancias a las muestras utilizadas. En el otro extremo, con un valor de  $\alpha$  muy grande, salvo para la muestra más cercana, los ponderadores serán muy cercanos a cero, por lo tanto, se considerará sólo el dato más cercano, tal como en el método del vecino más cercano. Esta técnica, si bien considera de mejor manera las muestras en el espacio, no utiliza información adicional sobre la geología del lugar en estudio, ni correlaciones entre las muestras, ni anisotropías.

## Kriging

El kriging, llamado así en honor a Danie Krige, uno de los padres de la geoestadística, intenta reunir las buenas propiedades que debe tener un estimador de esta naturaleza.

- Que utilice las distancias a las muestras.
- Que dé menos peso o importancia a datos agrupados (muestreo preferencial).
- Que utilice la continuidad espacial de los datos.
- Que sea un estimador insesgado.
- Que minimice la varianza del error.

Para construir el kriging se considera una serie de restricciones, las que le dan al kriging las propiedades mencionadas. Las restricciones son la de linealidad, la de insesgo<sup>1</sup> y la de optimalidad. El kriging trabaja sobre variables continuas, por lo cual no puede ser utilizado de manera directa para la estimación de variables categóricas. Sin embargo, para

---

<sup>1</sup>[sic]

el caso de este trabajo se utilizará como variable de estudio la distancia de una muestra en un cuerpo mineralizado a su contacto con el material estéril que lo rodea, lo cual permite utilizar el kriging sobre variables continuas para la estimación de volumen y forma de cuerpos geológicos. Esta idea no es extensible a más de dos categorías directamente, mas puede aplicarse jerárquicamente para modelar varias categorías.

**Restricción de Linealidad:** Esta restricción dice que el estimador debe ser una combinación lineal ponderada de los datos, como se muestra en la Ecuación 2.11.

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha}(x_0)Z(x_{\alpha}) \quad (2.11)$$

donde  $x_0$  es el sitio a estimar,  $\{x_{\alpha}\}_{\alpha=1}^n$  son los sitios con datos y  $\{\lambda_{\alpha}(x_0)\}_{\alpha=1}^n$  son los ponderadores asociados a dichos sitios. Estos ponderadores y el coeficiente  $a$  son las incógnitas del problema de kriging. El coeficiente  $a$  también está ligado al insesgo<sup>2</sup> que caracteriza al estimador.

El motivo por el cual la combinación de los datos es lineal recae en la posibilidad de calcular sólo los momentos hasta orden dos de manera simple. En caso de que no fuera lineal se requeriría poder caracterizar la función más allá del variograma, e incluso, hacer un poco más fuerte la restricción de estacionaridad, lo cual limitaría su aplicabilidad.

**Restricción de Insesgo<sup>3</sup>:** Esta restricción dice que el error de estimación tiene esperanza cero, Ecuación 2.12. Esta restricción no asegura errores bajos, sólo que la media global del error es cercana a cero.

$$\mathbb{E}[Z^*(x_0) - Z(x_0)] = 0 \quad (2.12)$$

**Restricción de Optimalidad:** Esta última restricción, especificada en la Ecuación 2.13, equivale, cuando hay insesgo<sup>4</sup> global, a minimizar el error cuadrático medio.

$$\min_{\lambda_{\alpha}(x_0)} \text{Var}\{Z^*(x_0) - Z(x_0)\} \quad (2.13)$$

**Construcción del estimador:** Para construir el estimador, lo primero que se necesita, en términos prácticos, es contar con un conjunto de datos, a los cuales se les denomina “datos condicionantes”. El estimador es una combinación lineal de datos ponderados. Sin embargo, no se limita el número de muestras a priori, por lo cual es un factor a considerar al

---

<sup>2</sup>[sic]

<sup>3</sup>[sic]

<sup>4</sup>[sic]

momento de estimar un yacimiento. El costo computacional es  $\mathcal{O}(n^3)$  donde  $n$  es la cantidad de datos que se utilice, pues se debe resolver un sistema de ecuaciones, con una ecuación para cada dato condicionante. Para este problema existen dos soluciones. La primera, más precisa, corresponde a utilizar todos los datos condicionantes disponibles, sin embargo, el costo computacional aumenta cúbicamente con la cantidad de datos utilizados y el supuesto de estacionaridad es muy fuerte. La otra solución es utilizar una vecindad móvil, tomando para cada sitio a estimar, una cantidad fija de datos condicionantes cercanos o los que estén en un área definida. Esta solución tiene la ventaja de ser más liviana, sacrificando la precisión pero permitiendo tener un estimador más flexible y robusto a variaciones locales de la media y la varianza. En términos prácticos la segunda opción suele ser la utilizada en kriging, pues es la que más se acomoda a lo que se necesita.

**Kriging Simple (media conocida):** El kriging simple, o kriging de media conocida, asume una media conocida y constante, valga la redundancia, y a partir de las restricciones, se genera el sistema de ecuaciones para realizar la estimación.

La restricción de linealidad no tiene una dependencia de la media, por ende, no sufre cambios.

Desarrollando la restricción de insesgo<sup>5</sup>, en la Ecuación 2.14, y asumiendo una media  $m$  conocida, se puede concluir una relación entre el coeficiente  $a$ , los ponderadores  $\lambda_\alpha(x_0)$  y la media  $m$ , tal como se muestra en la Ecuación 2.15. En este kriging, se da un efecto inverso entre la media y los datos ponderados, lo cual favorece a la media cuando los datos tienen ponderadores pequeños, es decir, cuando están muy lejos, y en caso contrario, el peso de la media en la estimación disminuye.

$$\mathbb{E}[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha \underbrace{\mathbb{E}[Z(x_\alpha)]}_m - \underbrace{\mathbb{E}[Z(x_0)]}_m = a + m \cdot \left( \sum_{\alpha=1}^n \lambda_\alpha - 1 \right) \quad (2.14)$$

$$a = m \cdot \left( 1 - \sum_{\alpha=1}^n \lambda_\alpha \right) \quad (2.15)$$

Por medio de la restricción de optimalidad se llega a un sistema de ecuaciones para obtener finalmente los ponderadores  $\lambda_\alpha$ . En la Ecuación 2.16 se muestra la regla de cálculo para la varianza de la suma ponderada de variables aleatorias, donde  $C(\cdot, \cdot)$  es la covarianza.

$$\text{Var} \left( \sum_{i=1}^n \lambda_i Z_i \right) = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k C(Z_j, Z_k) \quad (2.16)$$

En la Ecuación 2.17 se desarrolla la varianza del error en la estimación, aplicando la regla de cálculo de la Ecuación 2.16.

---

<sup>5</sup>[sic]

$$\begin{aligned}
& \text{Var} [Z^*(x_0) - Z(x_0)] \\
&= \text{Var} \left[ \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) - Z(x_0) \right] \\
&= \text{Var} \left[ \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) \right] + \text{Var} [Z(x_0)] - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(Z(x_{\alpha}), Z(x_0)) \\
&= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(Z(x_{\alpha}), Z(x_{\beta})) + C(Z(x_0), Z(x_0)) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(Z(x_{\alpha}), Z(x_0)) \\
&= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + C(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) \tag{2.17}
\end{aligned}$$

La restricción de optimalidad plantea que la varianza del error es la mínima, luego para minimizar dicha varianza se anulan las derivadas parciales respecto a los ponderadores  $\lambda_{\alpha}$  en el último paso de la Ecuación 2.17. El resultado de lo anterior se muestra en la Ecuación 2.18.

$$\sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha} - x_{\beta}) = C(x_{\alpha} - x_0) \quad 1 \leq \alpha \leq n \tag{2.18}$$

En la Ecuación 2.19 se muestra el mismo resultado en notación matricial.

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ C(x_n - x_1) & \cdots & C(x_n - x_n) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \end{pmatrix} \tag{2.19}$$

Una vez calculados los ponderadores  $\lambda_{\alpha}$  el estimador de kriging simple queda de la forma mostrada en la Ecuación 2.20.

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) + m \left( 1 - \sum_{\alpha=1}^n \lambda_{\alpha} \right) \tag{2.20}$$

La varianza mínima del error de estimación en el sitio  $x_0$ , o también conocida como *varianza de kriging* adquiere la forma de la Ecuación 2.21.

$$\sigma_{KS}^2(x_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) \tag{2.21}$$

donde  $\sigma^2$  es la varianza a priori de la función aleatoria. Se puede demostrar que la varianza de kriging es menor o igual que la varianza a priori. Notar que esta varianza no depende directamente del valor de los datos, sino que de sus posiciones.

**Kriging Ordinario (media desconocida):** El kriging ordinario o de media desconocida, genera un sistema de ecuaciones muy parecido al de kriging simple.

La restricción de linealidad, al no depender de la media, no sufre cambios.

A su vez, en la Ecuación 2.22, al desarrollar la restricción de insesgo<sup>6</sup>, se llega a una relación para la incógnita  $a$  y los ponderadores  $\lambda_\alpha$ , la media  $m$  no participa de la relación, pero su calidad de incógnita permite llegar a otra relación, y no a la misma del kriging simple (Ecuación 2.15). Esta relación se puede ver en la Ecuación 2.23.

$$\mathbb{E}[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha \underbrace{\mathbb{E}[Z(x_\alpha)]}_m - \underbrace{\mathbb{E}[Z(x_\alpha)]}_m = a + m \cdot \left( \sum_{\alpha=1}^n \lambda_\alpha - 1 \right) \quad (2.22)$$

$$a = 0 \quad \wedge \quad \sum_{\alpha=1}^n \lambda_\alpha = 1 \quad (2.23)$$

La restricción de optimalidad en este caso requiere, para la optimización, utilizar un multiplicador de Lagrange  $\mu$  para introducir la relación obtenida de la restricción de insesgo<sup>7</sup>. La función objetivo que se minimiza queda de la forma que muestra la Ecuación 2.24.

$$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) + C(0) - 2 \sum_{\alpha=1}^n \lambda_\alpha C(x_\alpha - x_0) + 2\mu \left( \sum_{\alpha=1}^n \lambda_\alpha - 1 \right) \quad (2.24)$$

Al derivar parcialmente respecto a los ponderadores  $\lambda_\alpha$  y el multiplicador de Lagrange  $\mu$  se obtiene las Ecuaciones 2.25 y 2.26.

$$\sum_{\beta=1}^n \lambda_\beta C(x_\alpha - x_\beta) + \mu = C(x_\alpha - x_0) \quad 1 \leq \alpha \leq n \quad (2.25)$$

$$\left( \sum_{\alpha=1}^n \lambda_\alpha - 1 \right) = 0 \quad (2.26)$$

En la Ecuación 2.27 se muestra lo anterior en notación matricial.

---

<sup>6</sup>[sic]  
<sup>7</sup>[sic]

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) & 1 \\ \vdots & & \vdots & 1 \\ C(x_n - x_1) & \cdots & C(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \\ 1 \end{pmatrix} \quad (2.27)$$

Debido a que las ecuaciones indican que el coeficiente  $a = 0$ , el estimador corresponde a la suma ponderada de las muestras como se muestra en la Ecuación 2.28.

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) \quad (2.28)$$

La varianza de kriging en el caso del kriging ordinario queda de la forma mostrada en la Ecuación 2.29.

$$\begin{aligned} \sigma_{KO}^2(x_0) &= \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) - \mu \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(x_{\alpha} - x_0) - \mu \end{aligned} \quad (2.29)$$

donde  $\sigma^2$  es la varianza a priori de la función aleatoria, la cual equivale también a  $C(0)$ .

Es importante notar en la Ecuación 2.29 que la varianza de kriging no depende de la varianza a priori  $\sigma^2$ , y en consecuencia, el kriging ordinario aplica para caso de variogramas sin mesetas.

**Kriging de Indicadores:** El kriging de indicadores corresponde al grupo de los kriging no lineales, donde la idea es modificar el kriging original para calcular la probabilidad de satisfacer la condición de un indicador dado. Este enfoque también permite que diferentes tipos de información puedan ser utilizadas de manera conjunta. En la ecuación 2.30 se muestra la variable indicador sobre la cual se realiza el kriging de indicadores.

$$I(x) = \begin{cases} 1 & \text{si } x \in C \\ 0 & \sim \end{cases} \quad (2.30)$$

Esta variable indicador asigna el valor 1 cuando la muestra satisface la condición y 0 en caso contrario. En el caso de dos o más categorías es necesario crear una variable indicador  $I_k(x)$  para cada una de las categorías  $C_k$  y, en consecuencia, un variograma  $\gamma_k(h)$  para cada categoría.

Para el caso de los indicadores el tema de la estacionaridad es un problema, pues los indicadores generalmente se agrupan en ciertos sectores preferenciales del espacio, por lo tanto no existe la misma proporción de categorías en todos los sectores de dominio. Este comportamiento indica que los datos no se ajustan al modelo o que se viola de cierta manera el supuesto de estacionaridad, generando resultados que no son los esperados.

**Kriging Factorial:** El kriging factorial[9] consiste en la descomposición de la estimación de kriging en la suma de diferentes factores, lo cual permite también descomponer la covarianza en una suma de igual cantidad de términos. Esta descomposición permite filtrar las componentes espaciales que se desee, introduciendo cambios en el lado derecho del sistema de kriging como se muestra en la Ecuación 2.31.

$$\begin{aligned} \sum_{\beta=1}^n \lambda_{\alpha} \cdot C(x_{\alpha} - x_{\beta}) + \mu &= C(x_{\alpha} - x) - \sum_{l=0}^{l_0-1} C_l(x_{\alpha} - x) \\ \sum_{\beta=1}^n \lambda_{\alpha} &= 1 \end{aligned} \quad (2.31)$$

Los efectos aplicar estos filtros, son equivalentes a los obtenidos al trabajar con filtros en imágenes, donde al filtrar las componentes de alta frecuencia se difumina o suaviza el resultado y; al filtrar las componentes de baja frecuencia, se obtiene un resaltamiento de la variabilidad a corta distancia, o también conocido como detector de bordes. Un uso de interés para este trabajo es filtrar las componentes de alta frecuencia, en este caso el efecto pepa, para eliminar las discontinuidades que se generan en los puntos muestreados. Al aplicar este filtro la interpolación deja de ser exacta, mas no se genera discontinuidades, siendo las circunstancias del problema las que inclinarán la balanza a uno u otro lado.

### 2.1.3. Simulación geoestadística de variables continuas

La simulación geoestadística consiste en construir una variable regionalizada similar a la que se estudia. Esta simulación corresponde a otra realización de la función aleatoria con la que se trabaja, es decir, se caracteriza la variable regionalizada mediante sus estadísticos y, posteriormente, se genera una nueva realización de dicha función. La nueva realización comparte las características de continuidad y los valores en los sitios con datos, cuando la simulación es condicional. La simulación, a diferencia de kriging no es precisa, es decir, no minimiza el error, pues se centra en respetar las estadísticas de la variable regionalizada. Esto permite que en una simulación los valores no se vean suavizados como en el kriging, que exista valores fuera del rango que definen los valores en los sitios muestreados y la aparición de estructuras que no se pueden ver en el kriging y que son coherentes con la caracterización de la variable en estudio. Estas propiedades de la simulación permiten utilizar las diferentes realizaciones para hacer un análisis de riesgo sobre los modelos obtenidos, estudiar la sensibilidad de la variable y colocarse en distintos escenarios. Este tipo de simulación se utiliza tanto para variables continuas como para discretas cuantificables. Ejemplos de estas variables son la ley de un mineral, porosidad, densidad, recuperación metalúrgica, cantidad de árboles en una zona, entre otras.

Para realizar una simulación es necesario caracterizar la función aleatoria, lo cual, a diferencia del kriging, no basta con la distribución univariada y bivariada, sino que se debe conocer la distribución de probabilidad multivariada de la variable regionalizada. Luego, para poder realizar una simulación se debe trabajar con un modelo que caracterice todas estas distribuciones multivariadas.

## Modelo multi-gaussiano

El modelo multi-gaussiano asume que la variable regionalizada tiene una distribución de probabilidad multivariable gaussiana (multi-gaussiana o multi-normal). En la Ecuación 2.32 se muestra la distribución multinormal para los puntos  $\{x_i\}_{i=1}^n$  y una matriz de covarianza  $C$ .

$$g(y_1, \dots, y_n) = \frac{1}{\sqrt{2\pi}^n \cdot \sqrt{\det(C)}} e^{-\frac{y \cdot C^{-1} \cdot y^t}{2}} \quad (2.32)$$

De la Ecuación 2.32 se puede observar que, en este caso particular, la matriz de covarianza es suficiente para definir la distribución multivariable, luego caracteriza de manera total a la función aleatoria gaussiana.

## Anamorfosis

Es bastante difícil poder asumir que la variable en estudio se distribuya como una gaussiana, esto se puede notar fácilmente al ver la distribución univariable (histograma), el cual generalmente no es simétrico o no es compatible con una gaussiana. Para estos casos existe una función de transformación llamada anamorfosis, la cual a partir de un histograma cualquiera genera un histograma gaussiano estándar, con media 0 y varianza 1.

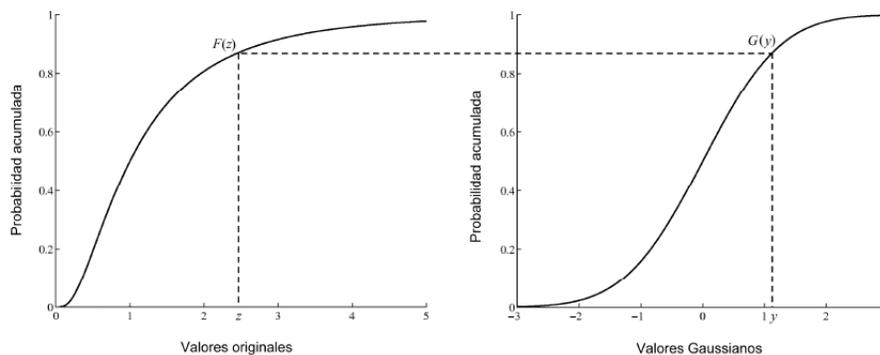


Figura 2.6: Idea tras la anamorfosis para histograma acumulado

En la Figura 2.6 se puede ver la transformación para histogramas acumulados. En este caso, para cada valor  $z$  de la variable original se le asocia el valor  $y$  con la misma frecuencia

acumulada, es decir  $F(z) = G(y)$ , donde  $F$  es el histograma original y  $G$  es el histograma gaussiano estándar. La función de anamorfosis gaussiana queda definida entonces como la composición de la inversa de  $F$  con la función  $G$ , es decir  $\phi(x) = [F^{-1} \circ G](x)$ . A la variable  $y$  se le llama “Normal Score de la variable  $x$ ”.

## Simulación secuencial gaussiana

La simulación secuencial gaussiana es el algoritmo más simple para construir nuevas realizaciones de una función aleatoria. El primer paso dentro de este algoritmo es definir un camino aleatorio que será el orden en que se visitarán los nodos de la grilla a simular. El proceso para obtener el valor simulado es igual para todos los nodos.

Sea  $u_i$  el punto a estimar,  $\{x_i\}_{i=1}^n$  los sitios con datos condicionantes y  $\{u_i\}_{i=1}^m$  los nodos ya simulados.

- Utilizando los valores en los sitios muestreados  $\{Y(x_1), \dots, Y(x_n)\}$  y en los nodos ya simulados  $\{Y(u_1), \dots, Y(u_{i-1})\}$  se selecciona una vecindad y se realiza un kriging para obtener el valor estimado  $Y^*(u_i)$  y su desviación estándar  $\sigma^*(u_i)$ .
- Generar un valor aleatorio  $g_i$ . Estos valores aleatorios son independientes entre si y se distribuyen como una gaussiana con media 0 y varianza 1.
- Calcular el valor simulado para el nodo por medio de la fórmula

$$Y_S(u_i) = Y^*(u_i) + g_i \cdot \sigma^*(u_i)$$

Se puede ver que este método es sumamente simple, además no requiere una etapa de condicionamiento como otros métodos. El procedimiento explicado es sólo una parte dentro de todo el proceso de la simulación. Típicamente los pasos a seguir son los siguientes:

- Desagrupamiento de los datos originales, lo cual genera pesos para cada muestra y permite conocer la distribución representativa de los datos.
- Transformar los datos en datos gaussianos, tomando en cuenta los pesos de las muestras obtenidos al desagrupar.
- Realizar el análisis variográfico sobre los datos transformados.
- Simular la grilla.
- Volver en la transformación, es decir aplicar  $\phi^{-1}$ .

Este método es el más simple para la simulación de variables continuas y dado que el código de este algoritmo está disponible en GSLIB se utilizó para simular el valor de la variable distancia al contacto y así simular cuerpos geológicos.

## Método de las bandas rotantes

El método de las bandas rotantes[8] reduce el problema de realizar simulaciones en  $\mathbb{R}^n$  a realizar simulaciones en  $\mathbb{R}$ , lo cual es menos costoso. El procedimiento consiste en la generación de  $N$  vectores  $U_i$  regularmente distribuidos en una bola unitaria  $S_n$  en  $\mathbb{R}^n$ , para realizar simulaciones a lo largo de dichas rectas y sumar las contribuciones de éstas, tal como se muestra en la ecuación 2.33.

$$Y(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i(\langle x, U_i \rangle) \quad (2.33)$$

donde  $\langle \cdot, \cdot \rangle$  es el producto interno o producto punto.

El método de las bandas rotantes genera una simulación multigaussiana no condicionada, por lo cual es necesario realizar una etapa de condicionamiento sobre el resultado de la simulación, la cual se muestra en la Ecuación 2.34. Este condicionamiento se efectúa mediante un kriging sobre la diferencia entre los sitios con valores medidos y el valor simulado no condicionado para los puntos muestreados. Posteriormente el resultado de la estimación se suma al valor de la simulación no condicional.

$$Y_{CS}(x) = Y_S(x) + [Y(x) - Y_S(x)]^{SK} \quad (2.34)$$

donde  $Y_S$  es el resultado de la simulación no condicional,  $Y(x)$  son los sitios muestreados y  $[\cdot]^{SK}$  corresponde a la estimación realizada mediante kriging simple.

Este método es bastante bueno, sin embargo, no es muy simple ni tampoco está contenido en GSLIB. Estas razones hicieron que se escogiera la simulación secuencial gaussiana en desmedro de ésta. No obstante, las ideas detrás de este trabajo, en principio, debiesen ser aplicables sin ningún problema en este método.

### 2.1.4. Simulación geoestadística de variables categóricas

A continuación se muestra algunas metodologías para la simulación de atributos categóricos, es decir, atributos no cuantificables, y por lo tanto, que no existe una relación de orden entre los valores. Ejemplos de variables categóricas son el tipo de roca, tipo de alteración, mineralización, entre otras. En el caso del tipo de roca los posibles valores que puede tomar la variable pueden ser andesita, calcopirita, covelina, por nombrar algunos. Estos valores posteriormente se identifican con números, sin embargo, no existe una relación de orden, es decir, la categoría 1 no es mayor (ni menor) que la categoría 2. Estos métodos no requieren, a diferencia de variables continuas, de una transformación previa, sin embargo, éstos tienden a necesitar más parámetros para su operación, pues se debe caracterizar cada categoría independientemente.

## Simulación secuencial de indicadores

La simulación secuencial de indicadores consiste en determinar la categoría a la que pertenece un punto a partir de las probabilidades obtenidas mediante kriging de indicadores. El algoritmo es bastante similar a la simulación secuencial gaussiana, pues opera de manera análoga: Obtener las muestras en la vecindad de búsqueda, para posteriormente realizar un kriging de indicadores para cada categoría  $C_k$ . Con las probabilidades de pertenencia a cada una de las categorías ya calculadas se crea una distribución de probabilidad para, finalmente, determinar un valor mediante una simulación de Montecarlo y mapear este valor a la categoría correspondiente.

Este algoritmo, al igual que el kriging de indicadores, pues lo utiliza internamente, presenta problemas respecto a la estacionaridad, pues los datos no se ajustan al modelo, el cual exige que estos cumplan con dicha restricción de estacionaridad. Este problema hace que se generen artefactos en la simulación, afectando la calidad del resultado.

## Simulación gaussiana truncada

Este método consiste en la simulación de un valor a través de una variable aleatoria gaussiana, la cual posteriormente, se mapea a una categoría mediante la truncación de ésta según diferentes umbrales. Para realizar la simulación se debe definir los umbrales mediante la fórmula en la ecuación 2.35.

$$y_i = G^{-1} \left\{ \sum_{j=1}^i p_j \right\} \quad i = 1, \dots, K - 1 \quad (2.35)$$

donde  $p_k$  corresponde a la proporción de cada clase  $G^{-1}$  a una gaussiana inversa y  $K$  la cantidad de categorías.

La variable categórica en este caso se calcula a partir de la truncación de la variable aleatoria gaussiana  $Y$  como se muestra en la ecuación 2.36.

$$I(x) = \begin{cases} c_1 & \text{si } x < y_1 \\ \vdots & \\ c_i & \text{si } y_{i-1} < x < y_i \\ \vdots & \\ c_K & \text{si } y_{K-1} < x \end{cases} \quad (2.36)$$

Este algoritmo es bueno para situaciones donde el orden de las clases se puede extraer del contexto, como el ejemplo de capas sedimentarias que se muestra en la Figura 2.7, donde los contactos entre categorías son evidentes y el orden de los correspondientes umbrales salta

a la vista. En casos de contactos que no son claros o cuando los contactos son más complejos, este método deja de funcionar de buena manera.

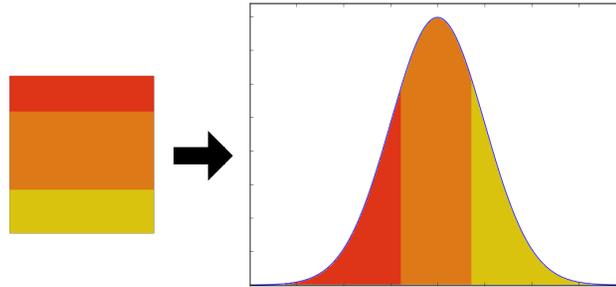


Figura 2.7: Ejemplo de capas sedimentarias con sus umbrales para gaussiana truncada.

### Simulación plurigaussiana

La simulación plurigaussiana[19] corresponde a la extensión del método de la gaussiana truncada. Para el método de la gaussiana truncada las categorías deben estar ordenadas, es decir, una al lado de la otra, lo cual obliga a que las categorías limiten sólo con otras 2: una por arriba y una por abajo. Esto no permite modelar cuerpos intrusivos o estructuras geológicas complejas, como por ejemplo, contactos entre tres categorías. La simulación plurigaussiana utiliza el mismo concepto, esta vez simulando varias variables gaussianas y truncando la tupla según corresponda.

En el caso de una bigaussiana, la función de truncamiento se conoce como *bandera*, pues los umbrales de truncamiento dependen de dos variables gaussianas  $Y_1$  e  $Y_2$ , visualmente parecido a una bandera. Esta bandera permite entender de manera gráfica e intuitiva la concordancia o discordancia entre los contactos. En la Figura 2.8a se muestra una bandera donde la clase de color azul puede limitar con las clases naranja y amarilla, sin embargo, las clases amarillas y naranja no pueden ser vecinas, a estas dos clases se les llama discordantes, pues no se tocan. En la realización asociada a dicha bandera se puede ver como en ningún caso las clases amarilla y naranja se tocan, pues está la clase azul siempre entre ellas. En la Figura 2.8b se puede ver que la bandera indica que todas las clases pueden limitar con las demás, es decir, naranja con azul, naranja con amarillo y azul con amarillo, a esto se le llama clases concordantes. En la realización asociada se puede observar que, en efecto, todas las clases limitan con las otras.

Este algoritmo tiene una serie de problemas prácticos como el condicionamiento de la simulación a datos reales, el cual utiliza un muestreador de Gibbs. Otro tema complejo es el análisis variográfico, pues se debe relacionar el variograma de indicadores con el variograma de la gaussiana.

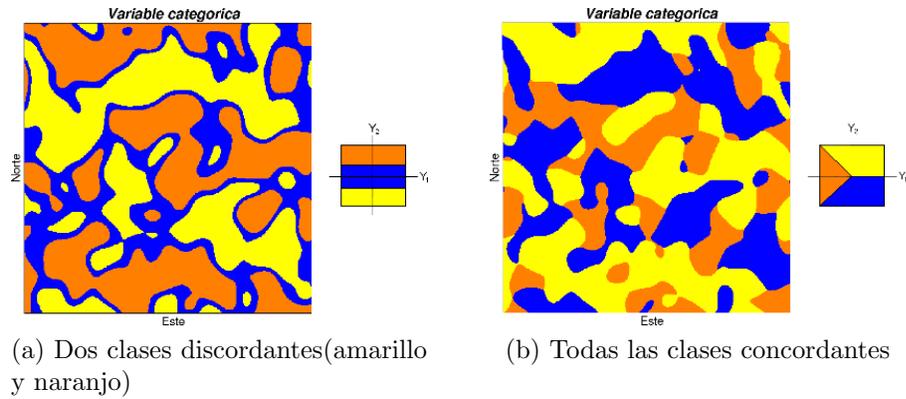


Figura 2.8: Ejemplos de banderas y una realización asociada a ella[19].

### 2.1.5. Simulación Multipunto

La simulación multipunto, o MPS por sus siglas en inglés, es una metodología en la cual se prescinde de una caracterización paramétrica de la función aleatoria, cambiándola por una imagen de entrenamiento explícita, que describe la geología de la zona de estudio. Esta simulación nace en respuesta a las limitaciones de la geoestadística convencional, donde sólo se capturan las relaciones entre pares de puntos. MPS permite utilizar relaciones de orden superior al momento de simular, por medio de patrones recolectados en la imagen de entrenamiento, sin embargo, se pierde la rigurosidad matemática de la geoestadística tradicional. A pesar de la falta de un formalismo tan fuerte para MPS, ésta tiende a reproducir las estadísticas univariadas y bivariadas de la función aleatoria (se suele hacer este test para verificar la correctitud de la simulación).

Los algoritmos de este tipo son todos muy similares, pues todos recolectan información de los patrones en la imagen de entrenamiento, para después simular un valor, utilizando el patrón más parecido a la configuración de los datos condicionantes para el nodo. El primer algoritmo para simulación multipunto [10] consistió en introducir la búsqueda de patrones en la simulación secuencial directamente, siendo una alternativa muy demandante en tiempo, pues para cada nodo a simular se busca directamente la probabilidad de ocurrencia de las clases en la imagen de entrenamiento, siendo esta recorrida una gran cantidad de veces. El algoritmo SNESIM [22] perfeccionó la idea anterior por medio de un preproceso a la imagen de entrenamiento, almacenando en una estructura de árbol la probabilidad condicional para cada patrón, con esto se reduce sustancialmente el tiempo de ejecución para la simulación. Existen diversos algoritmos[26][28][13][18] que siguen introduciendo nuevos conceptos a la simulación multipunto, como la extracción de características y posterior búsqueda en espacios multidimensionales, el uso de filtros como características para la simulación de variables continuas y el muestreo directo sobre la imagen de entrenamiento de manera inteligente.

Esta metodología ha demostrado que tiene un gran potencial, pues permite la simulación de estructuras complejas que no son reproducibles con la geoestadística convencional, sin embargo, carece de un formalismo matemático fuerte, además de una gran dependencia de la imagen de entrenamiento, que genera una alta susceptibilidad a malas interpretaciones.

## 2.2. Modelamiento Geológico

Un modelo geológico consiste en una representación en 2 ó 3 dimensiones, de los atributos existentes en el subsuelo. Este modelo puede estar compuesto de modelos más detallados, donde se ahonda en diversos aspectos, siendo de los más utilizados en evaluación de yacimientos de minerales metálicos, aquellos que definen los dominios geológicos y los modelos de leyes. Los modelos de leyes, utilizando una discretización en bloques, permiten el cálculo fácil y rápido de curvas tonelaje-ley y permite hacer una planificación a corto, mediano y largo plazo. Los dominios geológicos definen zonas donde la geología y las estadísticas de los atributos son homogéneas[6]. Esta característica de los dominios implica una estacionaridad por zonas, lo cual permite utilizar técnicas geoestadísticas para obtención de los modelos de leyes o de otros atributos.

Este trabajo se enfoca en la definición de dominios mediante la construcción de cuerpos geológicos, por lo cual se muestran las técnicas que permiten obtener cuerpos geológicos a partir de datos de análisis químico sobre sondajes, y algunas técnicas usadas en la industria. En primer lugar se presenta el método determinístico en base a plantas y secciones, por ser el más popular. Los demás métodos corresponden a técnicas que utilizan procesos estocásticos, lo cual genera modelos diferentes para cada ejecución del algoritmo. De estos algoritmos los más importantes corresponden a *distance-based contour* y la simulación en presencia de campos de direcciones variables, pues estos son la base para el trabajo realizado.

### 2.2.1. Plantas y secciones

Una de las técnicas para generar cuerpos geológicos más usada es el de plantas y secciones. Este método opera a través de la definición de perfiles en los planos XY, XZ e YZ, los cuales posteriormente se combinan para crear un entramado que, finalmente, permite generar una triangulación que define el cuerpo[6].

A simple vista es uno de los métodos más sencillos, sin embargo, es bastante lento por el poco nivel de automatización y la gran cantidad de ajustes que el usuario debe ir haciendo a medida que avanza el modelamiento. Los principales problemas de este método radican en las constantes iteraciones que se deben hacer para calzar los perfiles en distintos planos, como se muestra en el ejemplo de la Figura 2.9. Este es un proceso bastante tedioso y lento, pues como se dijo, no es automático.

En [27] se muestra un método automatizado que genera un modelo utilizando perfiles en la misma dirección. Este método es bastante mejor que aquél que utiliza perfiles en distintas direcciones, pues no necesita iterar para hacer calzar los modelos. El algoritmo genera una triangulación a partir de polígonos, en la misma dirección, como se muestra en la Figura 2.10. Los polígonos representan el contorno del cuerpo al realizar un corte en éste. Luego, la idea es reconstruir todo el borde a partir de estos contornos. Para esto se genera una triangulación entre cada par de polígonos consecutivos, como los pares  $P1 - P2$ ,  $P2 - P3$  y  $P3 - P4$  de la Figura 2.10. Para generar esta triangulación, se impone la restricción que cada triángulo debe tener un punto de un polígono y dos del otro, como se muestra en la Figura 2.11. El

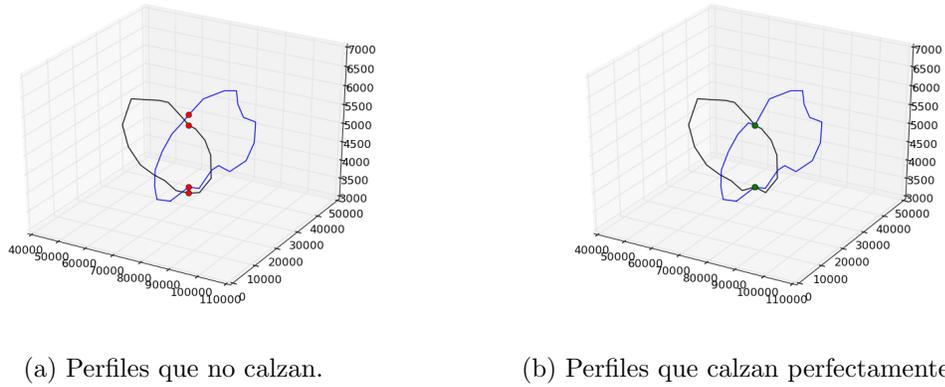


Figura 2.9: Ejemplo de dos perfiles ( $x=80000$  e  $y=20000$ ).

método se basa en la premisa que, dados 2 polígonos, la triangulación que mejor las une es la que tiene menor área. Esto tiene bastante sentido, pues la triangulación con menor área usa los triángulos cuyos puntos son más cercanos. En la Figura 2.12 se muestra la triangulación entre dos polígonos, base del método.

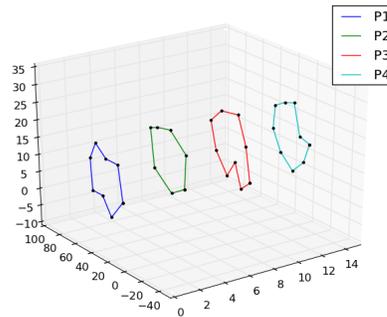


Figura 2.10: Ejemplo de secciones para algoritmo semi-automático.

No obstante la mejora en términos de automatización, este método hereda ciertos problemas del modelamiento clásico usando plantas y secciones. Algunos de estos problemas son la alta dependencia de las plantas y/o secciones, las cuales no están estrictamente ligadas a los datos. También existe la posibilidad de generar triángulos “aguja”, es decir, muy largos y delgados, pues las distancias entre perfiles son de decenas o centenas de metros y los puntos de un polígono están separados por metros o decenas de metros. El problema de estos triángulos “aguja” es que generan dificultades al momento de calcular áreas o volúmenes. Otros problemas bastante frecuentes son producto del muestreo poco regular, es decir, usar muchos puntos en zonas con mucho detalle y pocos en zonas simples, como por ejemplo, el cruzamiento de aristas, que introduce anomalías en la triangulación y, en consecuencia, problemas con el modelo.

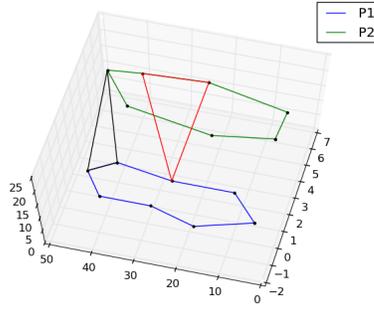
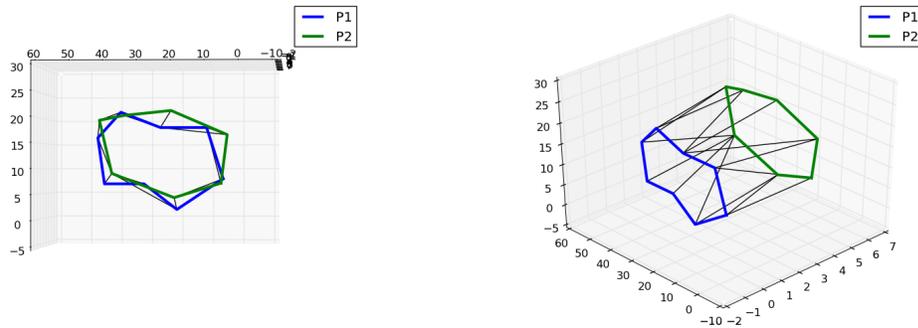


Figura 2.11: Ejemplo de la restricción sobre los triángulos para la triangulación.



(a) Vista frontal

(b) Vista lateral

Figura 2.12: Ejemplo de triangulación entre 2 polígonos.

## 2.2.2. Simulación basada en procesos

Esta familia de técnicas se basa en una reconstrucción del depósito guiada por la geología. Estas técnicas corresponden a una reconstitución de escena a una escala temporal mucho más rápida que como ocurrió. La idea es modelar la superposición de capas sedimentarias, erosiones, pliegues, fallas, intrusiones entre otros fenómenos que se suceden o sucedieron desde miles o millones de años para generar el yacimiento actual.

Existen técnicas para poder modelar yacimientos con esta técnica[14][11] y que respetan datos geológicos reales tomados del subsuelo, sin embargo, son sumamente costosos en términos computacionales, además que no se considera la relación espacial entre los datos, como las direcciones de anisotropía. Estos algoritmos sirven para generar modelos preliminares con una cantidad de datos muy exigua, sin embargo, para etapas más avanzadas dejan de ser realistas.

### 2.2.3. Simulación basada en distancia entre objetos (ODSIM)

Este algoritmo, explicado en [12], está diseñado para modelar estructuras complejas, utilizando un esqueleto de las estructuras y completando el resto de manera estocástica. El algoritmo genera un campo de distancias  $D(p)$  donde para cada punto de la grilla de simulación  $G$ , calcula su distancia al punto más cercano del esqueleto. Esta distancia se trunca con un umbral estocástico  $\psi(p)$  de las mismas dimensiones de la grilla, es decir,  $I_B(p) = D(p) < \psi(p) \quad \forall p \in G$ , generando cuerpos irregulares, pero que mantienen la tendencia del esqueleto.

Este algoritmo es sumamente novedoso, sin embargo, tiene dos problemas. El primero tiene relación con la alta dependencia de la definición del esqueleto. Es fácil inferir que, si el esqueleto no es el correcto, es muy poco probable que el resultado sea el esperado. El segundo problema es que el condicionamiento a datos reales es bastante lento y, en consecuencia, poco práctico.

### 2.2.4. Estimación directa de la probabilidad multivariada

Esta técnica, la cual se detalla en [16], utiliza un enfoque bastante similar al de kriging de indicadores. En ambos métodos, la idea es utilizar la probabilidad condicional de un punto respecto de sus vecinos para realizar la estimación. La diferencia entre estos métodos es la forma en que se obtiene esta probabilidad. Para calcular la probabilidad condicional con este nuevo método, se utiliza un modelo de Markov en tres dimensiones. La diferencia de este modelo con aquél en una dimensión es que al ser un modelo en más dimensiones pierde la simplicidad de depender sólo del estado anterior. Con este modelo de Markov multipaso y multidependencia la probabilidad condicional corresponde a una función aplicada sobre la probabilidad bivariada entre los puntos involucrados, como se muestra en la Ecuación 2.37, lo cual permite ser calculada de manera más simple.

$$P(x_0|x_1, \dots, x_n) = f(p(x_i, x_j) \quad i, j = (0, \dots, n)) \quad (2.37)$$

Este algoritmo tiene algunos problemas en la práctica. el primero de ellos es la cantidad de datos condicionantes, pues más datos condicionantes implica mayor tiempo de cálculo, y a su vez, menos datos condicionantes implica una estimación más pobre de la probabilidad, generándose, según los propios autores, algunos artefactos en los resultados de la estimación. El otro problema es que la proporción de datos de la distribución univariable no se reproducen, lo cual no es un comportamiento esperable para algoritmos de este tipo.

### 2.2.5. Estimación y simulación basada en la distancia al contorno

En la mayoría de los casos, la simulación de variables categóricas se hace en base al porcentaje de pertenencia a cierta categoría o la probabilidad de pertenecer a ella. En el caso

de variables binarias, es decir dos categorías, existe la alternativa de usar las técnicas geoestadísticas clásicas sobre la variable distancia, la cual mide la distancia al contacto[23][25][4]. Al utilizar esta variable, la cual es continua, se evita ciertos efectos colaterales de usar kriging y simulación de indicadores, como la creación de artefactos en diversas partes del dominio.

Para utilizar la variable distancia al contacto en el modelamiento de una unidad geológica se debe calcular, para cada muestra, la distancia al contacto más cercano. Sin embargo, se debe tener en cuenta que es bastante difícil que los nodos de la grilla tengan distancia 0, es decir, que estén exactamente sobre el contacto. Esto indica que es necesario marcar una diferencia entre los datos al interior del contacto y al exterior, pues en caso contrario no existe una isodistancia con valor 0 que delimite el cuerpo. Para esto se modifica la noción de distancia, permitiendo que existan distancias negativas para así indicar la pertenencia o no al interior del cuerpo en función del signo.

En este trabajo se usa esta técnica con kriging ordinario y con la simulación secuencial gaussiana. Ambos métodos fueron modificados para incluir información blanda en el modelamiento de los cuerpos geológicos.

### 2.2.6. Simulación en presencia de direcciones variables

Incorporar direcciones variables en la estimación y simulación geoestadística convencional es una forma de ayudar el modelamiento cuando la geología se vuelve irregular y las relaciones de pares de puntos no son suficientes. Esta estimación y simulación ocupan el mismo nicho de la simulación multipunto.

Esta técnica fue utilizada en este trabajo para incorporar información blanda en el modelo. La idea de utilizar direcciones variables es que el usuario pueda a través de éstas, indicar las direcciones de mayor continuidad que su experiencia u otras fuentes de información muestran.

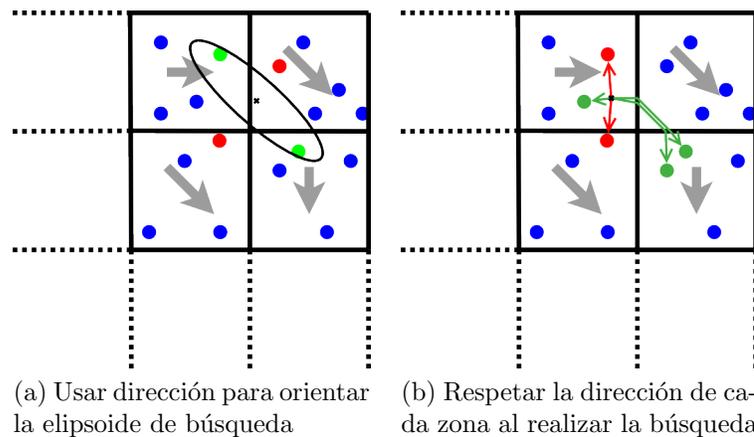


Figura 2.13: Ejemplos de formas de utilizar las direcciones variables en la búsqueda de muestras

## Cambio global en la dirección

Esta primera alternativa[15] utiliza el campo de direcciones variables para cambiar la orientación de la búsqueda y las direcciones del variograma en la simulación secuencial gaussiana y en el kriging. En la Figura 2.13a se muestra lo anteriormente descrito. Este cambio, es suficiente en caso en que el campo de direcciones no posea cambios abruptos, pues sólo se utiliza la dirección del campo en el punto a simular, ignorando las direcciones del campo en otros sectores. Esta solución es bastante simple y aporta mucho al sentido geológico del yacimiento a modelar. El campo de direcciones para este algoritmo está representado por 3 ángulos los cuales indican la rotación respecto a los ejes cartesianos X, Y y Z. Las razones de anisotropía entre los ejes se definen para cada dirección en el campo.

## Cambio local en la dirección

La idea de este algoritmo[17][3][2] es utilizar todo el campo de direcciones variables al momento de trabajar con las distancias entre los puntos. Para esto se define una distancia nueva, la cual corresponde a la distancia en el camino de menor resistencia, donde resistencia se refiere a ir contra la dirección que indica el campo, tal como se muestra en la Figura 2.13a.

En la práctica, para el cálculo de la distancia anisotrópica se utiliza el algoritmo de caminos mínimos en grafos de Dijkstra. Una vez calculadas las distancias se debe cambiar a un espacio de alta dimensionalidad, de manera que la distancia euclidiana para cada par de puntos en el nuevo espacio corresponda a la distancia anisotrópica en el espacio original. Este cambio de espacio es para satisfacer uno de los requisitos de kriging, pues éste opera con distancias euclidianas. Una vez hecho el cambio de espacio el algoritmo continúa de la manera usual. Este algoritmo recibe el campo de direcciones variables por medio de 5 valores para cada celda, los tres primeros corresponden a los ángulos *azimuth*, *dip* y *plunge* y; los dos últimos, a la razón entre la primera y segunda direcciones, y entre la primera y tercera direcciones. En la Figura 2.14 se muestra las rotaciones y escalamientos que generan los valores recién descritos.

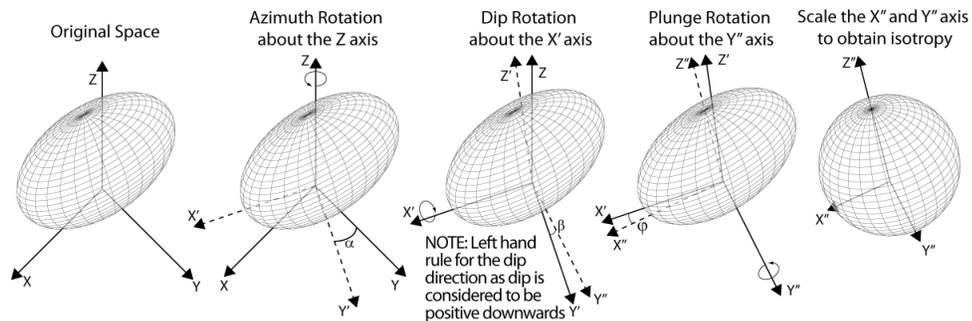


Figura 2.14: Transformación que generan las direcciones variables[2]

## 2.3. Manejo de rugosidad del contacto entre el cuerpo y el exterior

El contacto entre unidades geológicas generalmente suele ser una zona bastante difusa, por lo cual es bastante difícil determinar una superficie que separe de manera absoluta la pertenencia de un bloque a una unidad o a otra. En estas circunstancias, las técnicas que se manejan para determinar esta frontera entre unidades generan contornos de manera paramétrica, permitiendo calibrar los parámetros que la definen mediante técnicas de validación cruzada o *jack-knife*.

La primera técnica revisada[24] introduce un término constante  $C$  que se suma a la distancia al contacto en las muestras, generando un efecto de erosión o dilatación sobre el resultado de la estimación o simulación. Como ya se mencionó, para encontrar el valor correcto para este parámetro se utiliza un mecanismo de calibración mediante *jack-knife*.

La segunda técnica estudiada[21] trabaja sobre el variograma, modificando el modelo utilizado. La idea es considerar el variograma exponencial, gaussiano y pepa puro como casos particulares del modelo que se muestra en la ecuación 2.38, donde el variograma de efecto pepa puro, exponencial y gaussiano corresponden a  $s \rightarrow 0$ ,  $s = 1$  y  $s = 2$ , respectivamente.

$$\gamma(h) = C \cdot \left(1 - e^{-\left(\frac{|h|}{a}\right)^s}\right) \quad (2.38)$$

Este parámetro el cual se mueve entre 0 y 2, genera, para los valores cercanos a 0, cuerpos bastante irregulares e incluso desagregados y; para los valores cercanos a 2, cuerpos más suaves, con fronteras menos rugosas. Al igual que el método anterior, la calibración de este parámetro se efectúa mediante validaciones.

En este trabajo se utiliza una idea similar, donde en lugar de utilizar un parámetro  $s$  como exponente en el variograma, se utilizó la combinación lineal entre un variograma exponencial, uno gaussiano y el efecto pepa. Esta alternativa evita el “acarreo” del parámetro  $s$  a la función que calcula el variograma.

# Capítulo 3

## Modificación de algoritmos y generación de información blanda

### 3.1. *GSLIB*

*GSLIB*[1][5] es un conjunto de programas escritos en la Universidad de Stanford, los cuales realizan una serie de operaciones geoestadísticas como cálculo de variogramas experimentales, variogramas modelados, kriging, varios tipos de simulaciones, anamorfosis, desagrupamiento, entre muchas otras. Adicionalmente posee herramientas básicas para la visualización de muestras, modelos de bloques, variogramas, etc. Estas rutinas se han convertido en una de las principales herramientas para realizar prototipos de nuevos algoritmos y pruebas de concepto. La principal razón de esto es la simplicidad, pues no posee interfaz gráfica, ya que toda la información necesaria se provee mediante un archivo de parámetros. Los programas de *gslib* están escritos en Fortran 77 (existe también una versión en Fortran 90), lo cual dificulta un poco la extensión de ésta debido al bajo nivel del lenguaje y no tener soporte para clases o estructuras tipo C.

En la Figura 3.1 se muestran las componentes de esta librería, junto con la cantidad de líneas de código de cada uno. En la parte superior aparece un módulo llamado *gslib.a* que contiene una serie de funciones que son utilizadas por los ejecutables. En esta librería aparecen con borde rojo aquellas importantes y que son utilizadas por los ejecutables modificados en este trabajo. También se muestra, con relleno amarillo, una nueva función creada e incorporada a la librería. En la parte inferior se encuentran los ejecutables. Estos ejecutables reciben un archivo de parámetros que contiene toda la información necesaria para ejecutarse correctamente. Con relleno verde aparecen los ejecutables modificados en este trabajo. En estricto rigor se creó nuevos ejecutables, sin embargo, estos nuevos ejecutables fueron creados a partir de aquellos indicados con relleno verde, por esta razón se indica como una modificación a los ejecutables y no como nuevas rutinas. Los archivos modificados y sus respectivas cantidades de líneas de código se muestran en la Tabla 3.1.

gslib.a							
ACORNI 53 líneas	CHKTITLE 70 líneas	GAUINV 59 líneas	HEXA 24 líneas	NSCORE 85 líneas	PSFILL 61 líneas	RESC 10 líneas	SQDIST 47 líneas
BACKTR 72 líneas	COVA3 127 líneas	GCUM 24 líneas	HPFYFLTR 147 líneas	NUMTEXT 29 líneas	PSLINE 83 líneas	SCAL 306 líneas	SRCHPATH 145 líneas
BEYOND 363 líneas	DLOCATE 40 líneas	GETINDX 43 líneas	KSOL 137 líneas	ORDREL 106 líneas	PSTEXT 117 líneas	SETROT 76 líneas	SRCHSUPR 175 líneas
BLUE 38 líneas	DPOWINT 20 líneas	GETZ 71 líneas	KTSOL 194 líneas	PICKSUPR 96 líneas	RAND 49 líneas	SETSUPR 113 líneas	STRLIST 40 líneas
CHKNAM 61 líneas	DSORTEM 226 líneas	GREEN 45 líneas	LOCATE 42 líneas	POWINT 18 líneas	RED 45 líneas	SORTEM 225 líneas	STRLEN 17 líneas

ADDCORD 316 líneas	BIVPLT 864 líneas	GAM 1041 líneas	IK3D 1303 líneas	NSCORE 402 líneas	POSTSIM 777 líneas	SCATPLT 618 líneas	SISIM_LM 2008 líneas
ANNEAL 900 líneas	COKB3D 1502 líneas	GAMV 1315 líneas	KB2D 975 líneas	PFSIM 513 líneas	PROBPLT 589 líneas	SCATSMTH 1796 líneas	TRANS 685 líneas
BACKTR 331 líneas	DECLUS 568 líneas	GTSIM 389 líneas	KT3D 1496 líneas	PIXELPLT 771 líneas	QPPLT 671 líneas	SGSIM 2058 líneas	VARGPLT 1141 líneas
BICALIB 694 líneas	DRAW 318 líneas	HISTPLT 939 líneas	LOCMAP 696 líneas	PLOTEM 329 líneas	ROTCOORD 263 líneas	SISIM 2118 líneas	VARMAP 1441 líneas
BIGAUS 449 líneas	ELLIPSIM 344 líneas	HISTSMTH 1034 líneas	LUSIM 980 líneas	POSTIK 616 líneas	SASIM 2895 líneas	SISIM_GS 2154 líneas	VMODEL 250 líneas

Figura 3.1: Esquema de los componentes de *gslib*. Con borde rojo aparecen las subrutinas utilizadas; con relleno amarillo, las subrutinas creadas y con relleno verde, los ejecutables modificados.

	KT3D	SGSIM
ORIG	1496 líneas	2058 líneas
SDH	1719 líneas	2113 líneas
LVO	1664 líneas	2350 líneas

Tabla 3.1: Cantidad de líneas de código para KT3D y SGSIM, para algoritmos originales (ORIG), con pseudosondajes (SDH) y con orientaciones variables (LVO).

## 3.2. Glosario

A continuación se explica algunos acrónimos y rutinas que se utiliza a lo largo de éste y los siguientes capítulos.

- **Sondaje:** Un sondaje corresponde a una perforación realizada en el yacimiento. Estos sondajes tienen principalmente dos funciones. La primera de ellas, que es la de interés para este trabajo, es obtener material del suelo, el cual es analizado químicamente y

que permitan realizar la evaluación del depósito y de la factibilidad para la explotación. La segunda tarea en que se necesita de estos sondeos es para la tronadura. En esta etapa se necesita realizar perforaciones para introducir explosivos en ellos y fragmentar la roca.

- **Pseudosondeo:** Sonda artificial creado a partir de la experiencia del usuario y que explicita la continuidad de la roca caja o del cuerpo mieneralizado.
- **Pozos de tronadura:** Los pozos de tronadura corresponden a perforaciones hechas en la mina para introducir explosivos. Los detritos o testigos obtenidos de esta perforación se pueden utilizar para obtener más información y mejorar el modelo inicial. En el caso de este trabajo se utilizan estos datos para realizar la validación del modelo.
- **Deriva:** La deriva corresponde a una tendencia en los datos que está ligada a la posición de éstos. La deriva indica que la variable en estudio no es estacionaria y puede que algunas técnicas no funcionen del todo bien en este escenario o que deban ser modificadas para poder lidiar con esto.
- **Realización:** Una realización corresponde a un experimento de una variable aleatoria. En el caso de la simulación se entiende por una realización a los valores simulados siguiendo el camino aleatorio. En general, en una ejecución del algoritmo se obtiene  $n$  realizaciones, es decir  $n$  valores simulados para cada nodo en la grilla de trabajo.
- **SDH:** corresponde a utilizar pseudosondeos en la estimación o simulación. La sigla corresponde a *Soft DrillHole*, que en español quiere decir, “sondeo blando”.
- **LVO:** corresponde a la utilización de direcciones variables para la búsqueda y en la orientación del variograma. La sigla significa *Locally Varying Orientation*, que corresponde a “Orientaciones localmente variadas” en español. Este acrónimo está basado en LVA, pues pretende ser un paso intermedio entre la estimación o simulación con una anisotropía global y la aplicación de direcciones variables de manera local y exhaustiva, que corresponde a LVA.
- **LVA:** es la sigla en inglés para “Anisotropías localmente variadas” (*Locally Varying Anisotropy*). Esta metodología utiliza direcciones variables en la estimación y simulación convencional. Más detalles se pueden encontrar en la Sección 2.2.6 del Capítulo “Antecedentes”.
- **SETROT:** Subrutina de gslib que dados 3 ángulos en grados y dos razones de anisotropía, crea una matriz de transformación, la cual se utiliza para crear la estructura de superbloques para la búsqueda y para calcular distancias considerando la anisotropía global.
- **SRCHSUP:** Subrutina de gslib que utiliza la estructura de superbloques para realizar una búsqueda sobre las muestras disponibles.
- **HPFYFLTR:** Subrutina creada que ordena las muestras en función de la distancia, considerando la matriz de rotación. Esta subrutina elimina las muestras que se encuentran fuera de la elipsoide que define la matriz de rotación y el radio de búsqueda. Idealmente debe ser utilizado sobre un conjunto pequeño de muestras.

### 3.3. Modificación de la rutina KT3D

La rutina KT3D es la implementación de kriging en 3D que provee *gslib*. Esta rutina permite realizar kriging simple, ordinario, con deriva, entre muchos otros. Como todo programa de *gslib* recibe un archivo con los parámetros para la ejecución de éste. En la Figura 3.2 se muestra el archivo de parámetros para la rutina KT3D. A continuación se explica los parámetros relevantes del archivo:

**Línea 5:** Ruta al archivo con las muestras.

**Línea 6:** Columnas para las variables  $x$ ,  $y$ ,  $z$ , la variable en estudio y variables secundarias.

**Línea 11:** Nivel de debugging, 0 el más bajo, 3 el más alto.

**Línea 13:** Ruta al archivo con los resultados.

**Líneas 14 a 16** Grilla de estimación. Cada línea tiene: cantidad de nodos, coordenada del primer bloque y tamaño del bloque. Las líneas corresponden a  $x$   $y$  y  $z$ , respectivamente.

**Línea 17:** Discretización del bloque. Corresponde a la cantidad de puntos de discretización con las que se estima el bloque.

**Línea 18:** Cantidad mínima y máxima de muestras para usar en la estimación.

**Línea 20:** Radios de búsqueda para los ejes rotados  $y''$ ,  $x''$  y  $z''$ .

**Línea 21:** Rotación de los ejes de búsqueda siguiendo la convención de la Figura 2.14.

**Líneas 27 en adelante:** Variograma usado para la estimación. En la primera línea se detalla la cantidad de estructuras anidadas y la magnitud del efecto pepa. Las dos líneas siguientes indican los parámetros para la estructura: el primer valor corresponde al tipo (esférico, exponencial, gaussiano o potencia), el segundo a la contribución de meseta (ponderador que multiplica a la estructura), a continuación la rotación de los ejes para dicha estructura y, en la línea siguiente, los alcances en las direcciones  $y''$ ,  $x''$  y  $z''$ . Estas dos últimas líneas deben repetirse la cantidad indicada de estructuras. La cantidad de estructuras máxima es 4.

En la Figura 3.3a se muestra un diagrama simplificado del funcionamiento del algoritmo. Éste comienza con la lectura de los parámetros y, posteriormente, de los archivos con los datos a utilizar. El siguiente paso corresponde a la creación de la matriz de rotación (en estricto rigor es rotación y escalamiento), la cual se utiliza para crear una estructura de superbloques que se usa en la búsqueda. Esta transformación también se utiliza para calcular la distancia entre dos puntos considerando la anisotropía global. Para esto basta aplicar la transformación sobre los dos puntos involucrados y después calcular la distancia de la manera usual. Después de la creación de la estructura de búsqueda se realiza la estimación para cada punto en la grilla que cubre la zona de estudio. El proceso de estimación para cada punto consiste en la búsqueda de muestras cercanas, las cuales se utilizan para construir el sistema de kriging (Ecuaciones 2.19 y 2.27). Por medio de este sistema se puede obtener los pesos  $\{\lambda_\alpha\}_{\alpha=1}^n$  que permiten, finalmente, calcular el valor estimado  $Z^*(x_0)$ .

```

1          Parameters for KT3D
2          *****
3
4  START OF PARAMETERS:
5  ../data/cluster.dat          -file with data
6  0 1 2 0 3 0                 - columns for DH,X,Y,Z,var,sec var
7  -1.0e21 1.0e21              - trimming limits
8  0                             -option: 0=grid, 1=cross, 2=jackknife
9  xvk.dat                      -file with jackknife data
10 1 2 0 3 0                   - columns for X,Y,Z,vr and sec var
11 3                             -debugging level: 0,1,2,3
12 kt3d.dbg                    -file for debugging output
13 kt3d.out                     -file for kriged output
14 50 0.5 1.0                  -nx,xmn,xsiz
15 50 0.5 1.0                  -ny,ymn,ysiz
16 1 0.5 1.0                   -nz,zmn,zsiz
17 1 1 1                       -x,y and z block discretization
18 4 8                         -min, max data for kriging
19 0                             -max per octant (0-> not used)
20 20.0 20.0 20.0              -maximum search radii
21 0.0 0.0 0.0                 -angles for search ellipsoid
22 0 2.302                      -0=SK,1=OK,2=non-st SK,3=exdrift
23 0 0 0 0 0 0 0 0            -drift: x,y,z,xx,yy,zz,xy,xz,zy
24 0                             -0, variable; 1, estimate trend
25 extdrift.dat                -gridded file with drift/mean
26 4                             - column number in gridded file
27 1 0.2                        -nst, nugget effect
28 1 0.8 0.0 0.0 0.0           -it,cc,ang1,ang2,ang3
29 10.0 10.0 10.0              -a_hmax, a_hmin, a_vert

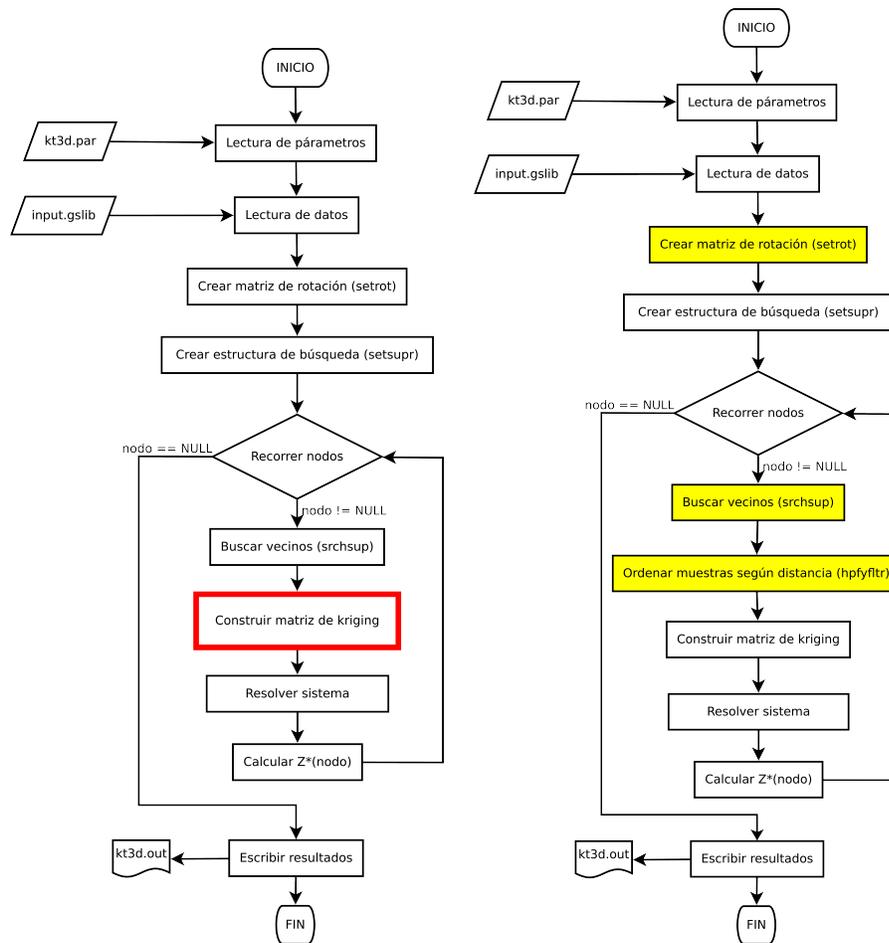
```

Figura 3.2: Archivo de parámetros de KT3D

### 3.3.1. Manejo de Datos

La primera modificación hecha a la rutina KT3D fue al archivo de parámetros. Éste se modificó con el fin de incluir nuevas fuentes de datos. Para el caso de KT3D\_SDH se agregó a los parámetros la ruta a un nuevo archivo, el cual contiene la información de pseudosondajes. Para este archivo se necesita también las columnas en las cuales se encuentra cada dato. Esta nueva información debe ser almacenada con cuidado, pues, posteriormente, en la construcción del sistema, se debe saber si el origen de las muestras es un pseudosondaje o un sondaje real. Para esto se almacenó los datos de la misma manera que los datos reales manteniendo una variable que indica su origen. Esta solución permite que se utilice la misma estructura para la búsqueda y no complica más de la cuenta el resto del procedimiento. El algoritmo nuevo asume que las coordenadas de la información blanda no coinciden con las coordenadas de la información real, en caso contrario el sistema de kriging se indefine. Para evitar que existan datos con las mismas coordenadas que otro se creó un script que descarta los datos a distancias menores a 1m de otro dato.

Para el caso de KT3D\_LVO sólo fue necesario agregar al archivo de parámetros la ruta al nuevo archivo, los índices de las columnas para los datos necesarios y los datos de la grilla donde se mapean las direcciones. Las variables que debe incluir el archivo corresponden a los ángulos *azimuth*, *dip* y *plunge* en grados, y a las razones de anisotropía en el plano  $X''Y''$  (rotado) y en la dirección ortogonal a éste. Estos datos se mapean a la grilla de la misma manera que la grilla de estimación, es decir, variando primero  $x$ , luego  $y$  y finalmente  $z$ . Es requisito que la grilla de las direcciones variables cubra la grilla de estimación, mas no es necesario que sean iguales (número de nodos, origen y tamaño de bloque). También se modificó la definición de la búsqueda y de la variografía, pues las direcciones y razones de



(a) Diagrama KT3D original. En rojo se indica donde se hace la modificación para trabajar con Pseudosondajes.

(b) Diagrama KT3D modificado para trabajar con direcciones variables (LVO).

Figura 3.3: Diagramas de funcionamiento para KT3D original, KT3D\_SDH y KT3D\_LVO

anisotropías se indican en el campo de direcciones para cada zona, luego es necesario sólo definir el radio de búsqueda y el alcance del variograma en la dirección principal ( $y''$ ), pues las direcciones restantes dependen del campo de direcciones en la zona que corresponda y se calculan internamente.

### 3.3.2. Nueva rutina KT3D\_SDH

En la Figura 3.3a se muestra el diagrama de la rutina KT3D original, en el cual aparece remarcado con color rojo dónde se realizan los cambios necesarios para que la nueva información sea utilizada de manera correcta. En la construcción del sistema que se debe resolver para obtener los pesos de kriging, es necesario poder discriminar los datos que son reales de los interpretativos. Una vez obtenidos las muestras se construye la matriz de la manera usual, mas para los datos provenientes de los pseudosondajes, se resta el efecto pepa

a la covarianza entre la muestra y la posición a estimar, en el lado derecho del sistema. En la Figura 3.1 se muestra el sistema modificado de kriging ordinario para estimar el punto  $x_0$ , donde las muestras  $\{x_i\}_{i=1}^n$  corresponden a los datos reales y  $\{x_i\}_{i=n+1}^m$  corresponden a los datos de pseudosondaje. En la Sección 2.1.2 del Capítulo “Antecedentes” se explica el kriging factorial, teoría que soporta estos cambios.

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_{n+1}) & \cdots & C(x_1 - x_m) & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ C(x_{n+1} - x_1) & \cdots & C(x_{n+1} - x_{n+1}) & \cdots & C(x_{n+1} - x_m) & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ C(x_m - x_1) & \cdots & C(x_m - x_{n+1}) & \cdots & C(x_m - x_m) & 1 \\ 1 & \cdots & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{n+1} \\ \vdots \\ \lambda_{n+m} \\ \mu \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_{n+1} - x_0) - C_0 \\ \vdots \\ C(x_m - x_0) - C_0 \\ 1 \end{pmatrix} \quad (3.1)$$

Es importante destacar que al tener dos fuentes de datos existe una mayor probabilidad de que un punto exista en ambas fuentes, lo cual trae un problema al kriging, pues habría dos filas de la matriz que serían linealmente dependientes, luego la matriz sería singular y el punto no sería estimado. Existe también la posibilidad que dos puntos no sean el mismo, sin embargo, se encuentren demasiado cerca de otro, lo cual genera la llamada singularidad no exacta, donde el determinante de la matriz no es cero, sin embargo, es muy pequeño. Esta singularidad no exacta trae problemas al momento de invertir la matriz, debido a errores relacionados a la coma flotante, generando pesos erróneos y, por lo tanto, una estimación incorrecta. Como se mencionó, el algoritmo asume que no existen datos duplicados o demasiado cercanos en los archivos en conjunto. Para satisfacer esta condición se realiza un preproceso donde se descarta los datos que coincidan o esten muy cerca de otro, y donde los datos blandos tienen la prioridad para ser descartados por no ser reales.

### 3.3.3. Nueva rutina KT3D\_LVO

Para entender de mejor manera los cambios, la Figura 3.3 muestra la estructura de rutina original y la rutina modificada. En la Figura 3.3b se muestran coloreados de color amarillo los cambios realizados respecto a la Figura 3.3a. La idea en esta nueva rutina es crear la estructura de búsqueda considerando que no existe anisotropía para descartar una gran cantidad de muestras y luego buscar con las anisotropías definidas localmente de manera exhaustiva sobre los puntos no descartados o candidatos. La primera búsqueda, o búsqueda isótropa, utiliza el radio mayor de la elipsoide, pues una bola de radio  $R$  contiene a todas las elipsoide cuyo radio mayor es  $R$ . Este procedimiento se explica en la Figura 3.4.

El primer cambio corresponde a la creación de la matriz de rotación antes de crear la estructura de superbloques. La matriz de rotación, como se explicó anteriormente, tiene efectos sobre la estructura de superbloques que utiliza la rutina para buscar las muestras. En este caso se crea la matriz con todos los ángulos iguales a  $0^\circ$ , y las razones de anisotropías iguales a 1, obteniendo una estructura de búsqueda sin anisotropías.

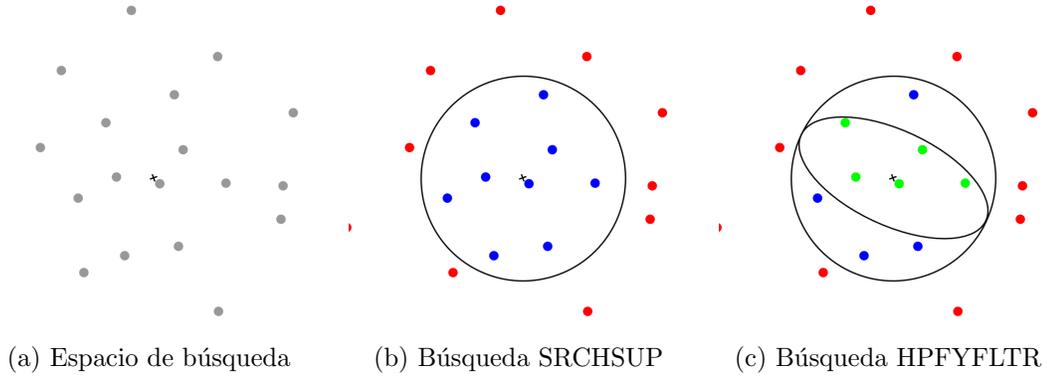


Figura 3.4: Ejemplo de búsqueda para rutina KT3D\_LVO en 2D

Como se puede intuir, el tercer cambio (considerando el orden que define el flujo de algoritmo) corresponde a la incorporación de la subrutina HPFYFLTR, la cual revisa los resultados de la búsqueda SRCHSUP, eliminando los puntos fuera de la elipsoide de búsqueda y ordenando crecientemente según la nueva distancia que considera las anisotropías correspondientes al nodo. El nombre de la subrutina es HPFYFLTR, pues se utiliza el algoritmo *heapsort* y se filtra en la etapa de creación del *heap* o *heapify*. Este proceso necesita que la matriz de rotación considere las anisotropías correspondiente al bloque, luego es necesario llamar previamente a la subrutina SETROT, que modifica la matriz de rotación. Dentro de las cosas que se propuso cambiar también se considera la orientación del variograma según las direcciones de anisotropía, sin embargo, el variograma utiliza también matrices de rotación para manejar sus anisotropías, luego sólo basta con modificar dichas matrices de rotación y los variogramas utilizarán las anisotropías locales. Como se mencionó, la idea es realizar la búsqueda original, que llamaremos SRCHSUP, con una matriz que represente un espacio isótropo, y posteriormente se revisa de manera exhaustiva los resultados obtenidos utilizando las anisotropías locales. El segundo cambio tiene relación con esta matriz de rotación y la búsqueda SRCHSUP, pues se necesita que la matriz de rotación sea isótropa (ángulos iguales a  $0^\circ$  y razones iguales a 1) para la búsqueda SRCHSUP, pero como más adelante en el algoritmo la matriz cambia, pues se consideran las direcciones locales, es necesario asegurar que la matriz de rotación utilizada antes de la búsqueda SRCHSUP sea isótropa, lo que se traduce en agregar una llamada a la rutina SETROT, para generar la matriz de rotación isótropa al inicio de cada ciclo.

En resumen, para cada nodo a estimar, se llama a la subrutina SETROT, para modificar la matriz de rotación, utilizando como argumentos de esta subrutina los ángulos y razones que definen un espacio isótropo. Posteriormente, se realiza la búsqueda con SRCHSUP la cual descarta todos los puntos que están fuera de la bola de radio  $R_{mayor}$  y, en consecuencia, también están fuera de la elipsoide de búsqueda. Estos corresponden a los puntos marcados con rojo en la Figura 3.4c. A continuación se vuelve a modificar la matriz de rotación global y las matrices de rotación de los variogramas, esta vez para que dichas matrices consideren la anisotropías locales para el nodo a estimar. Una vez modificada la matriz se utiliza la subrutina HPFYFLTR para ordenar las muestras según la distancia considerando las direcciones de anisotropía y eliminando las muestras que quedan fuera al utilizar esta nueva distancia. En la Figura 3.4c, los puntos de color azul corresponden a los nodos descartados por HPFYFLTR y

los verdes a los que finalmente son utilizados. En adelante el algoritmo sigue tal como antes, es decir, se utilizan las muestras para crear un sistema y se calcula el estimador con los pesos que resuelven dicho sistema.

Este algoritmo funciona bajo una serie de supuestos. El primero de ellos es que no es necesario utilizar las direcciones variables para calcular el variograma. Este no es un problema para el caso que se estudiará, pues la cantidad de datos es muy pequeña y, en lugar de calcular el variograma, se impone una estructura y se corrige con el variograma experimental. En caso contrario existe la posibilidad de utilizar una rutina adaptada para trabajar con campos de direcciones variables en el cálculo del variograma experimental, incluida en los archivos de LVA. Un supuesto también importante es que el efecto pepa se mantiene constante, al igual que el alcance en la dirección de mayor anisotropía. Estos últimos supuestos tienen sólo una razón: simplicidad, pues no es difícil de implementar, sin embargo, no existe tampoco motivos para hacer variar estos parámetros. Se asume que la zona en estudio corresponde a un espacio que, en el pasado, tuvo una anisotropía global y que, debido a los esfuerzos en la zona, se deformó generando pliegues y fallas. De la naturaleza de estas deformaciones se desprende que el variograma debiese ser en esencia el mismo, cambiando sólo las direcciones de anisotropía.

### 3.4. Modificación de la rutina SGSIM

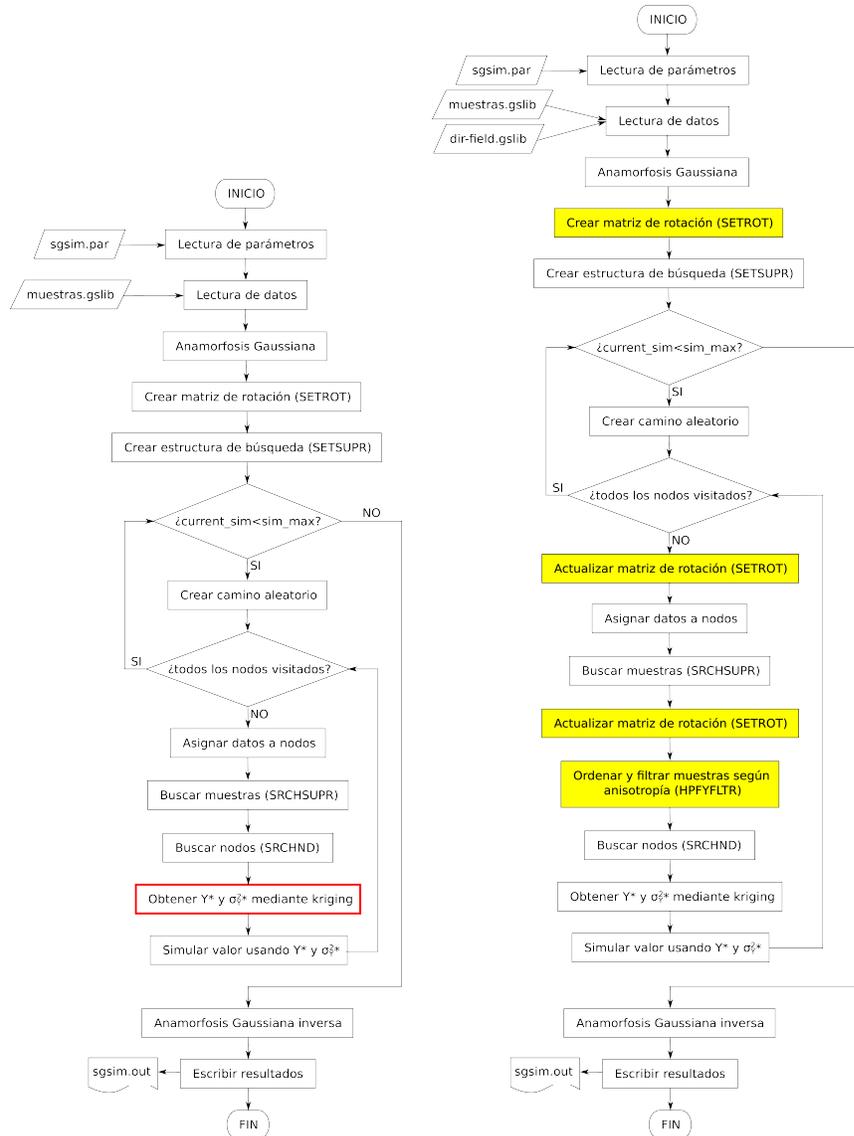
La rutina SGSIM, encargada de realizar la simulación secuencial gaussiana, utiliza el algoritmo de kriging para obtener un valor estimado y varianza para, posteriormente, simular un valor con una distribución gaussiana basada en estos parámetros para su media y varianza condicionales.

El proceso completo se muestra en la Figura 3.5a, donde se puede ver que es bastante similar a la rutina KT3D. Al igual que en ésta, el proceso comienza con la lectura de parámetros y de los archivos de datos. Posteriormente se crea la matriz de rotación y la estructura de búsqueda para dar paso a la simulación en sí. El primer cambio respecto a KT3D es que en este caso se puede efectuar varias realizaciones, es decir simular varias veces para cada nodo, pues las simulaciones tienen una componente pseudoaleatoria que permite que cada realización sea distinta a las otras, lo cual no ocurre en kriging. En cada realización de la simulación se debe recorrer los nodos de la grilla según un camino aleatorio. Para simular cada uno de estos nodos se busca las muestras cercanas y, adicionalmente, los nodos previamente simulados cercanos. Con estos valores se calcula el valor estimado  $Y^*(x)$  y la varianza de kriging  $\sigma_Y^*(x)^2$  para, finalmente, simular un valor según la fórmula de la Ecuación 3.2.

$$Y_{SGS}(x) = Y^*(x) + g \cdot \sigma_Y^*(x)^2 \quad g \sim N(0, 1) \quad (3.2)$$

La rutina SGSIM tiene como opción que las muestras se asignen al nodo más cercano, de esta manera, la búsqueda de muestras y de nodos ya simulados se reduce a una sola búsqueda. Además, dado que kriging es un interpolador exacto, asignar las muestras al nodo más cercano no genera mayores problemas, pues al calcularlo estrictamente el valor debiera

ser, aunque no necesariamente igual, bastante similar al asignado arbitrariamente. En principio, esta opción genera una mejora en el tiempo de cálculo, que depende del número de muestras a utilizar. También, al asignar las muestras a los nodos se pasa de nodos libres en el espacio a nodos en una grilla regular, lo cual permite búsquedas más inteligentes y rápidas. Además, asignar las muestras a nodos evita las ya mencionadas singularidades no exactas.



(a) Diagrama SGSIM original. En rojo se indica donde se hace la modificación para trabajar con Pseudosondajes. (b) Diagrama SGSIM modificado para trabajar con direcciones variables (LVO).

Figura 3.5: Diagramas de funcionamiento para SGSIM original, SGSIM.SDH y SGSIM.LVO

Para el caso de la SGSIM, los parámetros relevantes son los mismos que en el caso de KT3D, mas existen parámetros adicionales que es necesario indicar. A continuación se muestra dichos parámetros y su ubicación en la Figura 3.6, donde muestra el archivo de parámetros para SGSIM.

**Línea 16:** Cantidad de realizaciones.

**Línea 20:** Semilla para los números aleatorios utilizados en la simulación.

**Línea 22:** Cantidad de nodos previamente simulados que se utilizarán en la simulación.

**Línea 23:** 1 ó 0 que indica si se asigna previamente las muestras a nodos. Es recomendable, pues al buscar nodos y muestras por separado la matriz de kriging puede tener singularidades o singularidades no exactas.

```
sgsim.par
1  START OF PARAMETERS:
2  ../data/cluster.dat      -file with data
3  1 2 0 3 5 0             - columns for X,Y,Z,vr,wt,sec.var.
4  -1.0      1.0e21        - trimming limits
5  1                       -transform the data (0=no, 1=yes)
6  sgsim.trn              - file for output trans table
7  0                       - consider ref. dist (0=no, 1=yes)
8  histsmth.out          - file with ref. dist distribution
9  1 2                    - columns for vr and wt
10 0.0      15.0          - zmin,zmax(tail extrapolation)
11 1      0.0            - lower tail option, parameter
12 1      15.0          - upper tail option, parameter
13 1                       -debugging level: 0,1,2,3
14 sgsim.dbg             -file for debugging output
15 sgsim.out             -file for simulation output
16 1                       -number of realizations to generate
17 50 0.5 1.0            -nx,xmn,xsiz
18 50 0.5 1.0            -ny,ymn,ysiz
19 1 0.5 1.0            -nz,zmn,zsiz
20 69069                 -random number seed
21 0 8                   -min and max original data for sim
22 12                    -number of simulated nodes to use
23 1                     -assign data to nodes (0=no, 1=yes)
24 1 3                   -multiple grid search (0=no, 1=yes),num
25 0                     -maximum data per octant (0=not used)
26 10.0 10.0 10.0       -maximum search radii (hmax,hmin,vert)
27 0.0 0.0 0.0          -angles for search ellipsoid
28 51 51 11             -size of covariance lookup table
29 0 0.60 1.0           -ktype: 0=SK,1=OK,2=LVM,3=EXDR,4=COLC
30 ../data/ydata.dat     - file with LVM, EXDR, or COLC variable
31 4                     - column for secondary variable
32 1 0.1                 -nst, nugget effect
33 1 0.9 0.0 0.0 0.0    -it,cc,ang1,ang2,ang3
34      10.0 10.0 10.0  -a_hmax, a_hmin, a_vert
```

Figura 3.6: Archivo de parámetros de SGSIM

### 3.4.1. Manejo de Datos

Para trabajar con los datos de pseudosondajes se hizo de la misma forma que en la rutina KT3D, sin embargo, la forma de proveer estos datos fue distinta. En el caso de SGSIM utilizando pseudosondajes, los datos, tanto de muestras como de pseudosondajes, se entregan en el mismo archivo, en lugar de proveerlos en archivos separados, siendo necesario agregar una nueva columna a éste, la cual indica si es un dato duro o no. El motivo del cambio es la serie de procedimientos que se debe hacer previos a la simulación. Los pesos de desagrupamiento y la anamorfosis gaussiana se calculan sobre todo el conjunto de datos, luego se debe trabajar con ellos como un todo. Para el caso de las direcciones variables los datos se introducen de la misma forma que en la rutina KT3D, es decir, se debe agregar el archivo con el campo de direcciones y los índices de las columnas que contienen cada dato.

### 3.4.2. Nueva rutina SGSIM\_SDH

Para incorporar los datos de pseudosondajes se tuvo que trabajar sobre la grilla de simulación y sobre las muestras. En el caso de las muestras se efectúa de la misma manera que para la rutina KT3D: al construir la matriz se modifica el lado derecho del sistema si el dato corresponde a un pseudosondaje. Para el caso de la grilla el cambio consiste en crear una variable indicador al igual que en las muestras, donde se marca los nodos que fueron asignados a partir de una muestra cercana proveniente de un pseudosondaje. Al momento de visitar un nodo durante el camino aleatorio, se verifica que éste esté vacío, en cuyo caso se simula, en caso que el nodo no esté vacío se verifica que sea un dato duro, si no lo es también se simula y, posteriormente, se marca como dato simulado (o duro). Con esto, los datos blandos ayudan en la simulación de los nodos cercanos y, al momento de simular dicha posición, se rectifica. La simulación secuencial gaussiana utiliza el kriging para obtener el valor estimado y varianza y, posteriormente, simular el nodo valor utilizando estos valores. Sin embargo, en términos de código no es tan directo, pues la rutina SGSIM realiza el kriging internamente, luego es necesario modificar este kriging para introducir los cambios en la simulación.

### 3.4.3. Nueva rutina SGSIM\_LVO

La rutina SGSIM\_LVO, al igual que la rutina SGSIM\_SDH, tuvo que ser modificada directamente, pues contiene su propio kriging. Sin embargo, se tuvo que modificar otras piezas de código, pues la simulación utiliza algunas tablas para ahorrar tiempo. Una de estas tablas es la que maneja la covarianza, la cual está pensada en un contexto de una rotación única. La tabla de covarianza es una estructura que aparece en la simulación con el fin de evitar el calculo repetitivo de la covarianza a través del variograma modelado que se entrega como entrada. Esta tabla se vale de que la grilla de simulación es regular y, por lo tanto, los nodos vecinos siempre se ubicarán a distancias múltiplo del tamaño del bloque, generando distancias discretas y no en un continuo como las muestras. Así, la primera vez que se calcula la covarianza para un vecino a distancia  $i \cdot size_x$  en  $x$ ,  $j \cdot size_y$  en  $y$  y  $k \cdot size_z$  en  $z$ , este valor se almacena en la posición  $(i, j, k)$  de la tabla, para ser consultada posteriormente. Cabe mencionar la estacionaridad que se asume para la variable, asegura que esta covarianza no depende de la posición absoluta, sino de la relativa al nodo.

Para el caso de SGSIM\_LVO las direcciones de anisotropía cambian para cada zona, y en consecuencia, el vecino a distancia  $(i, j, k)$  ya no tiene la misma covarianza respecto del nodo anterior, pues la covarianza depende de las direcciones de anisotropía. Como las direcciones cambian constantemente las covarianzas de los vecinos deben ser calculada para cada nodo en la simulación, sin embargo, los modelos permitidos son bastante simples y el impacto de quitar esta optimización es leve.

### 3.4.4. Manejo de la rugosidad

Para implementar un sistema de control de la rugosidad se diseñó un algoritmo que realiza simulaciones variando el efecto pepa y las estructuras del variograma y, mediante

la validación tipo *jack knife* define la mejor combinación. En principio se pensó en crear o modificar una rutina para efectuar todas las realizaciones con todas las modificaciones a la vez, sin embargo, se desechó y se optó por trabajar con scripts en python que generasen archivos de parámetros con cambios en la variografía. Al generar archivos de parámetros con cambios en la variografía se puede comparar las realizaciones de manera directa, pues utilizando la misma semilla toda la secuencia de números aleatorios, ergo, el procedimiento se mantiene idéntico, excepto el variograma y, en consecuencia, la rugosidad.

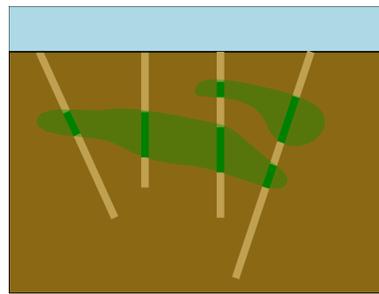
## 3.5. Generación de datos desde la interpretación del usuario

Para generar los datos artificiales se dispone, en principio, sólo de las muestras de sondajes. Éstas corresponden a lo que realmente ocurre en el subsuelo, lo objetivo. Sin embargo, esta información real es utilizada también por el geólogo, para generar una interpretación. Es esta interpretación la que, después de ser contrastada y ajustada con su propia experiencia, se convierte en una segunda fuente de información. Esta información corresponde a un prototipo de modelo que el usuario genera conceptualmente y que debe ser transferido al modelo. Frente a esta situación, el objetivo de los algoritmos que se proponen es visualizar los datos para que el usuario pueda definir en ésta su interpretación.

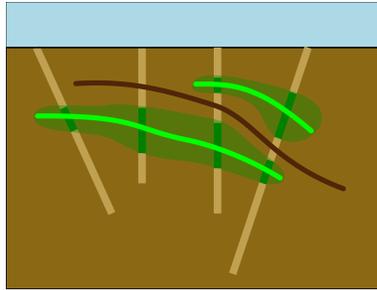
### 3.5.1. Pseudosondajes

La primera alternativa propuesta para que el usuario agregue su interpretación al modelo son los sondajes sintéticos o pseudosondajes. Estos nuevos datos deben tener, como mínimo, el mismo tipo de información que un sondaje real, es decir, coordenadas y un valor para la distancia al contacto, por lo tanto, los pseudosondajes deben estar ligados de alguna manera a los datos duros. Esta restricción indica que el valor de la distancia al contacto para las muestras de los pseudosondajes no deben ser muy distintas al valor en las muestras reales cercanas.

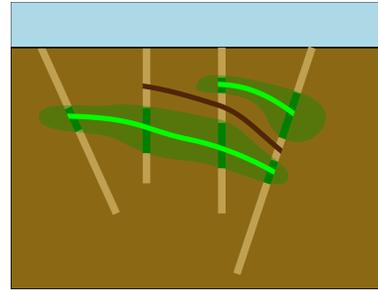
La idea tras el uso de un sondaje artificial es dar a los algoritmos de kriging y simulación secuencial gaussiana un refuerzo acerca de la continuidad en el yacimiento, tanto para el cuerpo mineralizado como para la roca de caja que rodea dicho cuerpo. En la Figura 3.7 se puede ver un esquema en 2D de cómo el usuario debe indicar la continuidad, dada su interpretación a partir de la información de los sondajes. En la Figura 3.7a se pueden ver los sondajes reales y una interpretación posible para el yacimiento. A continuación, en la Figura 3.7b se puede ver, adicionalmente, tres pseudosondajes: dos verde claro indicando la continuidad en los cuerpos mineralizados y uno café indicando la continuidad de la roca de caja. El pseudosondaje café indica que la zona cercana a él corresponde a roca de caja, luego el cuerpo mineralizado por sobre el pseudosondaje y aquél por debajo del pseudosondaje no pueden estar unidos. De la misma manera, cada uno de los pseudosondajes verdes indican, que los bloques en su sector conforman un solo cuerpo mineralizado conexo.



(a) Sondajes e interpretación del usuario



(b) Prototipo de pseudosondajes



(c) Pseudosondajes filtrados

Figura 3.7: Esquema en 2D que explica la idea tras los pseudosondajes.

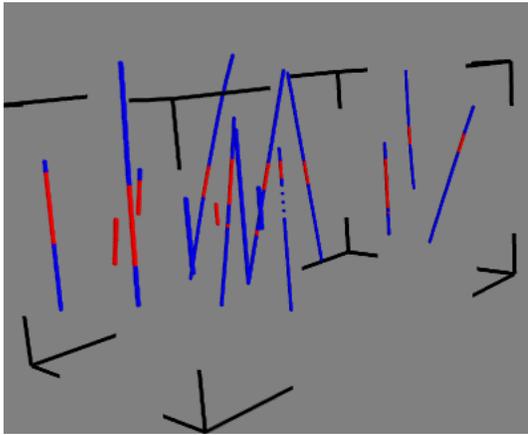
En el ejemplo anterior, se puede notar que los pseudosondajes se intersectan con los sondajes reales. Estas intersecciones separan los pseudosondajes en varios segmentos, estos se discretizan y, finalmente, para cada muestra se obtiene el valor de la distancia al contacto interpolando linealmente dado los valores en los extremos del segmento. En el caso de los dos trozos que quedan fuera, sólo se tiene un valor por lo cual no es posible generar los valores para la distancia al contacto, y por eso estos casos se excluyen del prototipo implementado (en la Figura 3.7c se muestra el ejemplo sin los trozos extremos).

Al extender la idea planteada al caso en 3D, espacio real de trabajo, la idea de dibujar ya no es practicable, pues se debe trabajar con una superficie pues una línea no es suficiente para reforzar la continuidad que se quiere. La extensión diseñada, utiliza una idea similar, que evita los extremos sin valores y que funciona para el espacio en 3D. La solución implementada puede ser en algunos aspectos bastante simple debido a que el yacimiento con el que se trabajó no requiere mayor sofisticación, sin embargo, en principio, el algoritmo es suficiente para cubrir de buena manera todos los casos.

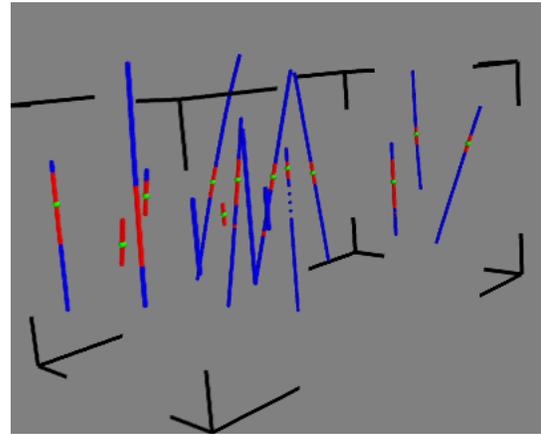
En los párrafos anteriores se impuso una condición sobre los pseudosondajes. Ésta dice que el pseudosondaje debe estar entre dos muestras reales, de manera de poder interpolar la distancia al contacto, utilizando los valores en dichas muestras. Para respetar esta condición, los pseudosondajes se crearán a partir de muestras que el usuario seleccione.

Dicho lo anterior, el primer paso en la creación de los pseudosondajes es la selección de muestras. Para esto se despliega los sondajes reales, donde el usuario puede, a través de la misma pantalla de despliegue, seleccionar a lo más una muestra por sondaje, es decir, para

cada sondaje se puede escoger una muestra o ninguna, como en la Figura 3.8. Esta etapa funciona para cualquier caso, pues sólo se selecciona los datos que se utilizarán para crear los pseudosondajes, lo cual no debiera ser complejo en ningún caso.



(a) Sondajes reales. Azul exterior del cuerpo, rojo el interior del cuerpo



(b) Sondajes reales con muestras seleccionadas en verde

Figura 3.8: Ejemplo de selección de sondajes.

Como se explicó anteriormente, al trabajar en 3D se debe utilizar una superficie para los pseudosondajes. Para generar esta superficie ya se dispone de las muestras seleccionadas por el usuario. El siguiente paso corresponde a generar una triangulación con estas muestras. Para lograr esta tarea se despliegan los datos seleccionados, de manera que el usuario pueda seleccionarlos para crear los triángulos. La idea de este procedimiento es crear los triángulos que definan el borde de la triangulación y el resto de los puntos se inserten de la manera usual. El inconveniente de este proceso es que no es automático, con todo lo que eso implica. Sin embargo, la idea principal es mantener el procedimiento simple y permitir al usuario el mayor poder de manipulación posible.

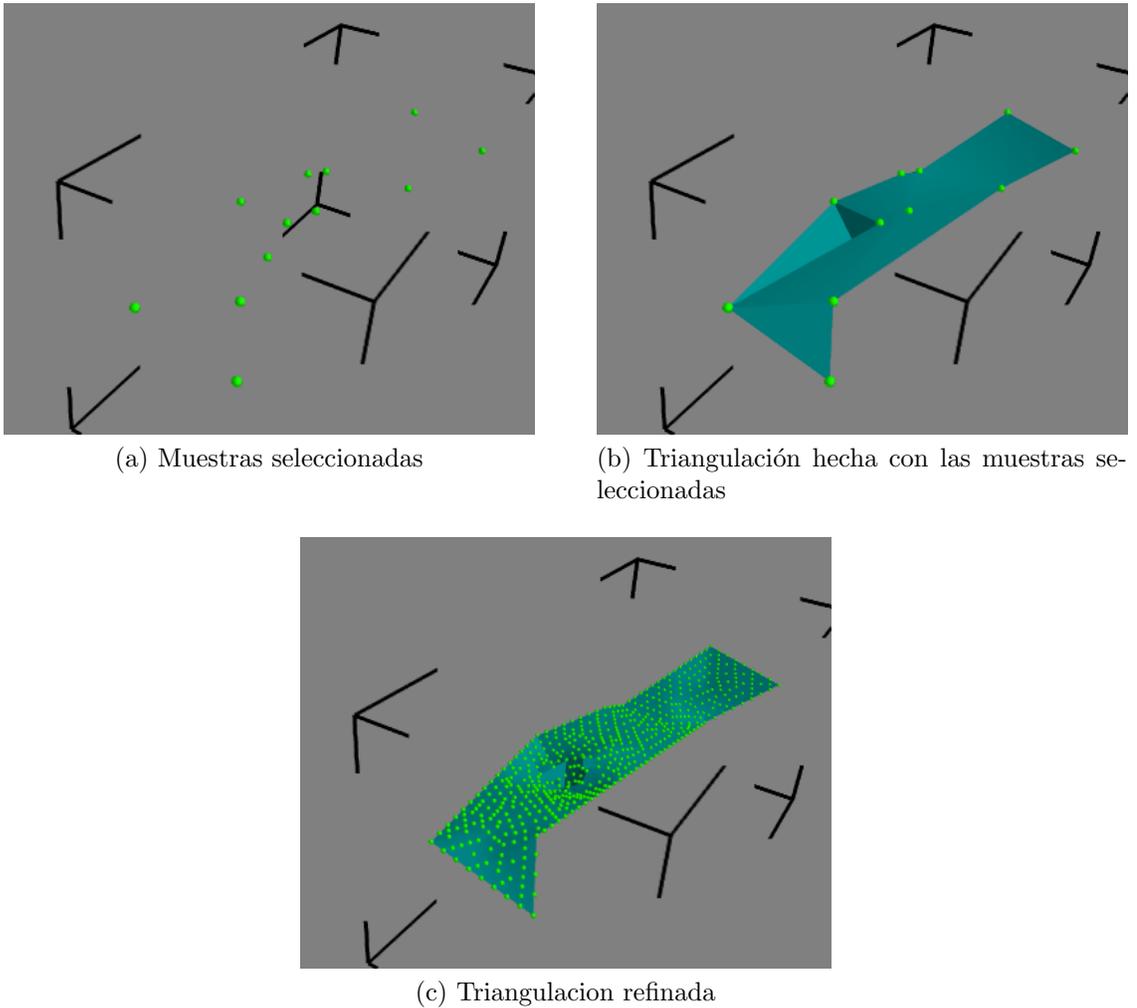


Figura 3.9: Ejemplo de generación de triangulación.

La triangulación obtenida utiliza sólo las muestras seleccionadas por el usuario, las cuales, en el contexto de pocos datos, están bastante espaciadas, además, en estricto rigor, no se ha generado nuevos puntos, sino que se han dispuesto los seleccionados como una triangulación. En este punto es donde comienza realmente la generación del pseudosondaje. Una alternativa es discretizar las aristas de los triángulos tal como en el caso 2D, sin embargo, esta idea es bastante limitada, pues el interior de los triángulos permanecen vacíos, lo que se contrapone a la idea de una superficie discretizada. La solución a esto pasa por refinar la malla, para lo cual existen muchas técnicas. La técnica elegida, por su simplicidad y sus buenos resultados, fue la bisección del lado más largo[20]. Esta técnica de refinamiento toma la arista más larga de la triangulación e inserta el punto medio de dicha arista, con esto el lado largo se divide a la mitad y se generan nuevos triángulos. Esta operación se repite hasta que el lado más largo no supera un umbral establecido. Al generar un nuevo vértice en la triangulación este tendrá, para la variable distancia al contacto, el promedio de la distancia al contacto en los extremos del segmento del que es punto medio el nuevo vértice.

Tanto en la generación de la triangulación como el refinamiento de ésta, no se ha

considerado la calidad de los triángulos obtenidos, pudiendo existir triángulos aguja, los cuales suelen causar diversos problemas. Sin embargo, no es necesario que la triangulación sea de gran calidad, pues finalmente son los vértices de los triángulos como puntos y no los triángulos en si mismos los que se utilizan en la estimación y la simulación. Estos puntos son sometidos a un post-proceso que descarte muestras muy cercanas a otras (arista corta de un triángulo aguja) para evitar las singularidades no exactas que pueden generarse en la matriz de kriging por esta cercanía<sup>1</sup>.

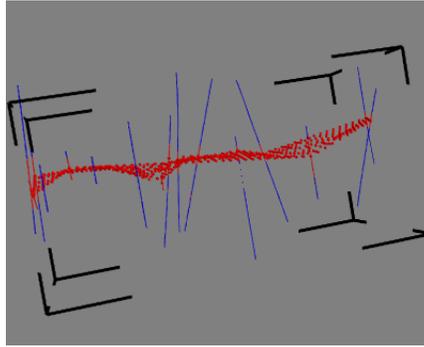


Figura 3.10: Ejemplo de pseudosondaje creado.

Al refinar la malla de triángulos se generan una gran cantidad de nuevos vértices. De dicha malla se recuperan los vértices nuevos que se crearon al refinar, y éstos representarán la discretización de la superficie que el usuario definió. Es importante recalcar que sólo los nuevos vértices se convertirán en la discretización, pues los vértices seleccionados en un principio del algoritmo ya existen en el conjunto de muestras reales, y si se agregaran también al conjunto de datos sintéticos, generarían que la matriz del algoritmo de kriging indefina el sistema. Posteriormente, se aplica el post-procesamiento mencionado en el párrafo anterior.

### 3.5.2. Campo de direcciones

En el caso de los pseudosondajes, la idea consiste en reforzar la continuidad del cuerpo por medio de muestras sintéticas. Usar un campo de direcciones, persigue la misma idea, definiendo las direcciones de anisotropía para cada punto. La diferencia respecto al caso anterior es que no se introduce muestras nuevas, lo cual puede ser malo, en el sentido que puede no ser suficiente para reforzar la continuidad. Sin embargo, esto tiene a su vez la ventaja que no afecta las estadísticas de los datos. Adicionalmente, la construcción del campo de direcciones está sujeta a menos condiciones, pues los datos son de distinto tipo.

Un punto en un campo de direcciones tiene 3 ángulos y 2 razones (se recomienda revisar la Figura 2.14), luego, para cada uno de los puntos en la malla que creará el usuario se debe generar estos 5 valores. Las razones de anisotropía, se asumen como globales, pues se considera que el yacimiento era, en un principio, un cuerpo con una anisotropía uniforme, y

<sup>1</sup>Al existir 2 puntos muy cercanos el determinante de la matriz es muy cercano a 0, ya que las filas correspondientes a estos puntos son casi linealmente dependientes.

por los posteriores esfuerzos, se deformó, pero manteniendo el módulo de la continuidad, esta vez sobre un camino no recto. Dado lo anterior, basta sólo con definir los 3 ángulos que giran el sistema de referencia en cada punto para definir completamente el campo de direcciones.

El algoritmo diseñado para definir los ángulos consiste en dos partes: la primera define la dirección de mayor anisotropía, y la segunda parte define la dirección con la segunda mayor anisotropía. La herramienta también considera una interfaz para que el usuario pueda transformar su interpretación a un campo de direcciones de la manera lo más limpia posible.

En la primera parte se define la dirección de mayor anisotropía para cada nodo en la malla. Al definir esta dirección se está definiendo además 2 de los 3 ángulos que caracterizan la transformación. Los ángulos involucrados son *azimuth* y *dip*. El valor de *azimuth* representa el ángulo en el plano (crece en contra de las manecillas del reloj) entre la dirección de mayor anisotropía (proyectada en el plano) y el eje  $Y$ . El valor de *dip* representa el ángulo en la vertical, entre la dirección de mayor anisotropía y su proyección en el plano (eje  $Y'$ ). Estos 2 ángulos definen la dirección de mayor anisotropía, sin embargo, los 2 ejes restantes, que determinan el plano perpendicular, no quedan definidos.

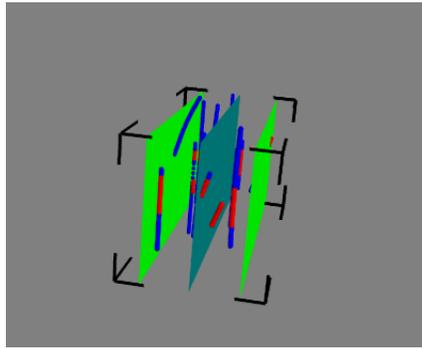
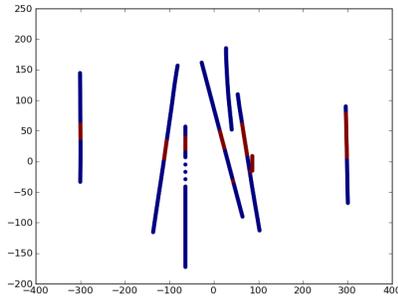


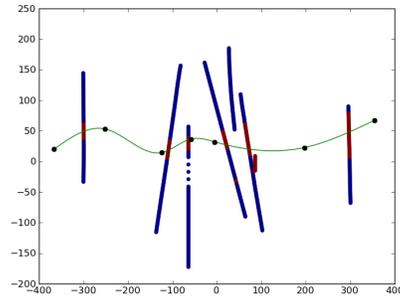
Figura 3.11: Ejemplo de plano de corte y planos que delimitan la zona donde las muestras se consideran.

La primera etapa es la más simple, pues basta con definir la dirección de mayor anisotropía en coordenadas cartesianas y a partir de éstas calcular los ángulos *azimuth* y *dip*. Para crear estas direcciones se elige un plano y se proyectan las muestras cercanas en él. Posteriormente se dibuja por medio de una spline cúbica<sup>2</sup> la dirección que el usuario puede inferir de la disposición de las muestras en el espacio. La curva obtenida se muestrea y se deriva en dichos puntos, con lo cual se puede obtener las coordenadas para el vector que representa la dirección. Estas coordenadas corresponden a muestras, las cuales se utilizan para estimar, mediante kriging ordinario, los nodos en la grilla del campo de direcciones. Para generar varias muestras se repite el proceso las veces que se estime pertinente. La estimación se realiza para cada componente del vector dirección. Posteriormente, se normaliza las componentes para finalmente obtener los ángulos y el plano que define la dirección calculada.

<sup>2</sup>En principio se pensó en curvas tipo Bezier para permitir flujos cíclicos, sin embargo, basta con splines, pues este tipo de flujos se puede crear mediante 2 o más ejecuciones (por ejemplo, un círculo: se hace primero la parte superior y después la inferior).



(a) Proyección de las muestras cercanas en el plano definido por el usuario



(b) Spline creada a partir de las muestras cercanas

Figura 3.12: Ejemplo de proyección de las muestras y spline generada.

La segunda etapa consiste en obtener la dirección secundaria mediante el mismo proceso: se definen las muestras, se realiza un kriging ordinario sobre cada componente de la dirección y, finalmente, se normaliza, obteniendo así esta nueva dirección en cada nodo de la grilla. El post proceso en este caso es un poco más complejo, pues las direcciones de anisotropía deben ser ortogonales, y las direcciones obtenidas no necesariamente lo son, luego se debe hacer un ajuste sobre estas direcciones. La dirección principal es la más importante, pues hace que las muestras en dicha dirección tengan mayor influencia, por lo tanto, esta dirección queda fija, y es la dirección secundaria la que se ajusta a la primera.

Como se mencionó, la segunda dirección debe estar contenida en el plano perpendicular a la dirección principal. Para obtener una dirección similar a la que se estimó que esté contenida en el plano, se proyecta ortogonalmente la dirección secundaria estimada en el plano donde debe estar (el plano que define la dirección principal). Una vez obtenida la dirección secundaria ajustada, la dirección que completa la base corresponde al producto vectorial entre las direcciones principal y secundaria. Con las tres direcciones calculadas se puede obtener los ángulos que definen la transformación. Este proceso se hace en cada nodo de la grilla.

# Capítulo 4

## Caso de estudio

El presente estudio tiene por objetivo mostrar resultados para los algoritmos desarrollados en un ejemplo con datos reales. Para esto se utilizó datos del yacimiento Cerro Colorado, perteneciente a la compañía BHP Billiton Pampa Norte. Para este estudio se tomó un porcentaje pequeño de los sondajes disponibles. La idea tras esta elección es recrear el escenario en el cual estos algoritmos tienen sentido, es decir, cuando la cantidad de datos es bastante pequeña. A medida que la cantidad de datos aumenta, la experiencia que el geólogo introduce en los algoritmos debiese hacer sentido con los nuevos datos, luego, la incorporación de información blanda debiese disminuir.

El procedimiento que se lleva a cabo en este capítulo corresponde al proceso típico para la evaluación de yacimientos. Todos los conceptos necesarios para entender el siguiente resumen se encuentran en la Sección 2.1 del Capítulo “Antecedentes”. El primer paso corresponde al estudio exploratorio de datos, en el cual se eliminan las muestras defectuosas y se muestran estadísticas e histogramas de los datos y se despliegan los datos en diferentes vistas. Para este trabajo, en el estudio exploratorio de datos no se descartan muestras, pues la zona de estudio corresponde a una pequeña porción del yacimiento completo y no presenta problemas de datos aberrantes. El siguiente paso en el proceso corresponde a identificar las direcciones de mayor anisotropía en la zona de estudio. Para esto se utilizó un criterio visual, debido a que las herramientas comunmente utilizadas necesitan de una gran cantidad de datos. En el caso de la simulación, el siguiente paso corresponde al desagrupamiento de los datos, es decir, dar pesos a las muestras, asignando pesos menores a las muestras clusterizadas, de manera de corregir muestreos preferenciales, y así poder realizar la transformación a una gaussiana sobre una distribución corregida. El desagrupamiento y transformación a una gaussiana se efectúan previo al análisis variográfico, pues la simulación trabaja sobre datos transformados y toda decisión, supuesto o lo que fuere, debe ser hecha sobre los datos transformados. Para kriging no se realiza las transformaciones mencionadas, porque kriging trabaja sobre los datos originales. Como se mencionó, posterior a la transformación de los datos, se realiza el análisis variográfico sobre éstos. En la práctica, esto consiste en obtener el variograma experimental, y a partir de éste, ajustar un modelo variográfico. Generalmente basta con lo descrito, sin embargo, en ocasiones se debe recurrir a técnicas más complejas. A continuación del análisis variográfico, o también llamado variografía, se ejecuta la estimación o simulación según corresponda, y, finalmente, sólo para el caso de la simulación, se aplica la

transformación inversa para volver a la variable original a partir de la variable gaussiana. El análisis sobre los resultados obtenidos es un tema aparte, pues es atinente a la aplicación. En el caso de este trabajo los análisis se harán respecto a la calidad de la respuesta para los distintos algoritmos en función de una serie de indicadores detallados en la Sección 1.4 en el Capítulo “Introducción”.

## 4.1. Cálculo de la distancia

El primer paso, previo a la selección de los sondeos a utilizar, es realizar el cálculo de la distancia al contacto. Para esto se utiliza la variable correspondiente al tipo de mineral (*minty*), cuyos valores posibles son *Cubierta*, *Lixiviado*, *Óxidos*, *Sulfuros supérgenos* (o simplemente *Sulfuros*), *Sulfuros secundarios y primarios* (o *Mixtos*) y *Sulfuros primarios* (o *Hipógenos*). Con esta variable se crea una nueva, con valor 1 para los óxidos y 0 para los demás tipos de mineral, llamada *oxide*. El contacto, necesario para calcular la distancia, se considera como el punto medio entre dos muestras consecutivas con valores distintos en esta nueva variable *oxide*. En caso de existir contactos, para todas las muestras en dicho sondeo se calcula la distancia al contacto más cercano y se multiplica por  $-1$  si la variable *oxide* es 1. En caso de no existir contactos para un sondeo, la distancia se calcula al más cercano de los contactos en otros sondeos.

En la Figura 4.1 se muestra un breve ejemplo del cálculo anterior para un sondeo con contactos. En la Figura 4.1a se muestra el sondeo, donde el color blanco indica el exterior; y el color negro, el cuerpo mineralizado. En este sondeo, el contacto (línea discontinua) se asume en el punto medio entre muestras vecinas que pertenezcan una al exterior y otra al interior del cuerpo. Para este ejemplo la distancia entre las muestras es de 1 unidad de longitud. En la Figura 4.1b se muestra la distancia entre cada muestra y el contacto más cercano, además las muestras se pintaron según una escala donde el color azul es que representa al valor más bajo y el color rojo al más alto, también se marcó con números rojos las muestras al interior del cuerpo. Finalmente, en la Figura 4.1c se muestra el resultado de dar valores negativos a las muestras al interior del cuerpo.

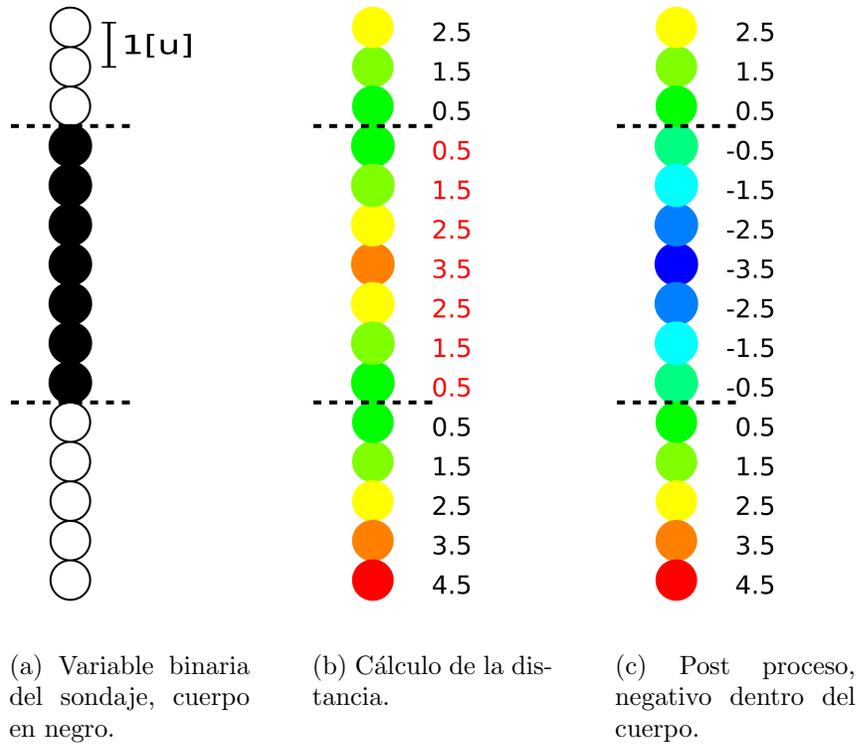


Figura 4.1: Ejemplo de cálculo de distancia.

## 4.2. Datos disponibles

Este trabajo está enmarcado en un proyecto de colaboración entre la Universidad de Chile y la compañía minera BHP Billiton. Dentro de los términos de esta cooperación la empresa facilitó datos para el desarrollo de diversos trabajos de memorias y tesis de postgrado. Los datos disponibles para estos trabajos corresponden a 4 yacimientos. Dos de ellos (Escondida y Escondida Norte) pertenecen a Minera Escondida, empresa que pertenece en un 57.5% a BHP Billiton y están ubicadas cerca de la ciudad de Antofagasta. Los otros dos yacimientos corresponden a Cerro Colorado, ubicado cerca de la ciudad de Iquique y Spence ubicada en el Desierto de Atacama. Estos últimos yacimientos pertenecen a BHP Billiton Pampa Norte, empresa cuyo único dueño es BHP Billiton.

Este trabajo en particular requiere de un yacimiento donde exista plegamientos lo más marcados posibles, de manera que la geometría del yacimiento sea un factor importante al momento de realizar la estimación y/o simulación. De los yacimientos facilitados, mencionados en el párrafo anterior, el que mejor se adapta a este requerimiento es Cerro Colorado. Este yacimiento corresponde a una mina a cielo abierto que ya está siendo explotada, por lo cual se tiene información tanto a priori como a posteriori. Para realizar este estudio se contó con 2 fuentes de información: Los datos reales y el conocimiento geológico.

### 4.2.1. Información Real

Esta información corresponde a los datos reales, proveniente de muestras de sondajes y pozos de tronadura en la zona de estudio. Las muestras de sondajes se dividieron en 2 conjuntos: 10 % para estimación y simulación, y 90 % para la validación de los resultados. Los datos de pozos de tronadura son más abundante, sin embargo, éstos se obtienen a posteriori, durante las fases de explotación, ergo sólo tienen sentido como conjunto de validación.

### 4.2.2. Información Blanda

Para el caso de este estudio, debido a la existencia de información a posteriori y al poco conocimiento geológico del autor, se revisó de manera cualitativa los datos de pozos de tronadura para encontrar tendencias, las direcciones en que se forman los cuerpos, entre otras cosas que permitan entender la estructura del yacimiento y simular el conocimiento experto. En un escenario real este conocimiento corresponde a la experiencia del geólogo, la presencia de estructuras como plegamientos, fallas, intrusiones, tendencia en las leyes, por nombrar algunos fenómenos, que estos profesionales puede inferir de los datos o del estudio previo del yacimiento. De esta información se extraen 2 datos importantes: los pseudosondajes y el campo de direcciones variables.

Para generar los datos de pseudosondajes se utilizó el algoritmo explicado en la sección 3.5.1 con pequeñas modificaciones. Estas modificaciones consistieron en seleccionar un sondaje completo en lugar de las muestras individuales. Estos sondajes tienen tres zonas marcadas, una zona de roca caja en la parte superior, una zona de cuerpo mineralizado y una zona de roca caja en la parte inferior. En la Figura 4.2 la roca caja está representada por las muestras de color café y el cuerpo mineralizado por las muestras de color verde. La idea consiste en seleccionar tres muestras automáticamente en cada uno de los sondajes elegidos. La primera muestra seleccionada se elige en la zona del cuerpo mineralizado y, las dos restantes: una en la roca caja sobre el cuerpo mineralizado, muestras en la elipse número 1) y la última en la roca caja debajo del cuerpo. En la Figura 4.2 se muestra algunos ejemplos de lo anterior. La idea es crear 3 pseudosondajes, uno como el esqueleto del cuerpo (utilizando las muestras que en la Figura 4.2 aparecen encerradas en un rectángulo rojo) uno por sobre el cuerpo (utilizando las muestras encerradas en un rectángulo azul) y uno por debajo (utilizando las muestras encerradas en un rectángulo gris). Al seleccionar un sondaje se elige el punto medio del corte con el cuerpo, en caso de existir dos o más cortes, se elige el más grande. Una vez seleccionado dicho punto, en la zona adyacente, tanto sobre como por debajo del corte con el cuerpo, se selecciona el punto cuyo valor absoluto sea igual al valor en la primera muestra seleccionada. En el sondaje DH2 en la Figura 4.2 se puede ver que la primera muestra, marcada con línea roja discontinua, tiene valor -2.5, luego en la zona adyacente superior se elige la muestra con valor 2.5. En caso de no existir un punto con el mismo valor absoluto que el de la primera muestra seleccionada, se selecciona la muestra con mayor valor en la zona adyacente. En el sondaje DH1 de la Figura 4.2 se puede ver que en los segmentos adyacentes no existe muestra con el valor absoluto igual al de la muestra marcada con rojo, que es -4.5, por lo tanto se selecciona el máximo en ambos segmentos adyacentes (ambos corresponden al mismo caso). El resto del proceso continúa de la forma normal.

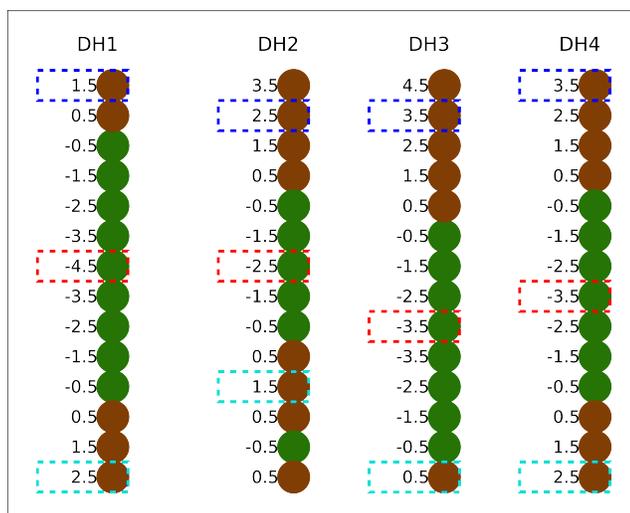


Figura 4.2: Ejemplo de selección de muestras modificado.

Para generar el campo de direcciones también se hizo modificaciones al algoritmo original 3.5.2, ya que las direcciones de anisotropía elegidas corresponden al plano horizontal, ambos con la misma razón de anisotropía, y la dirección normal a este plano, con una razón de anisotropía menor. Dada esta configuración, con dos ejes con la misma razón de anisotropía, basta sólo con estimar la normal al plano para definir la rotación completa. De todas maneras se debe definir los ejes del plano, sin embargo, basta con fijar uno y el otro se obtiene mediante el producto vectorial o producto cruz. En resumen, en lugar de realizar dos estimaciones, una para la dirección principal y una para la dirección secundaria; se realiza sólo una estimación sobre la dirección de menor anisotropía.

### 4.3. Análisis exploratorio de datos

En esta etapa del proceso se toman los datos de sondajes disponibles y se crean los dos conjuntos de trabajo, el primero con un 10 % de los datos para realizar el modelamiento y el segundo con el 90 % para la validación final. La zona de estudio, cuyas coordenadas se muestran en la Tabla 4.1, fue escogido por la curvatura presente en algunas de las capas, geometría para la cual se diseñó los algoritmos de estimación y simulación de distancia con información blanda.

El primer paso para la obtención de los conjuntos de trabajo es separar los sondajes que corten la zona de estudio, es decir, que alguna de las muestras de dicho sondaje pertenezca a esta zona. Del total de los sondajes disponibles sólo 128 cortan la zona de estudio. El resto de los sondajes se descarta. Posteriormente, se calcula la distancia al contacto para todos estos sondajes. Es importante que la distancia se calcule antes de filtrar las muestras que quedan fuera de la zona de estudio. En la Figura 4.3 se muestra un ejemplo donde se calcula la distancia al contacto más cercano antes y después de filtrar las muestras pertenecientes a la zona de estudio. En ésta se puede ver que las distancias cambian y que se

pierde información importante, pues no se considera el contacto que queda fuera de la zona de estudio, distorsionando las distancias. Este problema surge porque los sondeos no son estrictamente verticales. Como ya se mencionó, el siguiente paso corresponde a descartar las muestras que quedan fuera de la zona de estudio. Una vez descartadas dichas muestras se tiene sólo aquellas que pertenecen a la zona de estudio, cada una con su distancia al contacto más cercano. Finalmente, se toma, de forma aleatoria, el 10 % de los sondeos, es decir 13, conformando éste el primer conjunto, áquel utilizado para la estimación y simulación. El resto de los sondeos corresponde al conjunto de validación. Notar que el 10 % de los sondeos no corresponde necesariamente al 10 % de las muestras. Para los datos de pozos de tronadura, el proceso consistió sólo en descartar las muestras que no pertenecen a la zona de estudio, pues en la validación se compara la pertenencia al cuerpo y no la distancia al contacto.

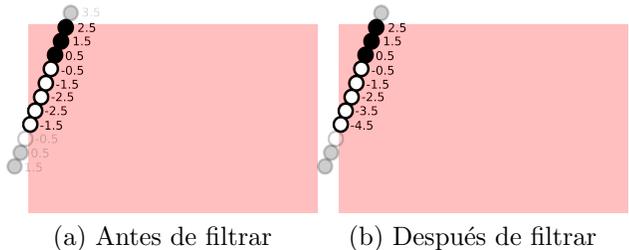


Figura 4.3: Ejemplo del cálculo de la distancia al contacto antes y después de filtrar las muestras.

En la Tabla 4.2 se observan la cantidad de muestras totales, la cantidad que corresponde al cuerpo y la que corresponde al exterior del cuerpo en estudio para distintos subconjuntos creados a partir de los datos de sondeos, en la Tabla 4.3 se muestra las mismas cantidades para los datos de pozos de tronadura.

	Mínimo	Máximo	Diferencia
Este	72000	72600	600
Norte	83450	83750	300
Elevación	2200	2500	300

Tabla 4.1: Coordenadas de la zona de estudio.

	Muestras	Cuerpo	Caja	Proporción $\frac{cuerpo}{total}$
Conjunto de trabajo (10 %)	1065	294	771	0.276
Conjunto de trabajo filtrado	964	294	670	0.305
Conjunto de validación (90 %)	8712	2104	6608	0.241
Conjunto de validación filtrado	7602	2009	5593	0.264
Todos los datos de sondeos (100 %)	9777	2398	7379	0.245

Tabla 4.2: Sondeos: cantidad de muestras.

	Muestras	Cuerpo	Caja	Proporción $\frac{cuerpo}{total}$
Pozos de tronadura	39495	9538	29957	0.318

Tabla 4.3: Pozos de tronadura: cantidad de muestras.

En la Figura 4.4 se muestra una vista isométrica de los datos utilizados, donde se puede observar los datos reales, los pseudosondeos y ambos juntos. En la Tabla 4.4 se muestran las estadísticas para las variables en los datos de sondeos utilizados, es decir, este, norte, elevación y distancia al contacto. Adicionalmente, en la Figura 4.6 se pueden ver los histogramas para estas variables. En la Figura 4.7 se muestran los datos utilizados para validación.

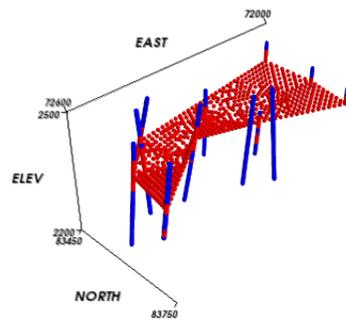
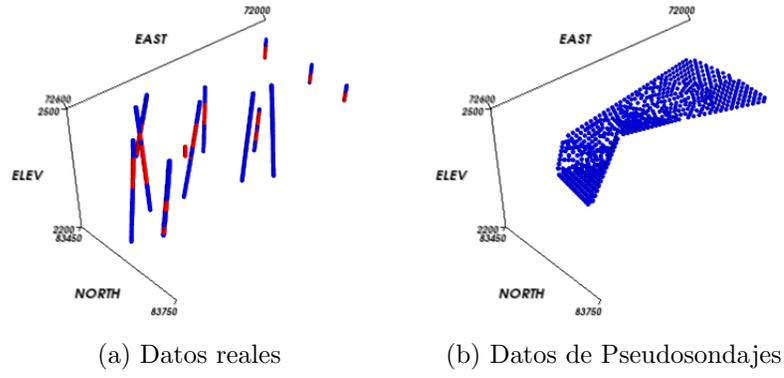


Figura 4.4: Vista isométrica de los datos para modelamiento.

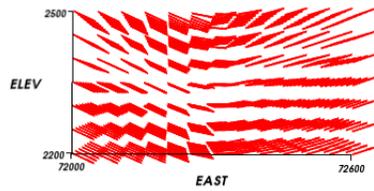


Figura 4.5: Campo de direcciones variables.

	Mínimo	Máximo	Media	Mediana	Desviación
Distancia	-45.00	137.00	25.00	19.00	36.75
Este	72000.78	72599.92	72370.73	72377.08.48	159.09
Norte	83455.15	83715.91	83591.52	83567.82	74.99
Elevación	2200.45	2499.82	2375.02	2379.62	66.38

(a) Conjunto de trabajo (10 %)

	Mínimo	Máximo	Media	Mediana	Desviación
Distancia	-45.00	137.00	25.00	19.00	36.75
Este	72000.78	72599.92	72370.73	72377.08	159.09
Norte	83455.15	83715.91	83591.52	83567.82	74.99
Elevación	2200.45	2499.82	2375.02	2379.62	66.38

(b) Conjunto de trabajo filtrado

	Mínimo	Máximo	Media	Mediana	Desviación
Distancia	-85.00	445.31	42.90	29.20	62.32
Este	71989.55	72601.92	72314.42	72327.76	181.30
Norte	83424.56	83749.54	83578.47	83580.00	76.92
Elevación	1874.20	2563.91	2372.57	2383.13	90.79

(c) Conjunto de validación (90 %)

	Mínimo	Máximo	Media	Mediana	Desviación
Distancia	-85.00	445.31	41.78	29.00	60.68
Este	71989.55	72601.92	72321.63	72334.73	180.14
Norte	83424.56	83749.54	83580.05	83579.93	76.86
Elevación	1874.20	2563.91	2373.34	2383.14	89.58

(d) Todos los datos (100 %)

Tabla 4.4: Estadísticas de las muestras de sondajes.

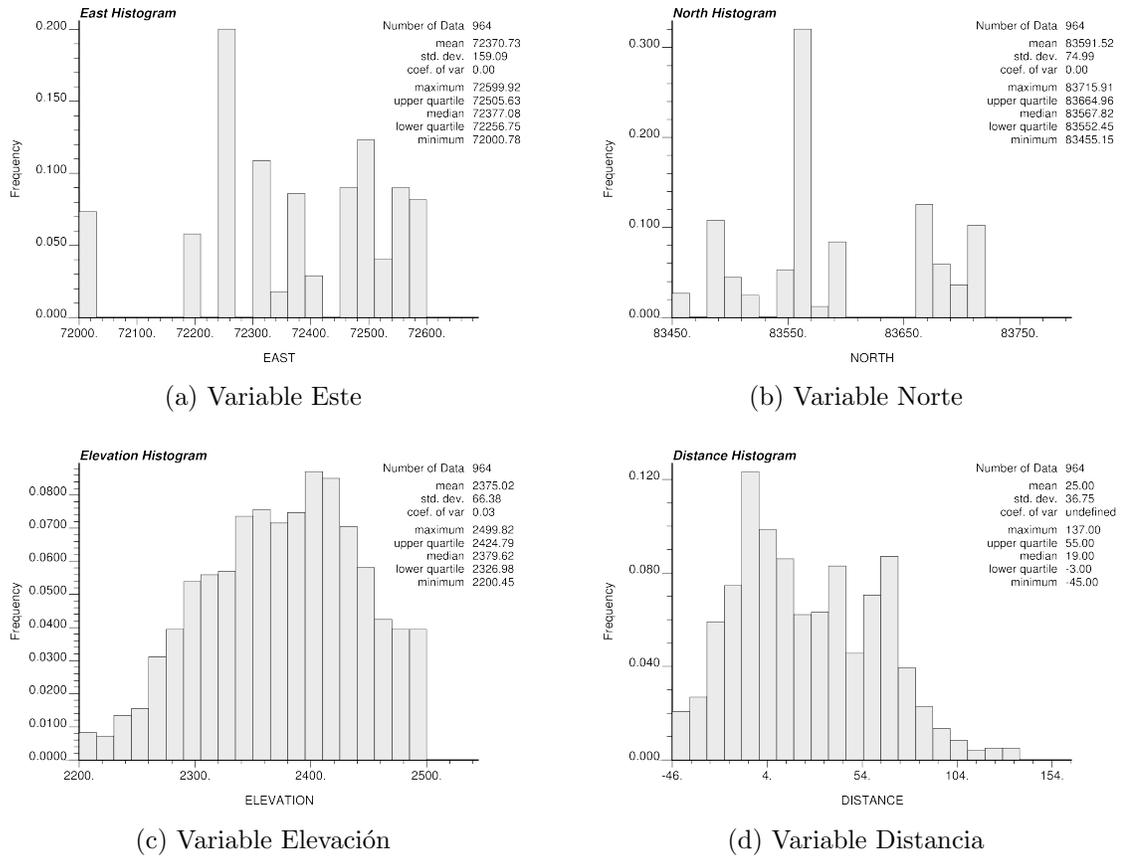


Figura 4.6: Histogramas de las variables utilizadas para datos de sondajes y pseudosondaje.

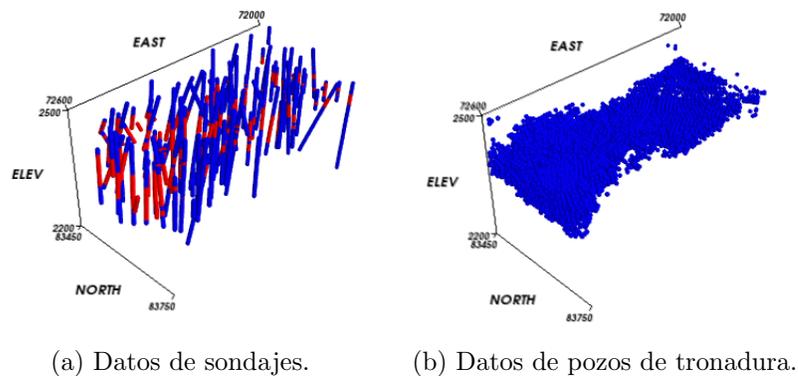
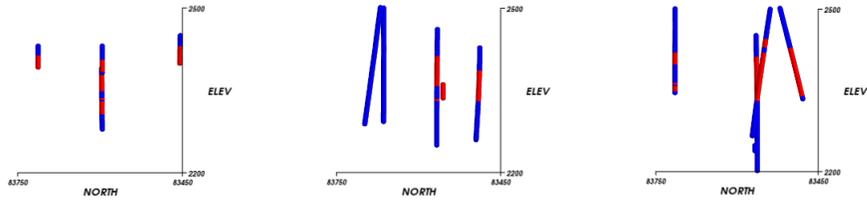


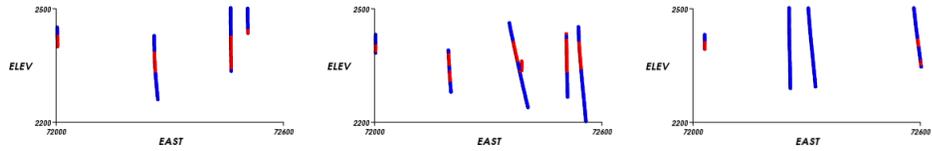
Figura 4.7: Vista isométrica de los datos de validación.

En las Figura 4.8 se muestran tres visualizaciones en perfil NORTE-ELEVACIÓN con profundidad de 200 metros cada una. En la Figura 4.9 se muestran también tres cortes, el perfil utilizado es ESTE-ELEVACIÓN y la profundidad es de 100 metros. Finalmente, en la Figura 4.10 se muestran tres cortes de profundidad 100 metros en el perfil ESTE-NORTE. En estas imágenes se muestran sólo perfiles para los datos reales.



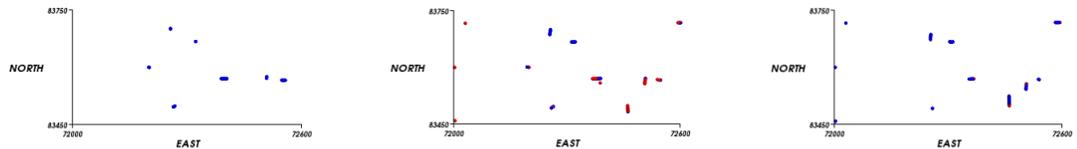
(a) ESTE 72000-72200      (b) ESTE 72200-72400      (c) ESTE 72400-72600

Figura 4.8: Visualización en perfil NORTE-ELEVACIÓN para datos reales



(a) NORTE 83450-83550      (b) NORTE 83550-83650      (c) NORTE 83650-83750

Figura 4.9: Visualización en perfil ESTE-ELEVACIÓN para datos reales.



(a) ELEVACIÓN 2200-2300      (b) ELEVACIÓN 2300-2400      (c) ELEVACIÓN 2400-2500

Figura 4.10: Visualización en perfil ESTE-NORTE para datos reales.

## 4.4. Kriging

### 4.4.1. Variografía

Dada la disposición en forma de capas, es decir, continuidad alta en el plano horizontal (ESTE-NORTE) y más baja en la vertical (ELEVACIÓN), de los minerales de la variable

*minty*, las direcciones de anisotropía elegidas para calcular el variograma experimental (Sección 2.1.1 del Capítulo “Antecedentes”) son la dirección en el plano (u omnihorizontal) y la dirección vertical. En la dirección omnihorizontal la separación entre sondajes no es constante, pues los sondajes del conjunto de modelamiento (10 %) son tomados aleatoriamente, luego la diferencia es también aleatoria. En la dirección vertical la separación está dada por el tamaño elegido al discretizar el sondaje, en este caso cada 2[m].

Es importante tener en cuenta que, debido a que se trabaja con pocos datos, los variogramas experimentales probablemente no sean representativos y tengan mucho ruido, lo cual dificulta el proceso de modelamiento de éstos. En este contexto se utiliza estructuras simples, sin anidación y que represente más el comportamiento esperado para la variable distancia que lo obtenido a partir de las muestras.

El variograma experimental se calcula utilizando los parámetros mostrados en la Tabla 4.5. Para el caso de la dirección omnihorizontal la elección de los parámetros relacionados a los pasos es mediante ensayo y error, pues no se cuenta con una referencia acerca de las distancias entre sondajes, al ser tomados éstos al azar. En el caso de la dirección vertical, se utiliza la distancia entre muestras como referencia y, posteriormente, se varía dicha distancia para obtener una gráfica más regular. En la gráfica de la Figura 4.11 (líneas discontinuas) se muestra los variogramas en el plano y en la vertical, la línea roja corresponde a la vertical y la azul al plano.

	Omnihorizontal	Vertical
N° Lags	15	15
Distancia Lag	25.0	5.0
Tolerancia Lag	12.5	2.5
Azimuth	90.0°	0.0°
Tolerancia Azimuth	90.0°	22.5°
Bandwidth Horizontal	10.0°	10.0°
Dip	0.0°	90.0°
Tolerancia Dip	22.5°	22.5°
Bandwidth Vertical	10.0°	10.0°

Tabla 4.5: Parámetros del variograma experimental para kriging.

Los modelos variográficos o variogramas modelados se confeccionan comenzando con la dirección vertical. El variograma experimental en la vertical (línea roja discontinua en la Figura 4.11) muestra un comportamiento bastante regular, luego es bastante fácil de modelar con un variograma gaussiano<sup>1</sup>. La dirección omnihorizontal se modela basándose en el modelo vertical (línea celeste continua en la Figura 4.11) y extendiendo el alcance del variograma

---

<sup>1</sup>Figura 2.4

para hacerlo calzar con el variograma experimental en el plano (línea azul discontinua en la Figura 4.11). Este procedimiento tiene como fin obtener un resultado similar al que se observa en la vertical, pero con un alcance mayor. Para el caso de la dirección vertical, el variograma experimental (curva roja de la Figura 4.11) muestra un comportamiento sumamente regular debido a que la variable distancia también lo es, pues las distancias están medidas en la dirección de los sondajes, es decir, la vertical. Este mismo comportamiento regular explica que el efecto pepa sea tan pequeño, pues el efecto pepa sólo recoge, para este caso, las muestras a distancias menores que la mitad del paso, es decir, cuyas distancias relativas por construcción son pequeñas. En la Ecuación 4.1 se muestra el modelo obtenido, su representación gráfica se puede ver en la Figura 4.11 (líneas continuas), donde la línea celeste corresponde a la dirección vertical y la verde a la omnihorizontal. Este variograma es utilizado tanto para el kriging original como para sus modificaciones. En el caso de las modificaciones que utilizan direcciones y razones de anisotropía variables, las razones se eligen de modo que se mantengan los radios de búsqueda y alcances en el variograma.

$$\gamma(h) = 20,0 \cdot pepa + 1100,0 \cdot gaussiano_{300,300,120}(h) \quad (4.1)$$

El modelo variográfico (Ecuación 4.1 y Figura 4.11) al utilizar la estructura de un variograma gaussiano, tiene un crecimiento lento cerca del origen, y a medida que se aleja más el crecimiento es más fuerte. Esto indica que a distancias<sup>2</sup> pequeñas la variabilidad es pequeña y a medida que la distancia aumenta la variabilidad crece cada vez más rápido. Para la vertical el crecimiento es más rápido que para el plano, pues en la vertical la variabilidad máxima se alcanza a los 120[m] y para el plano a los 300[m].

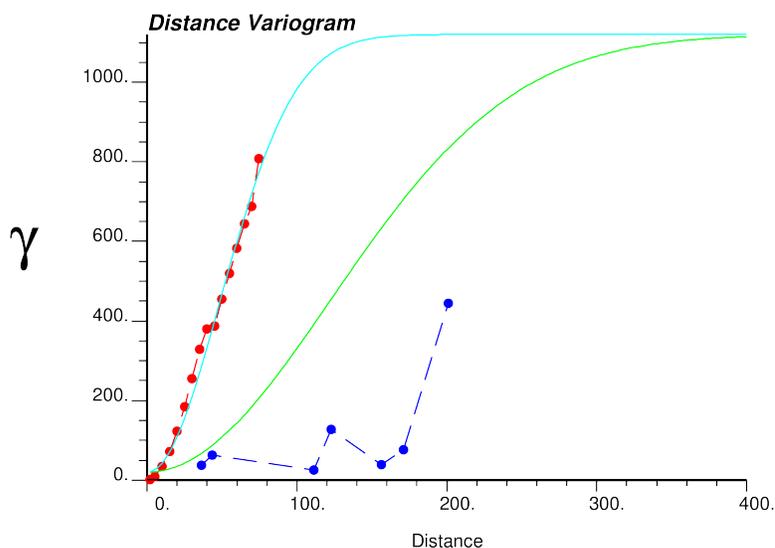


Figura 4.11: Variogramas para kriging.

---

<sup>2</sup>Es importante recalcar que en el variograma, la distancia corresponde a la distancia entre muestras y no a la distancia al contacto.

#### 4.4.2. Parámetros para la ejecución

En la Tabla 4.6 se detalla los parámetros de la grilla, la cual está compuesta por 450241 bloques. Los demás parámetros se muestran en la Tabla 4.7. La grilla definida para el campo de direcciones variables es idéntica a la grilla de estimación. Las dimensiones de esta grilla se eligieron pensando en las distancias entre las muestras y a la vez pensando en la cantidad de bloques del modelo, ergo se optó por un tamaño de bloque pequeño, pero que el modelo no sobrepase los 500000 bloques. Esta última restricción se impuso para poder utilizar la misma grilla en simulación, donde se deben efectuar una gran cantidad de realizaciones y los modelos con muchos bloques demoran demasiado tiempo.

	Nodos	Tamaño	Origen
<i>este</i>	121	5	72000
<i>norte</i>	61	5	83450
<i>elevacion</i>	61	5	2200

Tabla 4.6: Grilla para estimación, simulación y del campo de direcciones variables.

Parámetro	Valor
Radio búsqueda en plano	300
Radio búsqueda en vertical	120
N° mínimo de muestras	3
N° máximo de muestras	24
Restricción por octantes	No

Tabla 4.7: Parámetros para kriging.

Para el caso de orientaciones y anisotropías variables, el radio de búsqueda y el alcance del variograma se define mediante el radio o alcance en la dirección principal. En dichos casos los parámetros *ratio\_plane* y *ratio\_vertical* en el campo de direcciones corresponden a 1,0 y 0,4 respectivamente, de manera que concuerden con los radios y alcances del kriging original y utilizando pseudosondajes. Para el caso del kriging con pseudosondajes el variograma no se calcula nuevamente, pues es a través del condicionamiento que se modifica el kriging original y no a través de cambios en la función aleatoria.

## 4.5. Resultados de Kriging

A continuación se muestra los resultados para las cuatro versiones del kriging de distancias al contacto, éstas son: kriging sin modificaciones u original, kriging con pseudosondas, kriging con orientaciones variables y kriging con anisotropías variables. Para cada uno de los métodos se muestra una vista isométrica del cuerpo obtenido. Para obtener este cuerpo se etiqueta como “cuerpo mineralizado” a los bloques con distancia al contacto menor o igual que cero, y, como “roca caja”, a los mayores que cero. También se muestra el histograma de la variable distancia estimada al contacto y las estadísticas de ésta. Finalmente, se muestra la tabla de contingencia y las razones entre volumen modelado y real para los bloques del cuerpo y del exterior. Los indicadores son explicados en la Sección 1.4 de la “Introducción”.

### 4.5.1. Original

En la Figura 4.12 se muestra una vista isométrica del modelo y el histograma de la variable distancia estimada al contacto para el algoritmo de kriging de distancias original. En la Tabla 4.8 se muestran las estadísticas de la distancia estimada al contacto.

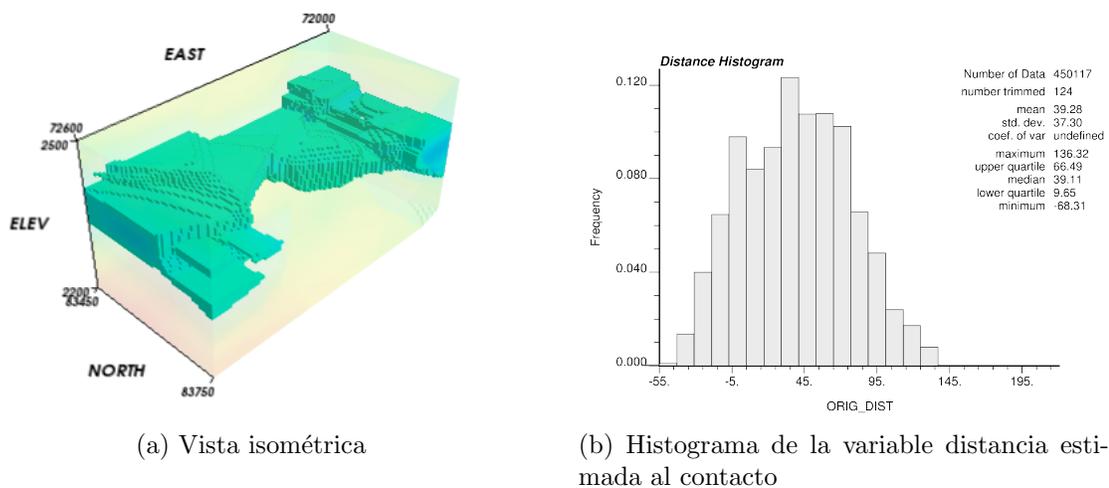


Figura 4.12: Vista del modelo e histograma de la distancia estimada al contacto para kriging original.

Mínimo	-68.31
Máximo	136.32
Media	39.28
Mediana	39.11
Desviación	37.30

Tabla 4.8: Estadísticas de la distancia estimada al contacto para kriging original.

En la Tabla 4.9 se muestran las tablas de contingencia del kriging de distancias original, para sondajes de validación y pozos de tronadura. En la Tabla 4.10 se pueden ver las razones entre volumen modelado y volumen real, para el cuerpo y para el exterior.

		Muestras					Muestras		
		Cuerpo	Exterior	Acierto			Cuerpo	Exterior	Acierto
Modelo	Cuerpo	1485	902	0.622	Modelo	Cuerpo	7300	4029	0.644
	Exterior	524	4687	0.899		Exterior	2238	25928	0.921
Cobertura		0.739	0.839	0.812	Cobertura		0.765	0.866	0.841

(a) Validación contra sondajes.

(b) Validación contra pozos de tronadura.

Tabla 4.9: Tablas de contingencia para validación de kriging de distancias original.

	Sondajes	Pozos
RVC	1.188	1.188
RVE	0.932	0.940

Tabla 4.10: Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias original.

#### 4.5.2. Con pseudosondajes: caso 1

Para este primer caso de uso de kriging de distancias con pseudosondajes se utiliza solamente un pseudosondaje, el cual indica la continuidad del cuerpo. Este sondaje se puede ver en la Figura 4.4.

En la Figura 4.15 se muestra una vista isométrica del modelo y el histograma de la variable distancia estimada al contacto para el algoritmo de kriging de distancias usando un

pseudosondaje sólo para indicar continuidad en el cuerpo. En la Tabla 4.14 se muestran las estadísticas de la distancia estimada al contacto.

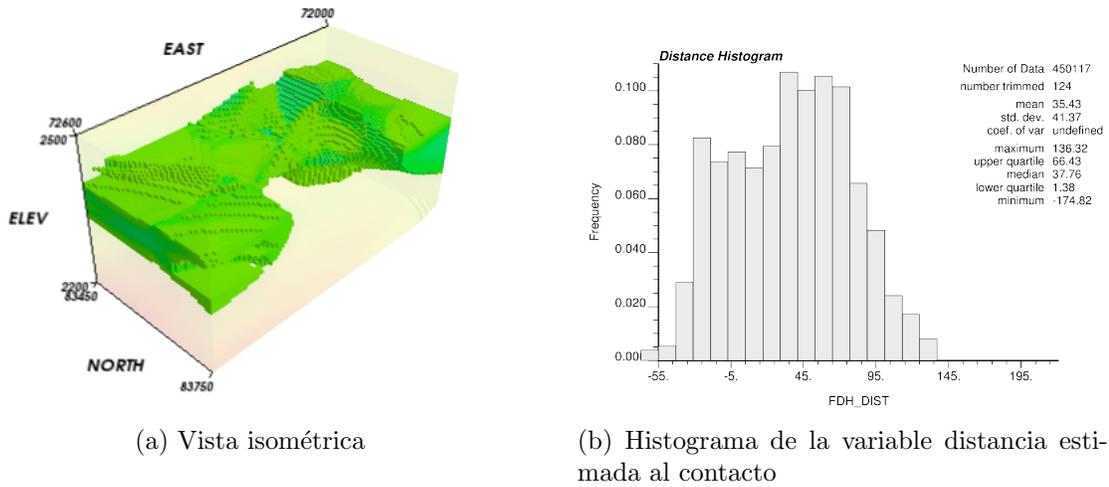


Figura 4.13: Vista del modelo e histograma de la distancia estimada al contacto para kriging con pseudosondajes (caso 1).

Mínimo	-174.82
Máximo	136.32
Media	35.43
Mediana	37.6
Desviación	41.37

Tabla 4.11: Estadísticas de la distancia estimada al contacto para kriging con pseudosondajes (caso 1).

En la Tabla 4.15 se muestran las tablas de contingencia del kriging de distancias con pseudosondajes, para sondajes no utilizados y pozos de tronadura. En la Tabla 4.16 se pueden ver las razones entre volumen modelado y volumen real, para el cuerpo y para el exterior.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	1742	1513	0.535
	Exterior	267	4076	0.939
	Cobertura	0.867	0.729	0.766

(a) Validación contra sondajes.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	8461	6905	0.551
	Exterior	1077	23052	0.955
	Cobertura	0.887	0.770	0.798

(b) Validación contra pozos de tronadura.

Tabla 4.12: Tablas de contingencia para validación de para kriging de distacias con pseudo-sondajes (caso 1).

	Sondajes	Pozos
RVC	1.620	1.611
RVE	0.777	0.805

Tabla 4.13: Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distacias con pseudosondajes (caso 1).

### 4.5.3. Con pseudosondajes: caso 2

Los resultados de la experiencia con sólo un pseudosondaje, muestran un aumento bastante grande del tamaño del sólido que describe el cuerpo mineralizado, llegando a un 160 % del tamaño real. Este crecimiento excesivo del volumen del cuerpo permite entender que si bien es importante indicar la continuidad del cuerpo a través de pseudosondajes, también es necesario indicar la continuidad de la roca caja que lo contiene. Para indicar al algoritmo que se debe limitar el crecimiento del cuerpo se crea dos nuevos pseudosondajes, uno por sobre el primero y uno por debajo (Figura 4.14).

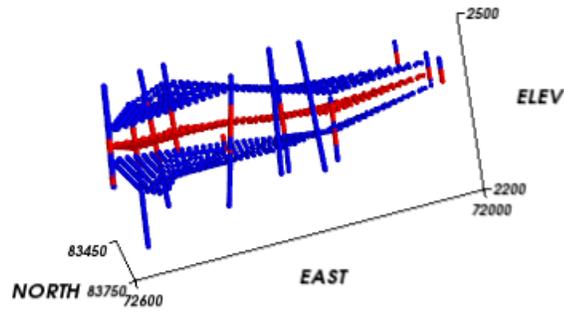
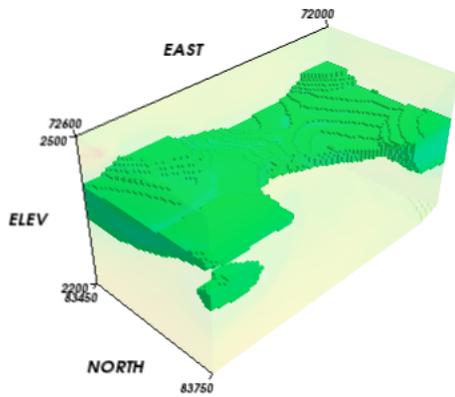
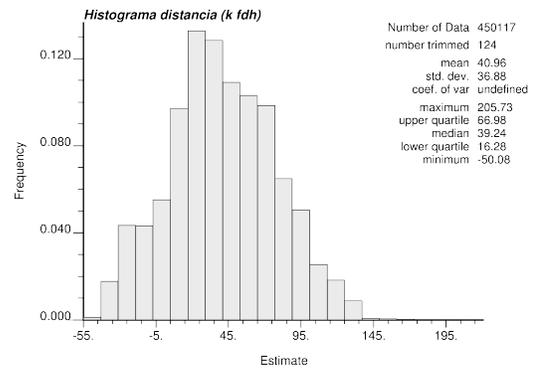


Figura 4.14: Vista de los nuevos pseudosondajes.

En la Figura 4.15 se muestra una vista isométrica del modelo y el histograma de la variable distancia estimada al contacto para el algoritmo de kriging de distancias usando pseudosondajes más restrictivos. En la Tabla 4.14 se muestran las estadísticas de la distancia estimada al contacto.



(a) Vista isométrica



(b) Histograma de la variable distancia estimada al contacto

Figura 4.15: Vista del modelo e histograma de la distancia estimada al contacto para kriging con pseudosondajes (caso 2).

Mínimo	-50.08
Máximo	205.73
Media	40.96
Mediana	39.24
Desviación	36.88

Tabla 4.14: Estadísticas de la distancia estimada al contacto para kriging con pseudosondajes (caso 2).

En la Tabla 4.15 se muestran las tablas de contingencia del kriging de distancias con pseudosondajes, para sondajes no utilizados y pozos de tronadura. En la Tabla 4.16 se pueden ver las razones entre volumen modelado y volumen real, para el cuerpo y para el exterior.

		Muestras					Muestras		
		Cuerpo	Exterior	Acierto			Cuerpo	Exterior	Acierto
Modelo	Cuerpo	1313	569	0.698	Modelo	Cuerpo	6251	2761	0.694
	Exterior	696	5020	0.878		Exterior	3287	27196	0.892
Cobertura		0.654	0.898	0.834	Cobertura		0.655	0.908	0.847

(a) Validación contra sondajes.

(b) Validación contra pozos de tronadura.

Tabla 4.15: Tabla de contingencia para validación de kriging de distacias con pseudosondajes (caso 1).

	Sondajes	Pozos
RVC	0.937	0.945
RVE	1.023	1.018

Tabla 4.16: Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distacias con pseudosondajes (caso 2).

#### 4.5.4. Con direcciones variables: Orientación de la búsqueda y variograma

En la Figura 4.16 se muestra una vista isométrica del modelo y el histograma de la variable distancia estimada al contacto para el algoritmo de kriging de distancias con

orientaciones variables. Para este caso el variograma es isotrópico y el alcance de éste corresponde al alcance en la dirección principal del variograma del caso original y son las variables *ratio\_plane* y *ratio\_vertical*, definidas localmente, las que determinan la anisotropía de la búsqueda y del variograma. En la Tabla 4.17 se muestran las estadísticas de la distancia estimada al contacto.

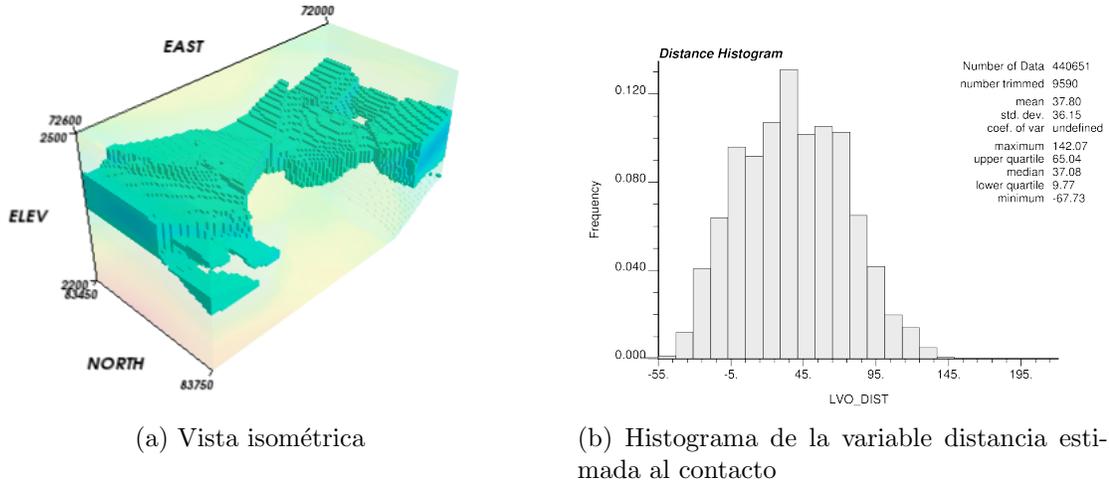


Figura 4.16: Vista del modelo e histograma de la distancia estimada al contacto para kriging con orientaciones variables.

Mínimo	-67.73
Máximo	142.07
Media	37.80
Mediana	37.08
Desviación	36.15

Tabla 4.17: Estadísticas de la distancia estimada al contacto para kriging con orientaciones variables.

En la Tabla 4.18 se muestran las tablas de contingencia del kriging de distancias con orientaciones variables, para sondeos no utilizados y pozos de tronadura. En la Tabla 4.19 se pueden ver las razones entre volumen modelado y volumen real, para el cuerpo y para el exterior.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	1443	766	0.653
	Exterior	566	4793	0.894
	Cobertura	0.718	0.862	0.824

(a) Validación contra sondajes.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	7057	3603	0.662
	Exterior	2481	26354	0.914
	Cobertura	0.740	0.880	0.846

(b) Validación contra pozos de tronadura.

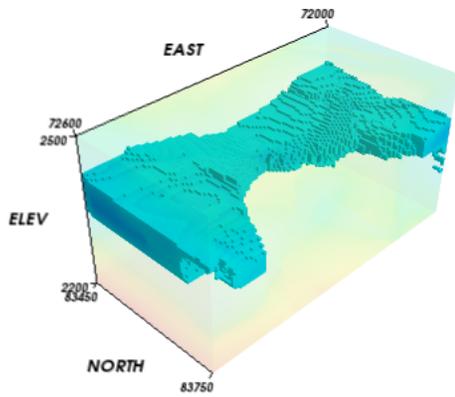
Tabla 4.18: Tabla de contingencia para validación de kriging de distancias con orientaciones variables.

	Sondajes	Pozos
RVC	1.100	1.118
RVE	0.964	0.963

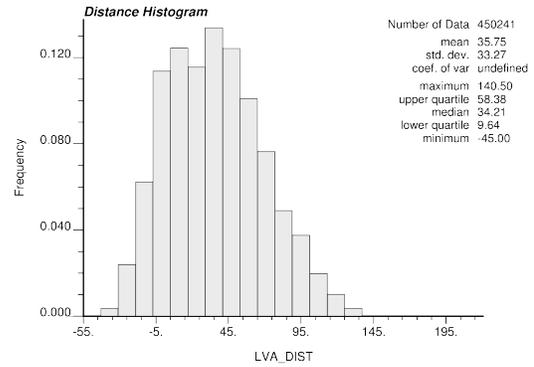
Tabla 4.19: Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con orientaciones variables.

#### 4.5.5. Con direcciones variables: Anisotropías locales

En la Figura 4.17 se muestra una vista isométrica del modelo y el histograma de la variable distancia estimada al contacto para el algoritmo de kriging de distancias con anisotropías variables. En este método las distancias se calculan considerando las direcciones variables, utilizando el camino entre dos puntos que respeta de mejor manera dichas direcciones. El variograma para este caso es isótropo y su alcance es el alcance en la dirección principal del variograma en el caso original y son las variables *ratio\_plane* y *ratio\_vertical*, definidas localmente, las que determinan la anisotropía de la búsqueda y del variograma. En la Tabla 4.20 se muestran las estadísticas de la distancia estimada al contacto.



(a) Vista isométrica



(b) Histograma de la variable distancia estimada al contacto

Figura 4.17: Vista del modelo e histograma de la distancia estimada al contacto para kriging con anisotropías variables.

Mínimo	-45.00
Máximo	140.50
Media	35.75
Mediana	34.21
Desviación	33.27

Tabla 4.20: Estadísticas de la distancia estimada al contacto para kriging con anisotropías variables.

En la Tabla 4.21 se muestran las tablas de contingencia del kriging de distancias con anisotropías variables, para sondajes no utilizados y pozos de tronadura. En la Tabla 4.22 se pueden ver las razones entre volumen modelado y volumen real, para el cuerpo y para el exterior.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	1494	744	0.668
	Exterior	515	4849	0.904
	Cobertura	0.744	0.867	0.834

(a) Validación contra sondajes.

		Muestras		
		Cuerpo	Exterior	Acierto
Modelo	Cuerpo	6991	3483	0.667
	Exterior	2547	26474	0.912
	Cobertura	0.733	0.884	0.847

(b) Validación contra pozos de tronadura.

Tabla 4.21: Tabla de contingencia para validación de kriging de distancias con anisotropías variables.

	Sondajes	Pozos
RVC	1.114	1.098
RVE	0.959	0.969

Tabla 4.22: Razón entre volumen modelado y volumen real para el cuerpo (RVC) y para el exterior (RVE), para kriging de distancias con anisotropías variables.

#### 4.5.6. Discusión de los resultados

Uno de los primeros resultados, mostrados en la Tabla 4.23, que saltan a la vista son los indicadores que se obtienen para el primer kriging con pseudosondajes (sdh 1), donde se ve que todos éstos empeoran excepto recall positivo y la precisión negativa. El valor del recall positivo indica que un mayor porcentaje del cuerpo real es reportado. Sin embargo, también se puede ver que esta mayor cobertura se debe a la sobrestimación del tamaño del cuerpo, pues el acierto disminuye y, la razón entre el volumen del cuerpo real y el modelo indica una sobrestimación evidente. En otras palabras, el aumento del cuerpo no es selectivo sino que sólo crece. Por otro lado, el caso del segundo kriging con pseudosondajes (sdh 2), va en la otra dirección, pues se está subestimando el tamaño del cuerpo. Si bien no es tan marcada se puede ver que la disminución del cuerpo tampoco es selectiva, pues el acierto aumenta, pero la cobertura baja al tramo 0.6 - 0.7. Es pertinente mencionar que la validación se realiza en base a la pertenencia o no al cuerpo en función de la distancia estimada al contacto<sup>3</sup>, y no directamente sobre esta última.

Para el caso de los kriging con direcciones variables se puede ver que ambos mejoran levemente la capacidad de predicción, sin embargo, no impactan en gran medida la cobertura, lo cual es bastante bueno. En estos métodos se sobrestima levemente el tamaño del cuerpo real, sin embargo, se sobrestima en menor medida al compararlo con el kriging original.

Un detalle a considerar es la media de la distancia estimada al contacto en todos los

<sup>3</sup>Pertenece al cuerpo si la distancia estimada al contacto es menor o igual a cero

kriging, la cual varía entre 35 y 41, que concuerda con las estadísticas de los datos de sondeos disponibles, la cual es aproximadamente 40. Sin embargo, el conjunto de datos de trabajo subestima esta media, pues el valor en este conjunto es de 25. Esta diferencia se explica por el agrupamiento de los datos y también por el sesgo espacial de éstos. En la Tabla 4.24 se puede ver las estadísticas de los datos desagrupados, donde la media aumenta respecto a la media del conjunto de trabajo. En las Figuras 4.8 y 4.9, se puede ver un ejemplo del sesgo espacial, ya que sólo uno de los sondeos llega a la parte inferior de la zona de estudio (2200[m]) donde se concentran los valores más altos de la distancia, generándose un sesgo hacia los valores más pequeños.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	orig	0.622	0.899	0.739	0.839	0.812	1.188	0.932
	sdh 1	0.535	0.939	0.867	0.729	0.766	1.620	0.777
	sdh 2	0.698	0.878	0.654	0.898	0.834	0.937	1.023
	lvo	0.653	0.894	0.718	0.862	0.824	1.100	0.964
	lva	0.668	0.904	0.744	0.867	0.834	1.114	0.959
Pozos	orig	0.644	0.921	0.765	0.866	0.841	1.188	0.940
	sdh 1	0.551	0.955	0.887	0.770	0.798	1.611	0.805
	sdh 2	0.694	0.892	0.655	0.908	0.847	0.945	1.018
	lvo	0.662	0.914	0.740	0.880	0.846	1.118	0.963
	lva	0.667	0.912	0.733	0.884	0.847	1.098	0.969

Tabla 4.23: Resumen de indicadores obtenidos para los distintos kriging.

## 4.6. Simulación

### 4.6.1. Desagrupamiento y Transformación a Gaussiana

Como se explicó en la Sección 2.1.3 del Capítulo “Antecedentes”, para realizar la simulación secuencial gaussiana se utiliza un modelo multi-gaussiano, el cual, a partir de la matriz de covarianza caracteriza la distribución multivariable de la función aleatoria. Para poder utilizar este modelo, la variable que se desea simular debe distribuirse como una gaussiana estándar, por lo cual se debe transformar desde la distribución original utilizando la función de anamorfosis gaussiana<sup>4</sup>. Un paso previo a esta transformación es desagrupar los datos, es decir, corregir problemas en el muestreo, como zonas con muchas muestra y otras con muy pocas. Este desagrupamiento permite tener estadísticas más representativas, y en consecuencia, una mejor función de anamorfosis.

<sup>4</sup>En el Apéndice A se muestra la validación de la hipótesis multigaussiana

Para desagrupar los datos de sondeos se utiliza la técnica de los polígonos de influencia. Se discretiza la zona de estudio en bloques de 1[m] en cada dirección. Para cada bloque, representado por su centro, se calcula la muestra más cercana a éste y, finalmente, para cada una de las muestras se cuenta la cantidad de veces que fue elegida como la muestra más cercana. Este valor normalizado al intervalo  $[0, 1]$ , corresponde al peso de la muestra. El procedimiento anterior se realizó para los datos reales y para los datos reales + datos de pseudosonajes. En la Figura 4.18 se muestran los histogramas desagrupados para la variable distancia al contacto para los dos conjuntos mencionados y; en la Tabla 4.24, las estadísticas de la misma variable sobre los datos desagrupados para los mismos conjuntos. La anamorfosis gaussiana se aplica también sobre los mismos conjuntos (datos reales y datos reales + pseudosonajes). La nueva variable, que corresponde a la transformación de la distribución original a una gaussiana se le conoce como “normal score”, así a la variable que se utiliza para la simulación en este trabajo se le denominará “normal score de la distancia al contacto”; y a la variable simulada, “normal score simulado de la distancia al contacto”.

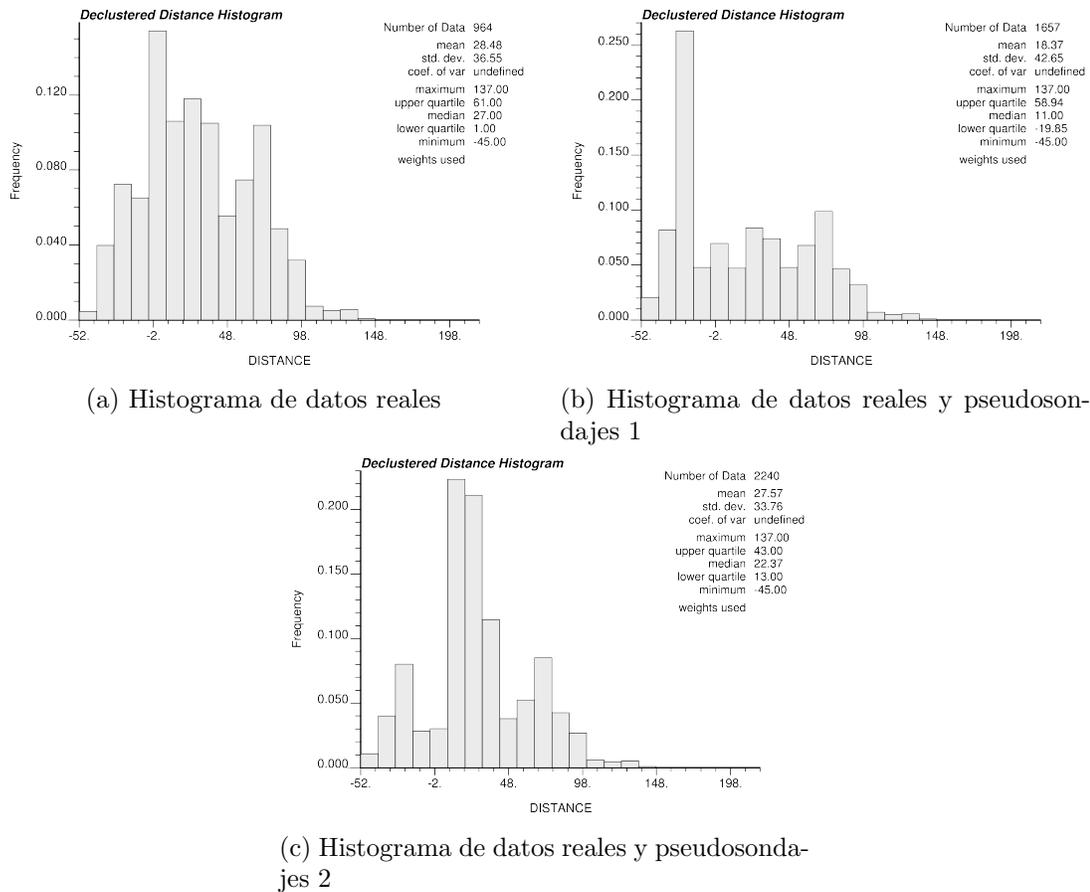


Figura 4.18: Histograma de la variable distancia al contacto para datos desagrupados.

	Datos reales	Datos reales + Pseudosondajes 1	Datos reales + Pseudosondajes 2
Mínimo	-45.00	-45.00	-45.00
Máximo	137.00	137.00	137.00
Media	28.48	18.37	27.57
Desviación	36.55	42.65	33.76

Tabla 4.24: Estadísticas de la variable distancia para los datos desagrupados.

#### 4.6.2. Variografía

Para la simulación se debe calcular un nuevo variograma, pues esta vez se trabaja sobre la variable distancia transformada. Este variograma se diferencia de aquel usado por kriging en que esta vez el variograma alcanza una meseta en 1, pues la variable corresponde a una gaussiana con media igual a 0 y varianza igual a 1, además la transformación que se aplica a la variable considera pesos de desagrupamiento, lo cual introduce, indirectamente, cambios en la forma del variograma.

En el caso de simulación también se utiliza la dirección omnihorizontal y la vertical como direcciones de anisotropía. En la gráfica de la Figura 4.19 (línea discontinua) se muestra los variogramas en el plano y en la vertical, la línea roja corresponde a la vertical y la azul al plano.

	Omnihorizontal	Vertical
N° Lags	15	15
Distancia Lag	25.0	5.0
Tolerancia Lag	12.5	2.5
Azimuth	90.0°	0.0°
Tolerancia Azimuth	90.0°	22.5°
Bandwidth Horizontal	10.0°	10.0°
Dip	0.0°	90.0°
Tolerancia Dip	22.5°	22.5°
Bandwidth Vertical	10.0°	10.0°

Tabla 4.25: Parámetros del variograma experimental para simulación.

En la Ecuación 4.2 se muestra el modelo obtenido, su representación gráfica se puede

ver en la Figura 4.19 (líneas continuas), donde la línea celeste corresponde la dirección vertical y la verde a la omnihorizontal. El procedimiento fue idéntico al que se hace para obtener el variograma usado en kriging, a saber: se modela la dirección vertical, la cual se ve bastante regular y, para la dirección omnihorizontal se utiliza la misma estructura con un alcance mayor, de manera que calce con el variograma experimental, al menos en algunos puntos.

$$\gamma = 0,1 \cdot pepa + 0,9 * gaussiano_{270,270,100}(h) \quad (4.2)$$

En el modelo variográfico de la Ecuación 4.2 se ve una tendencia similar a la del variograma utilizado en kriging: la variabilidad presenta un crecimiento lento que con la distancia<sup>5</sup> se hace más rápido, es decir, para distancias pequeñas se mantiene una variabilidad pequeña, mas cuando las distancias aumenta la correlación se pierde cada vez más rápido. En el nuevo variograma, confeccionado para la simulación, el efecto pepa se tiene que imponer a pesar de el mal calce con el variograma experimental, pues utilizar valores muy cercanos a cero genera problemas prácticos para calcular el valor estimado y la varianza de kriging.

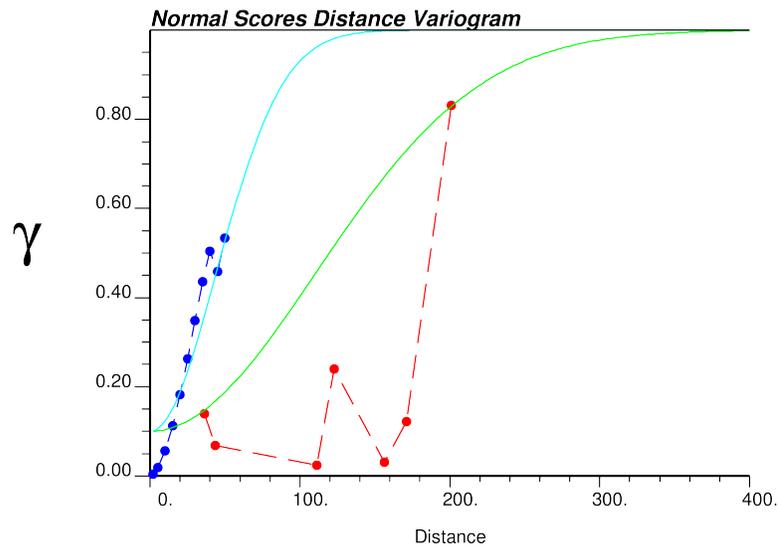


Figura 4.19: Variogramas para simulación.

### 4.6.3. Parámetros para la ejecución

En la Tabla 4.6 se detalla los parámetros de la grilla. La grilla de simulación, como ya se mencionó al describir el kriging, es idéntica a la grilla de estimación. En la Tabla 4.26 se muestran los parámetros utilizados para ejecutar la simulación.

<sup>5</sup>Recordar que esta distancia hace referencia a la distancia entre muestras y no a la distancia al contacto.

Parámetro	Valor
Radio búsqueda en plano	270
Radio búsqueda en vertical	100
N° mínimo de muestras	3
N° máximo de muestras	24
Nodos simulados a utilizar	24
Restricción por octantes	No
Tipo de kriging utilizado	Simple
Media para kriging simple	0.0
Realizaciones	200

Tabla 4.26: Parámetros para simulación.

La razón de anisotropía para simulación con orientaciones variables y anisotropías variables es 0,37, la cual corresponde a la razón entre los alcances en las direcciones del variograma utilizado en los otros métodos. Al igual que para kriging, la simulación con pseudosondajes utiliza el mismo variograma que la simulación original, pues se quiere modificar el comportamiento de la simulación por medio del condicionamiento y no por cambios en la variografía.

## 4.7. Resultados de Simulación

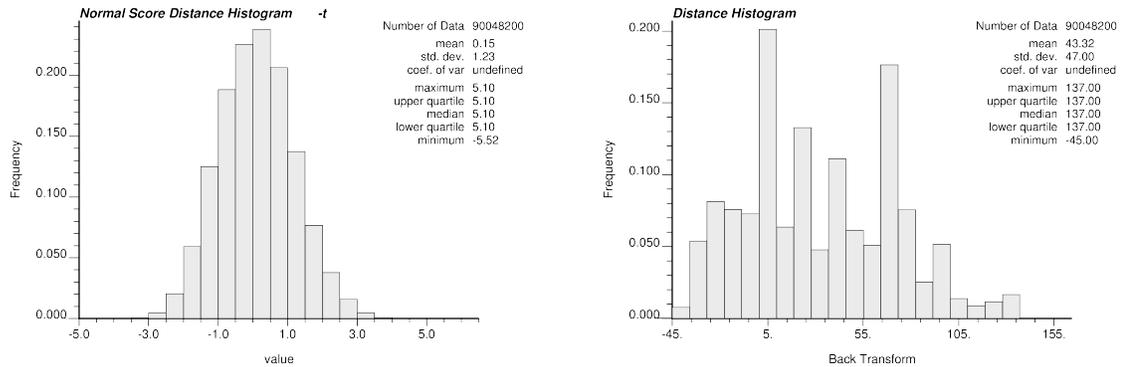
En esta sección se muestran los resultados para las distintas realizaciones de cada una de las versiones de la simulación secuencial gaussiana. A diferencia de kriging, no se utilizó la tabla de contingencia debido a que se debe informar el promedio, desviación y los valores extremos, lo cual complicaría de sobremanera la lectura y entendimiento de dicha tabla. En su lugar se muestra una tabla donde aparecen los valores medios, extremos y la desviación para cada indicador. También se muestran las estadísticas e histogramas para la variable normal score simulado de la distancia al contacto y para la variable distancia simulada al contacto.

### 4.7.1. Original

En la Tabla 4.27 se muestran las estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto. En la Figura 4.20 se muestran los histogramas para dichas variables.

	Normal score simulado de la distancia al contacto	Distancia simulada al contacto
Mínimo	-5.52	-45.00
Máximo	5.10	-137.00
Media	0.15	43.32
Desviación	1.23	47.00

Tabla 4.27: Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana original.



(a) Histograma de la variable normal score simulado de la distancia al contacto

(b) Histograma de la variable distancia simulada al contacto

Figura 4.20: Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana original.

En la Tabla 4.28 se muestra el valor mínimo, medio y máximo de los indicadores calculados sobre el conjunto de todas las realizaciones, también se muestra su desviación estándar.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	Mínimo	0.427	0.854	0.500	0.674	0.696	0.712	0.747
	Media	0.527	0.895	0.707	0.795	0.774	1.351	0.888
	Máximo	0.703	0.934	0.841	0.933	0.828	1.795	1.092
	Desviación	0.040	0.015	0.055	0.040	0.024	0.166	0.053
Pozos	Mínimo	0.423	0.865	0.547	0.654	0.688	0.782	0.719
	Media	0.559	0.901	0.718	0.816	0.793	1.295	0.906
	Máximo	0.700	0.936	0.835	0.925	0.838	1.882	1.069
	Desviación	0.042	0.014	0.050	0.038	0.024	0.155	0.049

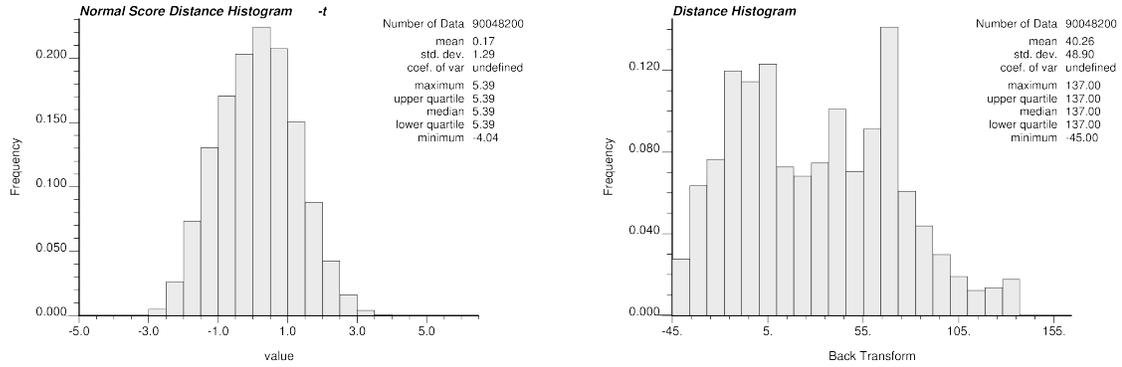
Tabla 4.28: Indicadores para simulación secuencial gaussiana original.

#### 4.7.2. Con pseudosondajes: caso 1

En la Tabla 4.31 se muestran las estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto. En la Figura 4.22 se muestran los histogramas para dichas variables.

	Normal score simulado de la distancia al contacto	Distancia simulada al contacto
Mínimo	-4.04	-45.00
Máximo	5.39	-137.00
Media	0.17	40.26
Desviación	1.29	48.90

Tabla 4.29: Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 1).



(a) Histograma de la variable normal score simulado de la distancia al contacto

(b) Histograma de la variable distancia simulada al contacto

Figura 4.21: Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 1).

En la Tabla 4.32 se muestra el valor mínimo, medio y máximo de los indicadores calculados sobre el conjunto de todas las realizaciones, también se muestra su desviación estándar.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	Mínimo	0.382	0.911	0.798	0.569	0.635	1.513	0.597
	Media	0.480	0.941	0.862	0.700	0.739	1.804	0.744
	Máximo	0.546	0.961	0.913	0.779	0.793	2.266	0.837
	Desviación	0.028	0.009	0.023	0.036	0.026	0.123	0.039
Pozos	Mínimo	0.371	0.909	0.818	0.554	0.620	1.468	0.609
	Media	0.500	0.946	0.870	0.720	0.757	1.749	0.762
	Máximo	0.565	0.959	0.910	0.797	0.805	2.227	0.851
	Desviación	0.034	0.008	0.019	0.042	0.031	0.137	0.044

Tabla 4.30: Indicadores para simulación secuencial gaussiana con pseudosondajes (caso 1).

### 4.7.3. Con pseudosondajes: caso 2

Al igual que en kriging, las simulaciones mostraron un crecimiento excesivo del volumen del cuerpo, por lo cual se agrega los dos nuevos pseudosondajes, uno por sobre y otro por debajo del cuerpo, para limitar su crecimiento.

En la Tabla 4.31 se muestran las estadísticas de la variable normal score simulado de

la distancia al contacto y de la variable distancia simulada al contacto. En la Figura 4.22 se muestran los histogramas para dichas variables.

	Normal score simulado de la distancia al contacto	Distancia simulada al contacto
Mínimo	-4.12	-45.00
Máximo	5.41	-137.00
Media	0.33	44.36
Desviación	1.16	45.73

Tabla 4.31: Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 2).

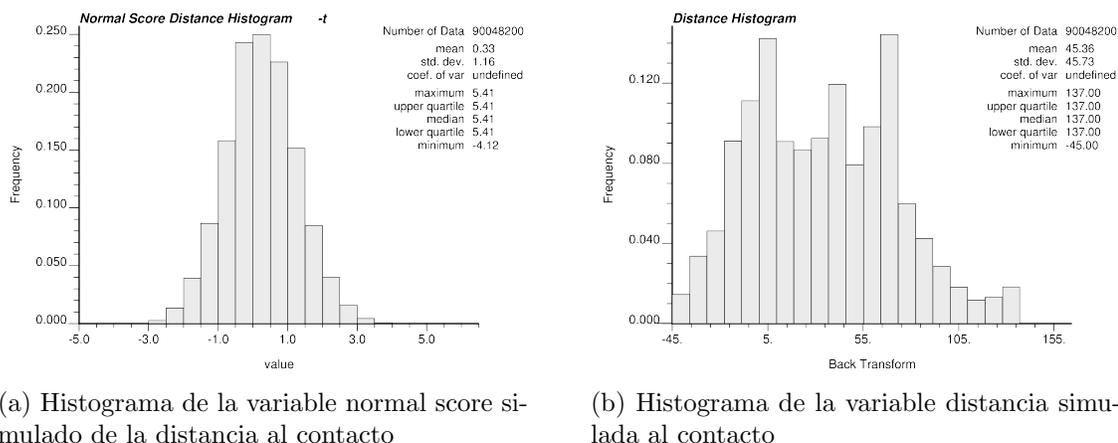


Figura 4.22: Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con pseudosondajes (caso 2).

En la Tabla 4.32 se muestra el valor mínimo, medio y máximo de los indicadores calculados sobre el conjunto de todas las realizaciones, también se muestra su desviación estándar.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	Mínimo	0.431	0.879	0.798	0.569	0.635	1.513	0.597
	Media	0.551	0.903	0.727	0.809	0.789	1.326	0.896
	Máximo	0.638	0.922	0.786	0.876	0.836	1.710	0.981
	Desviación	0.037	0.007	0.023	0.030	0.022	0.105	0.033
Pozos	Mínimo	0.451	0.879	0.662	0.723	0.722	1.046	0.807
	Media	0.561	0.903	0.725	0.817	0.795	1.300	0.904
	Máximo	0.654	0.917	0.776	0.883	0.838	1.606	0.985
	Desviación	0.044	0.006	0.020	0.035	0.025	0.118	0.037

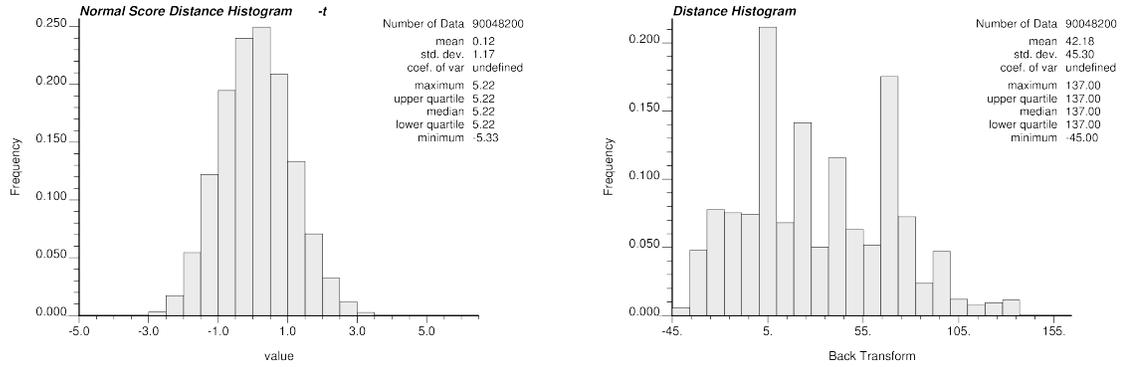
Tabla 4.32: Indicadores para simulación secuencial gaussiana con pseudosondajes (caso 2).

#### 4.7.4. Con direcciones variables: Orientación de la búsqueda y variograma

En la Tabla 4.33 se muestran las estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto. En la Figura 4.23 se muestran los histogramas para dichas variables.

	Normal score simulado de la distancia al contacto	Distancia simulada al contacto
Mínimo	-5.33	-45.00
Máximo	5.22	-137.00
Media	0.12	42.18
Desviación	1.17	45.30

Tabla 4.33: Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con orientación variable.



(a) Histograma de la variable normal score simulado de la distancia al contacto

(b) Histograma de la variable distancia simulada al contacto

Figura 4.23: Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con orientación variable.

En la Tabla 4.34 se muestra el valor mínimo, medio y máximo de los indicadores calculados sobre el conjunto de todas las realizaciones, también se muestra su desviación estándar.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	Mínimo	0.416	0.849	0.476	0.657	0.683	0.666	0.731
	Media	0.546	0.891	0.686	0.815	0.784	1.268	0.915
	Máximo	0.714	0.931	0.826	0.939	0.842	1.846	1.106
	Desviación	0.044	0.016	0.059	0.040	0.024	0.171	0.054
Pozos	Mínimo	0.430	0.857	0.513	0.672	0.698	0.722	0.743
	Media	0.565	0.895	0.693	0.826	0.794	1.238	0.924
	Máximo	0.710	0.932	0.821	0.933	0.844	1.807	1.088
	Desviación	0.048	0.015	0.055	0.041	0.027	0.163	0.052

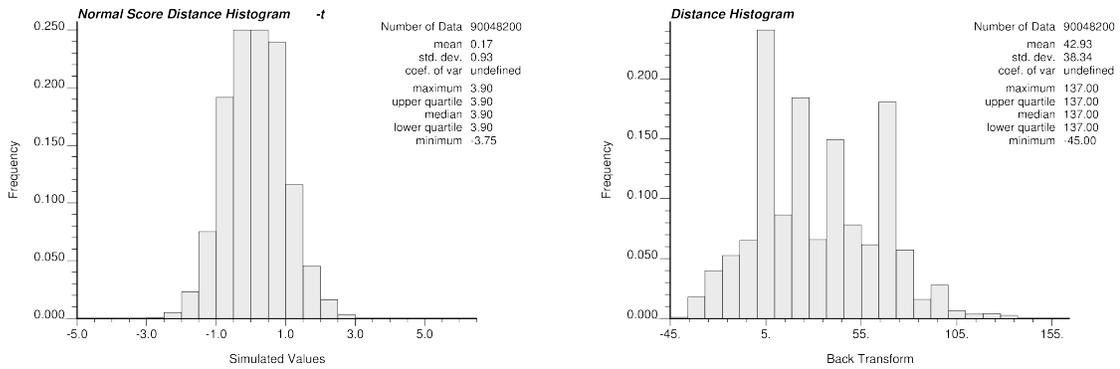
Tabla 4.34: Indicadores para simulación secuencial gaussiana con orientación variable.

#### 4.7.5. Con direcciones variables: Anisotropías locales

En la Tabla 4.35 se muestran las estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto. En la Figura 4.24 se muestran los histogramas para dichas variables.

	Normal score simulado de la distancia al contacto	Distancia simulada al contacto
Mínimo	-3.75	-45.00
Máximo	3.90	-137.00
Media	0.17	42.93
Desviación	0.93	38.34

Tabla 4.35: Estadísticas de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con anisotropías variables.



(a) Histograma de la variable normal score simulado de la distancia al contacto

(b) Histograma de la variable distancia simulada al contacto

Figura 4.24: Histograma de la variable normal score simulado de la distancia al contacto y de la variable distancia simulada al contacto para simulación secuencial gaussiana con anisotropías variables.

En la Tabla 4.36 se muestra el valor mínimo, medio y máximo de los indicadores calculados sobre el conjunto de todas las realizaciones, también se muestra su desviación estándar.

		PP	PN	RP	RN	AT	RVC	RVE
Sondajes	Mínimo	0.560	0.855	0.517	0.836	0.788	0.779	0.943
	Media	0.639	0.873	0.591	0.893	0.820	0.927	1.023
	Máximo	0.710	0.894	0.679	0.926	0.841	1.179	1.070
	Desviación	0.027	0.007	0.029	0.015	0.009	0.069	0.022
Pozos	Mínimo	0.572	0.858	0.529	0.844	0.795	0.777	0.952
	Media	0.642	0.871	0.581	0.896	0.820	0.906	1.030
	Máximo	0.704	0.886	0.661	0.927	0.835	1.152	1.071
	Desviación	0.021	0.005	0.023	0.013	0.006	0.059	0.019

Tabla 4.36: Indicadores para simulación secuencial gaussiana con anisotropías variables.

#### 4.7.6. Discusión de los resultados

Los resultados de la Tabla 4.37 muestran que la simulación tiene, en general, una precisión menor para reportar el cuerpo, manteniendo relativamente constante la precisión para el exterior. Al mirar la razón entre volumen del cuerpo informado y el real se puede ver que existe una sobrestimación para todos los métodos, exceptuando el método de anisotropías variables (Iva). Esta sobrestimación en las versiones de la simulación secuencial gaussiana original, con pseudosondajes y con orientaciones variables no genera mejoras para sus versiones análogas de kriging, por lo cual se puede decir sin lugar a dudas que los resultados de la simulación son peores que los de kriging. No obstante la disminución del porcentaje de acierto respecto a kriging, los resultados de simulación permiten cuantificar la incertidumbre de la variable aleatoria. Para el caso de la simulación con anisotropías variables (Iva) el indicador que baja considerablemente es la recall positivo, lo cual indica que se pueden estar perdiendo bloques con mineral. Sin embargo, se puede ver que, al menos para el acierto total, sigue siendo competitivo respecto al kriging original a pesar de subestimar el cuerpo.

En términos prácticos, a partir de la comparación de los resultados obtenidos para kriging y para la simulación secuencial gaussiana, el tema de la estacionaridad no es un problema mayor. Los resultados obtenidos en la simulación secuencial gaussiana, en comparación con kriging muestran una disminución en el acierto total, sin embargo, la diferencia entre el acierto de la simulación y la de su kriging respectivo varía, en el peor de los casos, menos de un 6%. Esta cota se aplica inclusive para los algoritmos que utilizan pseudosondajes, donde la diferencia entre el volumen estimado del cuerpo en simulación y en kriging es de un 40%, es decir, donde la diferencia entre los modelos es mayor. Otro elemento importante en esta discusión es que, a diferencia de otros tipos de simulación y estimación, como por ejemplo kriging y la simulación secuencial de indicadores, en los algoritmos estudiados no se generan artefactos. Ambos factores muestran que los efectos negativos de la no estacionaridad en la simulación se mantienen acotados, bastante bajos y sin artefactos, es decir, el efecto es casi

imperceptible.

			PP	PN	RP	RN	AT	RVC	RVE
Sondajes	orig	kriging	0.622	0.899	0.739	0.839	0.812	1.188	0.932
		simulación (media)	0.527	0.895	0.707	0.795	0.774	1.351	0.888
	sdh 1	kriging	0.535	0.939	0.867	0.729	0.766	1.620	0.777
		simulación (media)	0.480	0.941	0.862	0.700	0.739	1.804	0.744
	sdh 2	kriging	0.698	0.878	0.654	0.898	0.834	0.937	1.023
		simulación (media)	0.551	0.903	0.727	0.809	0.789	1.326	0.896
	lvo	kriging	0.653	0.894	0.718	0.862	0.824	1.100	0.964
		simulación (media)	0.546	0.891	0.686	0.815	0.784	1.268	0.915
	lva	kriging	0.668	0.904	0.744	0.867	0.834	1.114	0.959
		simulación (media)	0.639	0.873	0.591	0.893	0.820	0.927	1.023
Pozos	orig	kriging	0.644	0.921	0.765	0.866	0.841	1.188	0.940
		simulación (media)	0.559	0.901	0.718	0.816	0.793	1.295	0.906
	sdh 1	kriging	0.551	0.955	0.887	0.770	0.798	1.611	0.805
		simulación (media)	0.500	0.946	0.870	0.720	0.757	1.749	0.762
	sdh 2	kriging	0.694	0.892	0.655	0.908	0.847	0.945	1.018
		simulación (media)	0.561	0.903	0.725	0.817	0.795	1.300	0.904
	lvo	kriging	0.662	0.914	0.740	0.880	0.846	1.118	0.963
		simulación (media)	0.565	0.895	0.693	0.826	0.794	1.238	0.924
	lva	kriging	0.667	0.912	0.733	0.884	0.847	1.098	0.969
		simulación (media)	0.642	0.871	0.581	0.896	0.820	0.906	1.030

Tabla 4.37: Resumen de resultados para las distintas simulaciones con junto al kriging respectivo.

## 4.8. Control de la rugosidad

En el Apéndice B se muestra el estudio que se hizo respecto a la rugosidad del contacto en función del variograma, específicamente entre la estructura utilizada para modelar el variograma y el efecto pepa. En dicho estudio se muestra que los mejores resultados para el acierto total se dan al usar un variograma gaussiano con el menor efecto pepa. Estos

resultados dejan como conclusión que no se necesita de un algoritmo para sugerir que la mejor combinación, pues la tendencia indica que a menor rugosidad, mayor acierto.

# Capítulo 5

## Conclusiones

El trabajo desarrollado en esta tesis consistió en la modificación del algoritmo de kriging y el de simulación secuencial gaussiana para incorporar información blanda en el modelamiento de cuerpos geológicos. La primera modificación corresponde a la incorporación de pseudosondajes (sdh, por soft-drillhole), que corresponde a información experta (blanda) incorporada por el usuario y que indica tendencias que debieran seguir los cuerpos. La segunda modificación consiste en la incorporación de un campo de direcciones, el cual indica las direcciones de continuidad locales en la zona de estudio. Estas direcciones se utilizan para orientar la búsqueda de muestras y, posteriormente para orientar el variograma (lvo, por *locally varying orientation*). La tercera y última modificación, la cual no fue desarrollada en este trabajo, pero fue integrada a la comparación, corresponde al uso de direcciones variables para el cálculo de las distancias, respetando a nivel más local las direcciones de continuidad (lva, por *locally varying anisotropy*). Estos tres métodos se compararon contra un caso base, que corresponde a los algoritmos sin modificaciones. También se realizó un estudio de la rugosidad del contacto en función de parámetros de variograma.

El aporte de este trabajo va en la línea de crear nuevas formas en que los usuarios expertos en modelamiento geológico, puedan interactuar de manera más simple con los datos. Este trabajo en particular busca crear los mecanismos para incorporar información blanda en los algoritmos de estimación y simulación geoestadística tradicionales. El objetivo es comprobar que las modificaciones propuestas a los algoritmos de kriging ordinario y al algoritmo de simulación secuencial gaussiana, generan modelos consistentes y que se ajustan a las tendencias indicadas por los sondajes artificiales y las direcciones variables. Estos algoritmos están pensados para soportar un *front-end*, donde el usuario, mediante dibujos, pantallas táctiles u otros medios similares, pueda indicar las tendencias que su experiencia le sugiere y que transforme estas tendencias en datos de pseudosondajes o en un campo de direcciones variables.

Los algoritmos descritos fueron sometidos a prueba en un caso de estudio, utilizando datos reales de un yacimiento. El yacimiento utilizado fue elegido de entre cuatro posibles, debido a que éste posee una geometría más compleja que los otros candidatos y por lo tanto más adecuados para el experimento. En esta prueba se comparó ciertos indicadores que permiten evaluar desempeño a la hora de predecir la extensión y posición de un cuerpo mineralizado

a partir de información limitada. Los indicadores utilizados fueron la exactitud (*precision*), la cobertura (*recall*) y las razones volumen real - volumen del modelo. La precisión indica cuántos de los nodos pertenecientes al modelo fueron correctamente predichos, para el caso del cuerpo y del exterior. También se calculó el acierto total, es decir, cuántos nodos del total fueron predichos correctamente. La cobertura indica cuántos de los nodos correspondientes al cuerpo fueron predichos como tal, nuevamente para el caso del cuerpo como para el caso del exterior. Las razones entre los volúmenes corresponden al volumen del cuerpo estimado dividido por el volumen real, para el cuerpo y el exterior, sin importar el acierto en la posición de los nodos. Las conclusiones se hicieron en base a los datos de validación correspondiente a muestras de sondajes, las cuales, si bien son menores en cantidad respecto a las muestras de pozos de tronadura, están distribuidos de manera más uniforme, pues los datos de pozos de tronadura están localizados sólo en la zona del yacimiento que ya ha sido explotada. El yacimiento en estudio posee plegamientos, sin embargo, estos no son tan pronunciados como se quisiera y no sacarán todo el provecho que se espera a estos algoritmos.

Para el caso de kriging, los algoritmos estudiados tuvieron mejoras de menos del 3% para el acierto total, lo cual es bastante bajo, sin embargo, este comportamiento es esperable, pues kriging suaviza los modelos. No obstante, cabe destacar que kriging con lva mostró mejoras en el acierto total sin empeorar mayormente otros indicadores, a diferencia de kriging con sdh, donde se observa una baja considerable de la cobertura.

En la simulación los resultados difieren un poco respecto a los obtenidos con kriging. En este caso los tres algoritmos tuvieron un comportamiento adecuado al problema. La simulación con sdh y con lvo se comportaron bastante parecido, con mejoras de 2% y 1% en el acierto total, respectivamente. En el caso de lvo se puede ver también una disminución de un 2% en la cobertura. Para la simulación con lva el acierto total sube en casi un 5%, que corresponde a la mejora más importante en este trabajo, sin embargo, la cobertura baja en un poco más del 10%. Este comportamiento, que suba la precisión y baje la cobertura, o al revés, es bastante común. Para simulación con lva, este *trade-off* se inclinó en favor de la precisión, lo cual en palabras simples significa que el algoritmo reporta menos bloques, pero con una mayor seguridad en su respuesta. Luego, este algoritmo tiene una utilidad mayor en escenarios donde la economía está contraída y la inversión es menor.

En los experimentos se pudo ver que cuando se utilizó un pseudosondaje indicando la continuidad del cuerpo, la sobrestimación fue mucho mayor que para el caso base. El segundo experimento, donde adicionalmente se limitó el crecimiento del cuerpo con sondajes en el exterior, el algoritmo subestimó el tamaño del cuerpo, sin embargo, dado que el caso base ya sobrestimaba, el algoritmo mejoró para el indicador de volumen del cuerpo. Luego, el efecto del uso de pseudosondajes tanto en kriging como en la simulación es que puede introducir un sesgo en las estadísticas y, por lo tanto, una distorsión en los resultados. Este comportamiento, si bien en principio puede ser visto como una debilidad, también tiene sus ventajas, pues, como ocurrió en el segundo experimento, puede corregir la sobrestimación o subestimación de los volúmenes buscados.

Los resultados obtenidos al estudiar cómo afectaban el efecto pepa y las estructuras utilizadas para modelar el variograma a la rugosidad de las superficies mostraron que la combinación con mejores indicadores corresponde al variograma gaussiano con el menor efecto pepa posible, es decir, el contacto más suave entre exterior e interior del cuerpo. Corolario

de este resultado es que el ajuste de los parámetros del variograma no es necesario, pues no existe múltiples óptimos entre los cuales el usuario pueda elegir. También se puede concluir al respecto que tampoco es necesario un algoritmo para calibrar los parámetros o encontrar dicho óptimo, pues el óptimo es conocido: aquel que genera contactos más suaves, es decir un variograma gaussiano con el menor efecto pepa.

El algoritmo de kriging, como se explicó en los antecedentes, consiste, *grosso modo*, en recorrer cada elemento de la grilla ( $n$ ) y resolver un sistema lineal de tamaño acotado por el número de vecinos ( $v$ ). De lo anterior se desprende que el orden del algoritmo depende linealmente de la cantidad de nodos ( $n$ ) multiplicado por el orden de resolver el sistema ( $v^p$ ) ( $p$  es variable entre 2 y 3, según el algoritmo usado). Para el caso de la simulación este orden se debe multiplicar, adicionalmente, por la cantidad de realizaciones escogida por el usuario ( $r$ ), pues una realización equivale, en términos de la complejidad, a realizar kriging. A su vez, las modificaciones hechas a los algoritmos, introducen cambios en la búsqueda y manipulación de los vecinos, agregando un esfuerzo extra al procedimiento realizado en cada nodo, el cual, en ambos casos (sdh y lvo), no supera la complejidad de resolver el sistema, manteniendo así el orden original  $\mathcal{O}(n \cdot v^p)$ . En términos prácticos, los términos descartados por ser de menor orden, y los cambios en las constantes que se omite en el análisis, generan aumentos en el tiempo real de cálculo. Los tiempos empíricos medidos para kriging y simulación aumentan en promedio cerca de un 25 % para los algoritmos con pseudosondajes y alrededor de un 100 %, en promedio, para los algoritmos con orientación variable.

A modo de resumen, en kriging todas modificaciones hechas muestran mejoras, donde lva obtiene los mejores resultados, seguido de lvo y, finalmente, por sdh. En el caso de la simulación, el acierto total da mejores resultados para todos los algoritmos respecto del caso base, el algoritmo lva muestra los mejores resultados, seguido de sdh y lvo. Por los motivos expuestos, estos algoritmos surgen como una buena alternativa en escenarios tempranos de la exploración, donde la información dura, proveniente de sondajes, es exigua y donde la interpretación geológica es un factor importante. Es importante recalcar que un buen resultado dependerá de una buena interpretación y, por lo tanto, de buena información blanda.

En términos prácticos, la ventaja de los algoritmos desarrollados, frente a otros que también integran la experiencia del usuario en el cálculo, es la simplicidad al incorporar esta información blanda. Estos algoritmos constituyen el *back-end* para herramientas de software donde el usuario indique la continuidad basándose en el despliegue de los datos de sondajes. Por otro lado se tiene algoritmos como la simulación multipunto, donde esta integración se hace mediante una imagen, la cual no es fácil construir y se puede caer de manera muy fácil en inconsistencias; o como ocurre en otros métodos donde la información experta se debe transformar en una distribución de probabilidad que condiciona los resultados, donde la intuición y el mecanismo por el cual se incorpora la información no son directas.

Como trabajo futuro se sugiere el estudio del uso de los pseudosondajes como paso intermedio para la creación de un campo de direcciones. La generación de un campo de direcciones es un problema que aún está siendo estudiado al momento de este trabajo. La generación de la información blanda abre también un abanico de posibilidades a explorar. En este trabajo la generación de los datos blandos se mantuvo simple y totalmente guiado por el usuario, sin embargo, la metodología propuesta permite una gran cantidad de posibles mejoras. Entre éstas está la automatización de ciertas etapas, uso de algoritmos avanzados

para la generación de superficies, uso de más información para la confección de los datos blandos, como geología, alteración, litología, índices de dureza, etc.

# Bibliografía

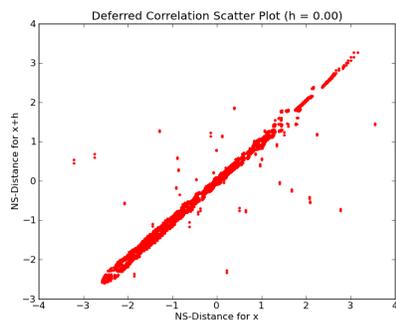
- [1] <http://www.gslib.com>.
- [2] BOISVERT, J. B. *Geostatistics with Locally Varying Anisotropy*. PhD thesis, University of Alberta, 2010.
- [3] BOISVERT, J. B., AND DEUTSCH, C. V. Programs for kriging and sequential gaussian simulation with locally varying anisotropy using non-euclidean distances. *Computers & Geosciences* 37, 4 (Apr. 2011), 495–510.
- [4] CÁCERES, A., EMERY, X., AEDO, L., AND GÁLVEZ, O. Stochastic geological modeling using implicit boundaries simulation. In *Geomin 2011* (June 2011).
- [5] DEUTSCH, C. V., AND JOURNAL, A. *Gslib: Geostatistical Software Library and User's Guide*. Oxford University Press, Inc., 1998.
- [6] DUKE, J. H., AND HANNA, P. J. Geological Interpretation for resource Modelling and Estimation. *Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice* (2001).
- [7] EMERY, X. *Apunte de Geoestadística*. Departamento de Ingeniería en Minas, Universidad de Chile, 2009.
- [8] EMERY, X., AND LANTUÉJOL, C. Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences* 32, 10, 1615–1628.
- [9] GOOVAERTS, P. *Geostatistics for natural resources evaluation*. Oxford University Press, USA, 1997.
- [10] GUARDIANO, F., AND SRIVASTAVA, R. Multivariate geostatistics: Beyond bivariate moments. In *Geostatistics Tróia '92*, A. Soares, Ed., vol. 5 of *Quantitative Geology and Geostatistics*. Springer Netherlands, 1993, pp. 133–144.
- [11] HASSANPOUR, R. M., AND DEUTSCH, C. V. An Introduction to Grid-Free Object-Based Facies Modeling. *CCG Annual Report 12* (2010).
- [12] HENRION, V., CAUMON, G., AND CHERPEAU, N. ODSIM: An Object-Distance Simulation Method for Conditioning Complex Natural Structures. *Mathematical Geosciences* 42, 8 (2010), 911–924.

- [13] HONARKHAH, M., AND CAERS, J. Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences* 42, 5 (2010), 487–517.
- [14] KAUFMANN, O., AND MARTIN, T. 3D geological modelling from boreholes, cross-sections and geological maps, application over former natural gas storages in coal mines. *Computers & Geosciences* 34, 3 (2008), 278–290.
- [15] LEIVA, A. Simulación geoestadística incorporando un campo de direccione variables. *Memoria de título* (2007).
- [16] LI, Y., AND DEUTSCH, C. V. Facies Modeling with Direct Multivariate Probability Estimation. *CCG Annual Report 12* (2010).
- [17] LILLAH, M., AND BOISVERT, J. B. Stochastic distance based geological boundary modeling with curvilinear features. *Mathematical Geosciences* 45, 6 (2013), 651–665.
- [18] MARIETHOZ, G., RENARD, P., AND STRAUBHAAR, J. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research* 46, 11 (2010).
- [19] ORTIZ, J. Apuntes del diplomado de evaluación geoestadística de yacimientos.
- [20] RIVARA, M. Mesh Refinement Processes Based on the Generalized Bisection of Simplices. *SIAM Journal on Numerical Analysis* 21, 3 (1984), 604–613.
- [21] SRIVASTAVA, R. Probabilistic modeling of ore lens geometry: An alternative to deterministic wireframes. *Mathematical Geology* 37, 5 (2005), 513–544.
- [22] STREBELLE, S. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology* 34, 1 (2002), 1–21.
- [23] WILDE, B., AND DEUTSCH, C. Kriging and simulation in presence of stationary domains: Developments in boundary modeling. In *Geostatistics Oslo 2012*, P. Abrahamsen, R. Hauge, and O. Kolbjørnsen, Eds., vol. 17 of *Quantitative Geology and Geostatistics*. Springer Netherlands, 2012, pp. 289–300.
- [24] WILDE, B. J., AND DEUTSCH, C. V. A New Way to Calibrate Distance Function Uncertainty. *CCG Annual Report 13* (2011).
- [25] WILDE, B. J., AND DEUTSCH, C. V. Simulating Boundary Realizations. *CCG Annual Report 13* (2011).
- [26] WU, J., BOUCHER, A., AND ZHANG, T. A sgems code for pattern simulation of continuous and categorical variables: Filtersim. *Computers & Geosciences* 34, 12 (Dec. 2008), 1863–1876.
- [27] XU, C., AND DOWD, P. Optimal construction and visualisation of geological structures. *Computers & Geosciences* 29, 6 (2003), 761–773.
- [28] ZHANG, T., SWITZER, P., AND JOURNAL, A. Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology* 38, 1 (2006), 63–80.

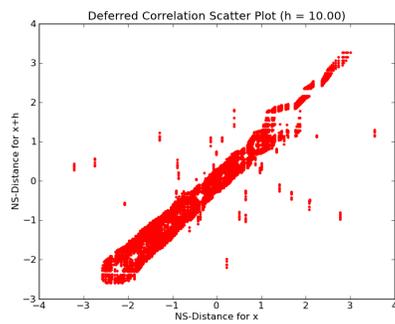
# Apéndice A

## Validación hipótesis Multigaussiana

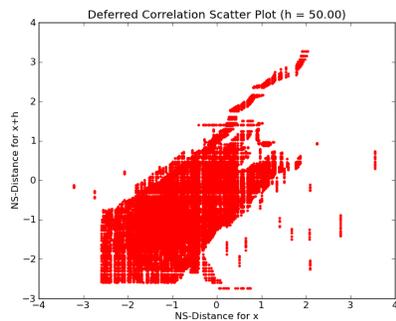
Para verificar que la variable transformada cumple con la hipótesis gaussiana se calculó la nube de correlación diferida, la cual se muestra en la Figura A.1. En ésta se puede ver que para las distancias pequeñas la nube, si bien no es tan dispersa, sigue la tendencia de agruparse en torno a la diagonal, indicando la correlación entre puntos a distancias pequeñas y, cuando la separación es mayor, comienza a dispersarse hasta tomar una forma circular, símbolo de que se pierde la correlación.



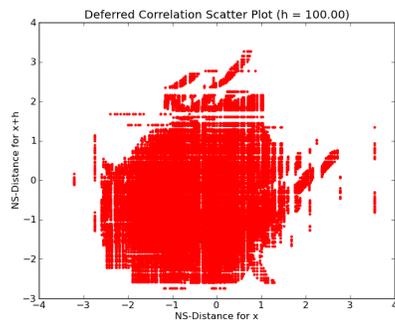
(a) Distancia: 0 m



(b) Distancia: 10 m



(c) Distancia: 50 m



(d) Distancia: 100 m

Figura A.1: Nubes de correlación diferida.

En la Figura A.2 se muestra la razón entre la raíz del variograma y el madograma, cuyo resultado debiese ser  $\sqrt{\pi}$ . La gráfica muestra como para la dirección omnihorizontal se mueve cerca del valor  $\sqrt{\pi}$  y, en el caso de la vertical describe una curva que tiende a acercarse al mismo valor.

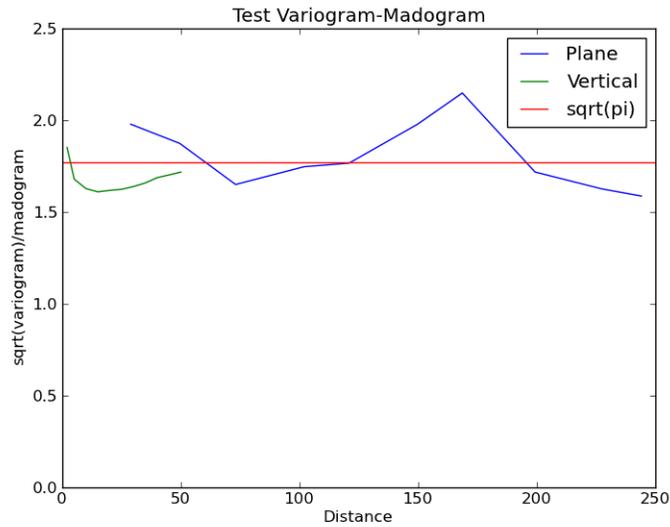


Figura A.2: Razón entre raíz del variograma y madograma.

# Apéndice B

## Control de rugosidad

El mecanismo de control de la rugosidad a través del variograma considera dos grados de libertad, estos corresponden al efecto pepa y la estructura utilizada en el modelo variográfico. En las Figura B.1 y B.2 se muestra como varía el borde del cuerpo en un corte realizado a una realización particular. En la primera se varía el efecto pepa de 0.05 a 0.50 en intervalos de 0.05, utilizando las estructuras exponencial y gaussiana en porcentajes 50 % y 50 %; en la segunda se utilizó un efecto pepa de 0.05, combinando las estructuras desde 100 % exponencial a 100 % gaussiano en intervalos de 10 %. Cabe destacar que las realizaciones se pueden comparar directamente, pues la semilla de los números aleatorios usados en la simulación se mantiene.

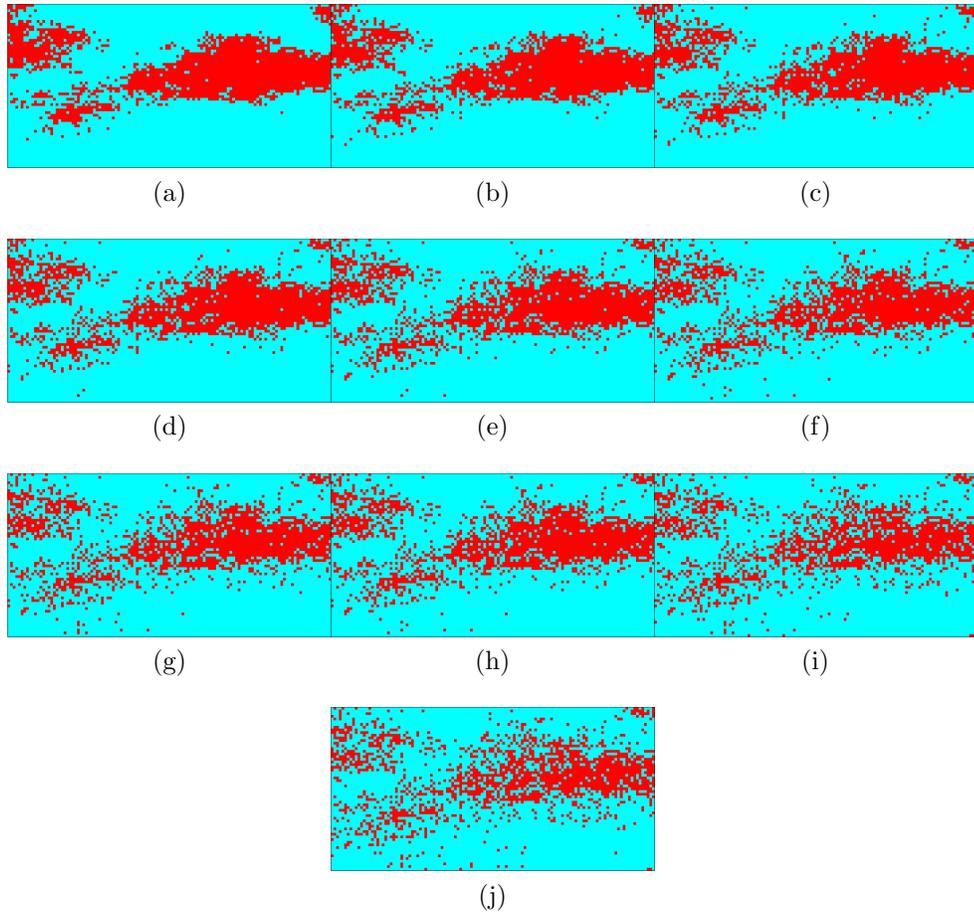


Figura B.1: Cortes para simulaciones con diferentes valores de efecto pepa, desde 0.05 (a) hasta 0.50 (j) en intervalos incrementales de 0.05.

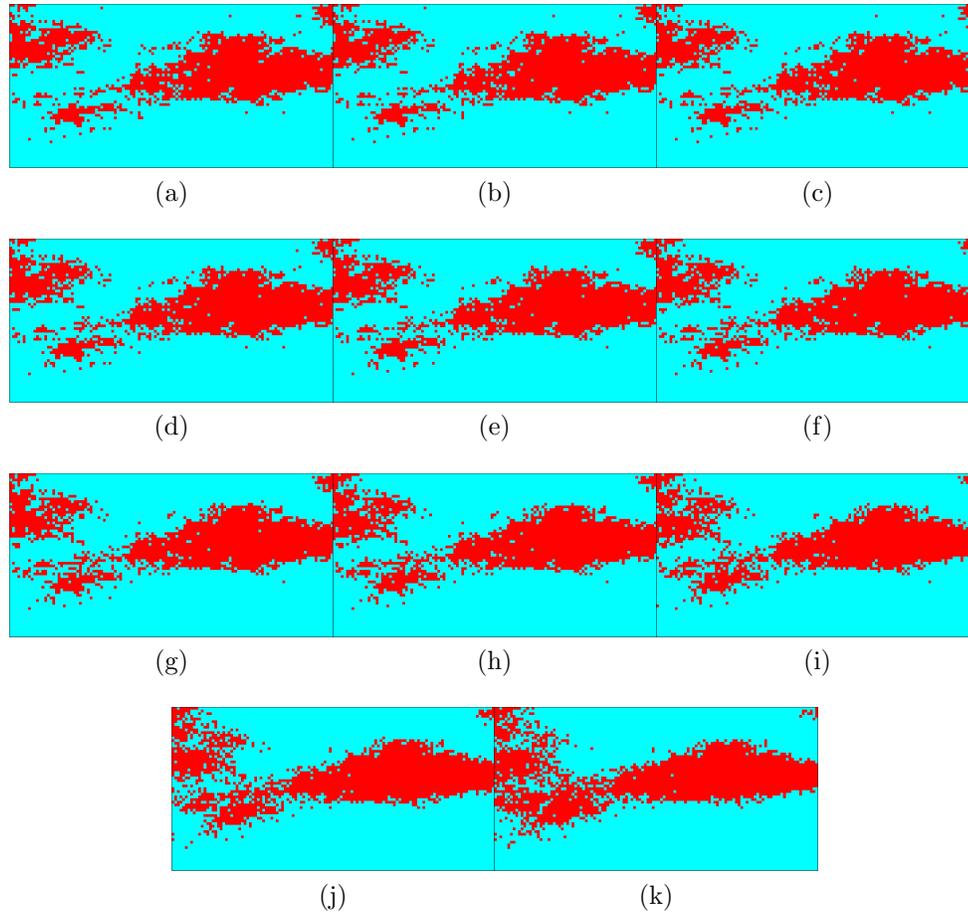


Figura B.2: Cortes para simulaciones con diferentes aportes de las estructuras, desde 100 % exponencial(a) a 100 % gaussiano(k) en intervalos de 10 %.

En la Figura B.3 se muestra un mapa de colores indicando el acierto total promedio al variar el efecto pepa (horizontal) y la estructura (vertical). Este gráfico muestra que el acierto máximo se da para un variograma gaussiano en un 100 % y con el menor efecto pepa, luego no es necesario crear un algoritmo para sugerir el nivel de rugosidad, pues se sabe cual es la combinación que tiene mejor acierto.

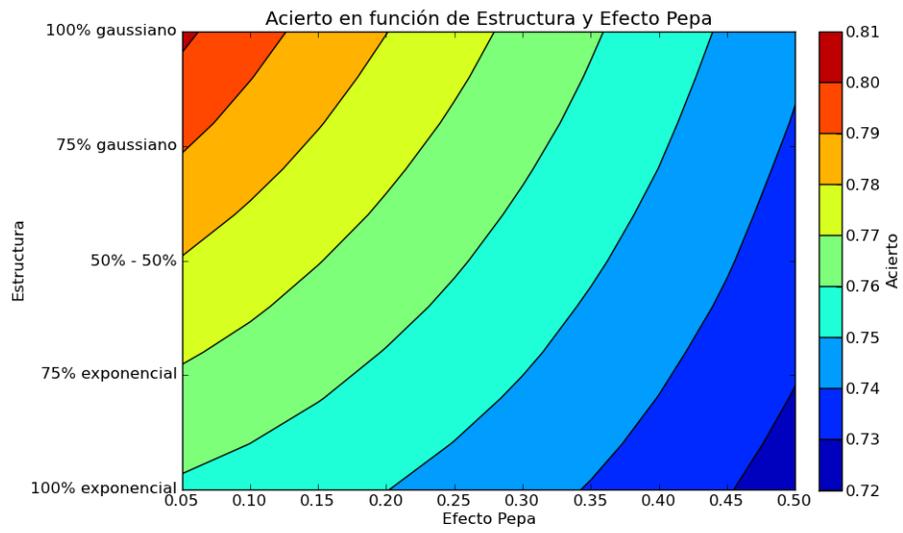


Figura B.3: Mapa de aciertos en función del efecto pepa y de la estructura.