

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA DE MINAS

**SIMULACIÓN MULTI-PUNTOS UTILIZANDO EL ALGORITMO FILTERSIM:
ANÁLISIS DE SENSIBILIDAD Y CASOS DE ESTUDIO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL DE MINAS

SEBASTIÁN ALEJANDRO ÁVALOS SOTOMAYOR

PROFESOR GUÍA:

JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:

XAVIER EMERY

EDUARDO MAGRI VARELA

SANTIAGO DE CHILE

ABRIL 2013

RESUMEN

La simulación de estructuras complejas de variables continuas basada en algoritmos de simulación multipuntos es un campo que comenzó su exploración hace no más de 20 años. La aplicación de estas técnicas en la caracterización de recursos de yacimientos mineros es aun escasa. No obstante, la propiedad de no suavizar las leyes estimadas lleva a reproducir la realidad de forma más fidedigna y, por consiguiente, tomar decisiones técnico-económicas con mejores resultados en los planes mineros. Dentro de los algoritmos multipuntos, FILTERSIM utiliza la generación de patrones estructurales a partir de la aplicación de filtros sobre imágenes de entrenamiento.

El objetivo de la presente memoria es analizar la sensibilidad de FILTERSIM a su conjunto de parámetros de entrada. Para ello se desarrolló un caso de estudio sintético considerando como criterios de análisis tanto el tiempo de cómputo como la calidad de los resultados. Los parámetros con mayor impacto fueron el tamaño de la grilla, que infiere los patrones de comportamiento estructural desde la imagen de entrenamiento, y la función que asocia dichos patrones a los eventos condicionantes en la grilla simulada.

Basado en los parámetros óptimos, como resultado del caso de estudio sintético, se decidió llevar a cabo un caso de estudio real, representado por datos de sondajes y posos de tronadura de un yacimiento cuprífero. Los resultados de la simulación fueron contrastados con las estimaciones mediante kriging ordinario. Se observó que la media de las realizaciones no tiende a la estimación de kriging, contradiciendo lo esperado. Por otro lado, los mapas de probabilidad entregaron un comportamiento de la variable similar al de los valores estimados mediante kriging.

Finalmente, se proponen dos alternativas para incorporar la media local al proceso de simulación utilizando FILTERSIM. La primera consiste en intervenir la forma en que se asocia el patrón estructural más cercano al evento condicionante. La segunda busca alterar el proceso de agrupación de los patrones inferidos de la imagen de entrenamiento.

ABSTRACT

The simulation of complex structures of continuous variables based in multipoint simulation algorithms is a field that began 20 years ago. The application of these techniques in relation to the characterization of ore deposits resources is still scarce. However, the simulations does not have the property of smoothing the grades, compared to the kriging methods, and with this can reproduce the reality more reliably, therefore, allowing to take technical and economic decisions with better results. Within these algorithms, FILTERSIM uses the structural pattern generation based on the use of filters upon training images.

The main goal of this work is to analyze the sensibility of FILTERSIM to its set of in-put parameters. In order to achieve it, a synthetic case of study was developed considering as criteria both computational time and quality of results. The parameters with more influence were the grid size, which infers the structural behavior pattern from the training image, and the function that associates these patterns to the conditional data event within the simulated grid.

Based on the optimal parameters, as a result of the synthetic case of study, it was decided to carry out a real case of study, represented by drill holes data and wells of a copper deposit. The simulation results were compared with estimations using ordinary kriging. It was showed that the average of the realizations does not tend to the kriging estimation, contradicting the expected behavior. Moreover, probability maps showed a variable behavior similar to the values estimated by kriging.

Finally, two alternatives are proposed in order to incorporate the local mean to the simulation process using FILTERSIM. The first one consists to intervene the way in which the closet pattern is associated to the conditional data event. The second one seeks to alter the clustering process of pattern inferred from the training image.

AGRADECIMIENTOS

Quiero partir agradeciendo al profesor Julián Ortiz por el apoyo y la dedicación entregados al desarrollo de la presente memoria desde la propuesta de tema hasta las últimas críticas y recomendaciones. Agradezco la contribución y el tiempo dedicado, por parte de Xavier Emery y Eduardo Magri, a construir este trabajo de una manera más sólida y realista.

A la Corporación Nacional del Cobre (CODELCO) por haberme beneficiado con la beca de excelencia académica a lo largo de la carrera. La cual fue un apoyo económico fundamental que repercutió favorablemente en muchos aspectos académicos y personales.

A la escuela de ingeniería de la Universidad de Chile, por haberme permitido estudiar en un ambiente pluralista, intelectual y de alta infraestructura. Elementos necesarios para un adecuado desarrollo académico.

Sin duda alguna agradezco el apoyo incondicional de mi familia durante todo el proceso universitario. Por los valores inculcados, la comprensión en situaciones difíciles y el amor por sobre todas las cosas. A mis amigos y amigas que en tantas oportunidades me han brindado su apoyo, alegría y desconexión de la rutina del estudio. En especial a Yerko Martínez, Yerko Yavar, José Fuenzalida y María Francisca Rubilar por estar siempre presentes tanto en los trabajos más difíciles como en los momentos más relajados.

Finalmente, agradezco la posibilidad de los viajes realizados, las experiencias adquiridas y a la naturaleza por tener siempre una forma nueva de mostrar que en la simpleza de las cosas está la belleza y que el equilibrio es fundamental en esta vida tan efímera.

ÍNDICE

1	INTRODUCCIÓN.....	1
1.1	Motivación del tema	2
1.2	Objetivos	4
1.2.1	Objetivo general	4
1.2.2	Objetivos específicos.....	4
1.3	Alcances.....	4
2	ESTADO DEL ARTE	5
2.1	Simulación estocástica multipuntos.....	6
2.1.1	Algoritmo SNESIM.....	7
2.1.2	Algoritmo FILTERSIM.....	13
2.1.3	Elección de algoritmo a utilizar.....	31
2.2	Stanford Geostatistical Modeling Software (SGeMS).....	32
2.2.1	Interfaz, estructura y utilidades	32
2.2.2	FILTERSIM	35
3	METODOLOGÍA.....	38
4	APLICACIÓN FILTERSIM CONVENCIONAL: CASO DE ESTUDIO SINTÉTICO.....	40
4.1	Metodología para el desarrollo del caso de estudio.....	40
4.2	Presentación del caso de estudio.....	41
4.2.1	Imagen de entrenamiento (T.I.)	42
4.2.2	Datos condicionantes.....	44
4.3	Variables a considerar.....	46
4.4	Criterios de evaluación	49

4.5	Resultado de las simulaciones	51
4.5.1	Promedio entre las 35 realizaciones	51
4.5.2	Tiempo de cómputo	59
4.5.3	Dispersión en el promedio y la desviación estándar, entre realizaciones.....	61
4.5.4	Promedio y varianza sobre el promedio de las realizaciones en cada simulación ..	65
4.5.5	Mapas probabilísticos	69
4.5.6	Análisis variográfico.....	75
4.6	Análisis de resultados	76
5	APLICACIÓN FILTERSIM CONVECCIONAL: CASO DE ESTUDIO REAL.....	78
5.1	Metodología	78
5.2	Descripción del yacimiento	79
5.3	Selección imágenes de entrenamiento (T.I.)	82
5.4	Selección de datos condicionantes: sondeos.....	85
5.5	Simulación	89
5.5.1	Parámetros de entrada FILTERSIM	89
5.5.2	Resultados utilizando la T.I. 1 sobre el banco 3860.....	90
5.5.3	Resultados utilizando la T.I. 2 sobre el banco 3850.....	92
5.5.4	Resultados utilizando la T.I. 3 sobre el banco 3830.....	94
5.6	Comparación con estimaciones mediante kriging ordinario.....	96
5.6.1	Resultados utilizando la T.I. 1 sobre el banco 3860.....	96
5.6.2	Resultados utilizando la T.I. 2 sobre el banco 3850.....	98
5.6.3	Resultados utilizando la T.I. 3 sobre el banco 3830.....	101
5.7	Curvas tonelaje-ley	104
5.7.1	Banco 3860.....	104
5.7.2	Banco 3850.....	105

5.7.3	Banco 3830.....	106
5.8	Análisis y discusión	107
6	ALTERNATIVAS DE MODIFICACION AL CÓDIGO FILTERSIM	108
6.1	Modificación de la función distancia.....	108
6.2	Consideración de la media local en la determinación del patrón final, nuevo filtro. ...	109
7	CONCLUSIONES.....	111
8	RECOMENDACIONES A FUTURO	113
9	BIBLIOGRAFÍA	114
10	ANEXO A	116
11	ANEXO B	118
12	ANEXO C	123
13	ANEXO D	135

ÍNDICE DE FIGURAS

Figura 1 Tres Imágenes de Entrenamiento con continuidad espacial distinta. Fuente: MPS short course (Jef Caers, Boucher, 2008).....	2
Figura 2 Variogramas Norte - Sur y Este - Oeste para los conjuntos estructurales de la figura 1. Fuente: MPS short course (Jef Caers, Boucher, 2008).....	2
Figura 3 Secuencia estándar de una simulación basada en estadística multipuntos. Fuente: (Zhang et al, 2006).....	3
Figura 4 Ejemplo ilustrativo de un evento condicionante con 3 posibles estados y 4 datos condicionantes	8
Figura 5 Ejemplo árbol de condicionamiento. Fuente: [Emery, 2012].	9
Figura 6 Ejemplo de función <i>cpdf</i>	10
Figura 7 Ejemplo, determinación del evento condicionante. (a) Vecindad móvil, (b) grilla de simulación, (c) evento condicionante.	12
Figura 8 Determinación del valor asociado al nodo central. (a) Uso de la <i>cpdf</i> dado el evento condicionante, (b) asignación del valor al nodo central.	12
Figura 9 Ejemplo de una realización con SNESIM (caso ficticio). (a) Grilla a simular con datos condicionantes, (b) mitad del proceso de simulación, (c) grilla de simulación completa.	12
Figura 10 Imágenes de entrenamiento. (a) T.I. de una variable binaria – porosidad media. (b) T.I. de una variable categórica – Depositación fluvial con cuatro categorías. (c) T.I. de una variable continua – Distribución de la porosidad. Fuente: Filter-Based training pattern classification for spatial pattern simulation (Zhang et al., 2006).....	16
Figura 11 Obtención de patrones a partir de una grilla de búsqueda desde una imagen de entrenamiento	16
Figura 12 Filtros asociados a la media. (a) Norte-Sur, (b) Este-Oeste, (c) Elevación. La escala representa los pesos asociados a cada nodo en la grilla	18
Figura 13 Filtros asociados al gradiente. (d) Norte-Sur, (e) Este-Oeste, (f) Elevación. La escala representa los pesos asociados a cada nodo en la grilla	18

Figura 14 Filtros asociados a la curvatura. (g) Norte-Sur, (h) Este-Oeste, (i) Elevación. La escala representa los pesos asociados a cada nodo en la grilla	19
Figura 15 Caracterización de un patrón, asignación de puntaje, al aplicar un filtro en específico	20
Figura 16 Ilustración del concepto de imagen erodada	20
Figura 17 Ubicación de un patrón en el espacio de puntuación de F dimensiones. (a) Imagen de entrenamiento, (b) F mapas de puntuación, (c) vector representativo de la ubicación de la puntuación del patrón en el espacio de F dimensiones.....	21
Figura 18 Ejemplo de una clase. (a) Prototipo que identifica a la clase, (b) patrones similares dentro de una misma clase.....	23
Figura 19 Ejemplo del proceso de agrupamiento utilizando el método <i>K-mean</i> . Se observan 3 centroides C1, C2 y C3 que llevan a generar las 3 líneas continuas que establecen los puntos equidistantes a los centroides más cercanos.	25
Figura 20 Ilustración de 3 multigrillas de tamaño 5x5 sobre una imagen de entrenamiento	31
Figura 21 Interfaz SGeMS Beta 2.1	32
Figura 22 FILTERSIM en SGeMS. Se muestran las pestañas: (a) General, (b) Condicionamiento, (c) Región, (d) Avanzado.	35
Figura 23 Distribución de los valores en la imagen de entrenamiento.....	42
Figura 24 (a) Imagen de entrenamiento, (b) Mapa variográfico sobre la T.I., (c) variogramas experimentales en las direcciones de principales de anisotropía.....	43
Figura 25 Distribución espacial de los datos condicionantes DC1, DC2, DC3 y DC4.....	44
Figura 26 Promedio y varianza de los datos condicionantes.....	45
Figura 27 Promedio de 35 realizaciones. Método distancia: filter score. DC 1	51
Figura 28 Promedio de 35 realizaciones. Método distancia: filter score. DC 2.....	52
Figura 29 Promedio de 35 realizaciones. Método distancia: filter score. DC 3	53
Figura 30 Promedio de 35 realizaciones. Método distancia: filter score. DC 4.....	54
Figura 31 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 1	55

Figura 32 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 2....	56
Figura 33 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 3....	57
Figura 34 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 4....	58
Figura 35 Tiempo de cómputo al simular considerando 35 realizaciones y utilizando el método <i>template pixels</i>	60
Figura 36 Tiempo de cómputo al simular considerando 35 realizaciones y utilizando el método <i>Filter scores</i>	60
Figura 37 Promedio de los valores internos de las simulaciones – Método: <i>Template pixels</i>	61
Figura 38 Varianza de los valores internos de las simulaciones – Método: <i>Template pixels</i>	62
Figura 39 Promedio de los valores internos de las simulaciones – Método: <i>Filter scores</i>	63
Figura 40 Varianza de los valores internos de las simulaciones – Método: <i>Filter scores</i>	64
Figura 41 Promedio sobre cada simulación. Método: template pixels.....	65
Figura 42 Promedio sobre cada simulación. Método: filter scores	66
Figura 43 Varianza sobre cada simulación. Método: template pixels.....	67
Figura 44 Varianza sobre cada simulación. Método: filter scores	68
Figura 45 Mapas de probabilidad. Configuración 11x11 – 7x7.....	69
Figura 46 Mapas de probabilidad. Configuración 21x21 – 7x7.....	70
Figura 47 Mapas de probabilidad. Configuración 21x21 – 13x13.....	71
Figura 48 Mapas de probabilidad. Configuración 41x41 – 7x7.....	72
Figura 49 Mapas de probabilidad. Configuración 41x41 – 13x13.....	73
Figura 50 Mapas de probabilidad. Configuración 41x41 – 25x25.....	74
Figura 51 Variogramas experimentales sobre una sola realización. Configuraciones: 11x11/7x7 y 41x41/25x25.	75
Figura 52 Visualización en 3 dimensiones del yacimiento completo	80
Figura 53 Vista YZ del yacimiento completo	80

Figura 54 Vista XZ del yacimiento completo	81
Figura 55 Vista XY del yacimiento completo	81
Figura 56 Selección primera imagen de entrenamiento – Banco 3952	82
Figura 57 Selección segunda imagen de entrenamiento – Banco 3940	83
Figura 58 Selección tercera imagen de entrenamiento – Banco 3928.....	84
Figura 59 (a) Vista en planta de los sondajes, (b) vista en 3D de los sondajes.....	85
Figura 60 Vista lateral en dirección YZ del conjunto de sondajes.....	85
Figura 61 Vista lateral en dirección XZ del conjunto de sondajes.....	86
Figura 62 Vista en planta de sondajes en secciones de a 10 metros entre 3,820 y 3,860.....	86
Figura 63 Vista en planta de sondajes en secciones de a 10 metros entre 3,860 y 3,890.....	86
Figura 64 Bancos escogidos a simular. A la izquierda los datos en grilla irregular y a la derecha en grilla regular.....	87
Figura 65 Histogramas de leyes en los datos condicionantes. (a) DC1, (b) DC2, (C) DC3.....	88
Figura 66 Histograma leyes media - Banco 3860.....	90
Figura 67 Seis realizaciones sobre el banco 3860	91
Figura 68 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3860.....	91
Figura 69 Histograma leyes media – Banco 3850.....	92
Figura 70 Seis realizaciones sobre el banco 3850	93
Figura 71 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3850.....	93
Figura 72 Histograma leyes media – Banco 3830.....	94
Figura 73 Seis realizaciones sobre el banco 3830	95
Figura 74 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3830.....	95
Figura 75 Variograma experimental omni-direccional sobre datos condicionantes 1	96
Figura 76 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3860....	97

Figura 77 Histograma de valores en el banco 3860 mediante kriging, simulacion y datos de pozo	97
Figura 78 Varianza condicional y varianza de kriging en banco 3860	98
Figura 79 Variograma experimental omni-direccional sobre datos condicionantes 2	99
Figura 80 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3850....	99
Figura 81 Histograma de valores en el banco 3850 mediante kriging, simulacion y datos de pozo	100
Figura 82 Varianza condicional y varianza de kriging en banco 3850	101
Figura 83 Variograma experimental omni-direccional sobre datos condicionantes 3	101
Figura 84 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3830..	102
Figura 85 Histograma de valores en el banco 3830 mediante kriging, simulacion y datos de pozo	102
Figura 86 Varianza condicional y varianza de kriging en banco 3830	103
Figura 87 Curvas tonelaje-ley en banco 3860	104
Figura 88 Curvas tonelaje -ley en banco 3850	105
Figura 89 Curvas tonelaje-ley en banco 3830	106
Figura 90 Septimo filtro	109

ÍNDICE DE TABLAS

Tabla 1 Ventajas y desventajas de los algoritmos de agrupamiento Cross y K-mean [Austin, 2007].....	26
Tabla 2 Dimensiones de la Imagen de Entrenamiento	42
Tabla 3 Estadística imagen de entrenamiento	42
Tabla 4 Variables de entrada - Mapa variográfico	43
Tabla 5 Estadística básica de los datos condicionantes	45
Tabla 6 Variables a analizar y sus valores, en la simulación	47
Tabla 7 Tiempo simulación, considerando 35 realizaciones. Software: SGeMS.....	59

1 INTRODUCCIÓN

La presente memoria se enmarca en la realización de dos grandes proyectos:



Su desarrollo nace bajo la necesidad de perfeccionar los métodos de simulación para estructuras complejas basadas en imágenes de entrenamiento utilizando simulación estocástica multipuntos.

- El proyecto Fondecyt 1090056 “*Multiple Point Geostatistics for the Evaluation of Uncertainty in Geological Attributes and Grades*” está ya terminado (Marzo 2009 a Febrero 2012). No obstante, la presente memoria guarda relación con la temática del proyecto tanto en conocimiento como en recursos (financiamiento).
- El proyecto CAM-G (BHP- Billiton) tiene por objetivo la generación de herramientas de modelamiento geológico semi-automático basado en la utilización de bases de datos e interpretaciones geológicas para el modelamiento de estructuras sólidas. Se busca además, dentro de las herramientas ya existentes, aumentar su velocidad en el proceso de cómputo para llevar a cabo evaluaciones de recursos en el menor tiempo posible.

1.1 Motivación del tema

Para realizar una simulación adecuada, previamente se debe haber caracterizado la estructura espacial de la variable a simular. El variograma es la herramienta que con mayor frecuencia se utiliza para caracterizar correlaciones espaciales, la cual está asociada principalmente a la estimación mediante kriging. No obstante, existen estructuras cuyo comportamiento no logra ser caracterizado utilizando tan sólo 2 puntos en el espacio. La figura 1 [Caers & Boucher, 2008] presenta el caso de 3 conjuntos de estructuras cuyas proporciones son idénticas y se observa que al realizar un análisis variográfico no se logra apreciar diferencias significativas entre un conjunto y otro.

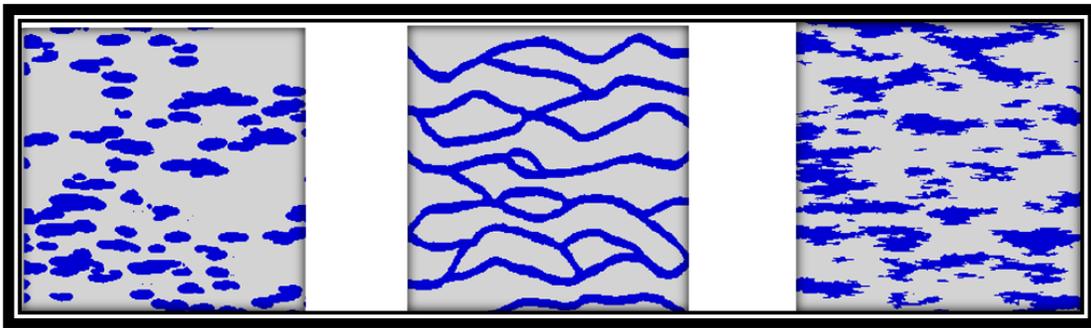


Figura 1 Tres Imágenes de Entrenamiento con continuidad espacial distinta. Fuente: MPS short course (Jef Caers, Boucher, 2008)

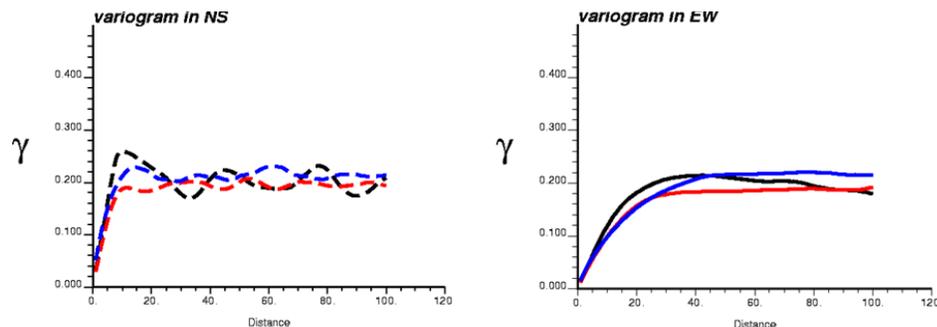


Figura 2 Variogramas Norte - Sur y Este - Oeste para los conjuntos estructurales de la figura 1. Fuente: MPS short course (Jef Caers, Boucher, 2008).

Como se observa en la figura 2, el alcance y la meseta para los tres variogramas tanto en dirección Norte-Sur como en Este-Oeste son muy similares, siendo que sus continuidades espaciales no lo son. Para lograr capturar estos patrones de comportamiento se utilizan las metodologías de simulación basadas en múltiples puntos. Para ello se requiere de una o más

imágenes de entrenamiento, **T.I.**, (por sus siglas en inglés, *Training Image*) que proveen la información de los posibles patrones de comportamiento de la estructura. Estas imágenes de entrenamiento pueden venir como una interpretación geológica de una zona, los resultados de una exploración geofísica, la reconstrucción de una zona ya estudiada similar a una desconocida, etc. Además, se debe contar con datos reales medidos en determinados puntos del espacio (por ejemplo, sondajes, muestras superficiales, etc.) considerados como datos duros. Con ambas informaciones, más la utilización de algún algoritmo (actualmente Snesim [Strebelle, 2002] o Filtersim [Zhang et al, 2006]), es posible llevar a cabo una simulación basada en estadísticas multipuntos.

A modo de simplificación, se observa en la figura 3 la secuencia lógica de simulación:

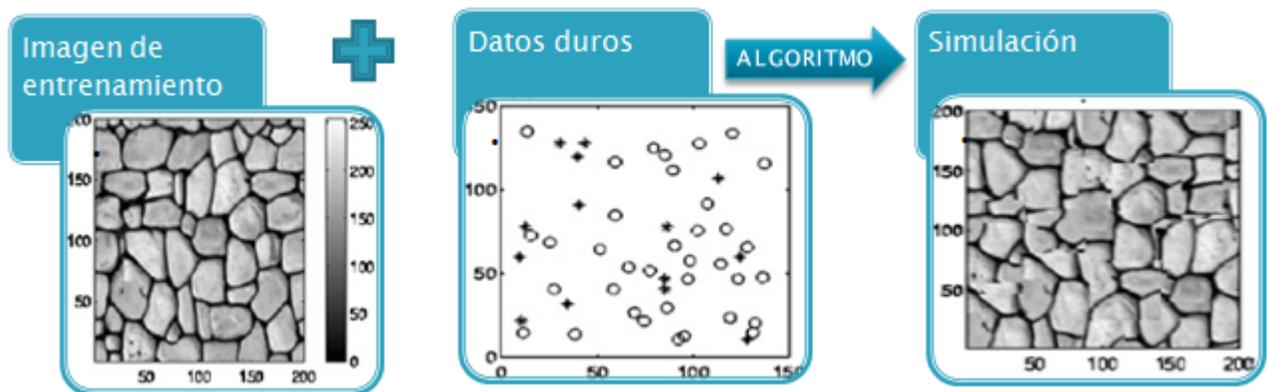


Figura 3 Secuencia estándar de una simulación basada en estadística multipuntos. Fuente: (Zhang et al, 2006)

Poder complementar este proceso de simulación con información secundaria llevaría a los resultados a un mayor acercamiento a la realidad. Amparado en esta idea, se plantea la posibilidad de integrar variabilidad en la media de manera local. Esto permitiría al usuario, basado en su propia experiencia o datos secundarios, poder variar la media local en las simulaciones resultantes.

1.2 Objetivos

Para llevar a cabo un estudio y análisis de los algoritmos existentes que guardan relación con la simulación estocástica multipuntos se despliegan los siguientes objetivos generales y específicos.

1.2.1 Objetivo general

- Comprender los procesos internos del algoritmo, análisis de sensibilidad bajo el desarrollo tanto de un caso de estudio sintético como de uno real (yacimiento cuprífero) y propuesta de incorporación de la media local utilizando siempre el algoritmo FILTERSIM.

1.2.2 Objetivos específicos

- Elegir el algoritmo adecuado para adaptar la incorporación de la media local.
- Determinar una forma adecuada para incluir la media local en el algoritmo definido.
- Desarrollar dos casos de estudio (sintético y real) utilizando las librerías de software geostatístico (GSLib) y los softwares SGeMS y U-mine.
- Analizar la repercusión en la calidad de los resultados de la simulación al variar los distintos parámetros de entrada de FILTERSIM.
- Definir la configuración de parámetros de entrada que lleve a mejores resultados en las simulaciones.
- Comparar entre estimación mediante kriging ordinario y simulación mediante FILTERSIM en el caso de estudio real (yacimiento cuprífero).

1.3 Alcances

De la presente memoria se establecen como alcances:

- Exponer el estado del arte respecto a la simulación multipuntos.
- Establecer las ventajas y desventajas de los algoritmos de simulación expuestos en el estado del arte.
- Proponer una metodología para incorporar la media local en la utilización de FILTERSIM generando así una versión modificada del algoritmo.
- Fijar los parámetros de entrada adecuados para el desarrollo de una posible comparación entre el algoritmo FILTERSIM original y la propuesta de modificación, basado en el mismo caso de estudio sintético expuesto en la presente memoria.

2 ESTADO DEL ARTE

La evaluación de los recursos naturales (minería, industria petrolífera, etc.), la interpretación de fenómenos naturales (sismología, geología estructural, etc.), entre otras ciencias aplicadas al campo de la tierra han llevado al desarrollo de la actual geoestadística. El concepto de geoestadística fue introducida por primera vez en 1960 por Matheron, pudiendo describirla como la interpretación estadística, a través de funciones aleatorias, de fenómenos naturales regionalizados. Desde entonces muchas técnicas de estimación se han ido desarrollando. Todas ellas utilizan como herramienta para interpretar la realidad el variograma, que viene siendo un indicador de la desestructuración espacial que presenta la variable regionalizada, considerando sólo 2 puntos en el espacio y la distancia entre ellos. No obstante, en la naturaleza se dan comportamientos estructurales, a pequeña y gran escala, que no logran ser interpretados al analizar sólo dos puntos en el espacio, tal es el caso de los meandros de los ríos, los paleocanales en el subsuelo, las columnas de sedimentos, los pliegues de estructuras, entre otros casos.

Por otro lado, es posible simular el comportamiento de esos fenómenos. Se entiende por simulación a la realización de una variable aleatoria regionalizada, $z(x)$, a través de una función aleatoria generadora, $Z(x)$. La simulación se diferencia respecto a la estimación en que esta última genera estimaciones cuyos valores presentan propiedades (momentos) distintos a las propiedades de los valores reales. En efecto, la estimación presenta las siguientes importantes limitantes [Emery, 2012]:

- Suaviza los valores reales: $dispersión_{estimada} < dispersión_{real}$.
- La variabilidad, a lo largo de la grilla estimada, no es uniforme. En efecto, en los extremos tiende a la media disminuyendo la variabilidad.
- No reproduce el histograma de los valores reales. Se pierden los valores extremos.

Por su parte, una única realización de la simulación lleva a reproducir espacialmente todas las propiedades (momentos) de la variable, en otras palabras, presenta una posible condición (realización) de la realidad. Varios algoritmos de simulación han sido desarrollados: bandas rotantes, secuencial gaussiana, gaussiana truncada, entre otras. Todas ellas, basadas en el uso del variograma.

2.1 Simulación estocástica multipuntos

Guardiano y Srivastava (1993), introducen por primera vez el concepto de la Estadística Multipuntos. En ella es posible incorporar las relación espacial que existe entre 3 o más puntos. En las herramientas de estimación basadas en el uso de variogramas es posible dar una interpretación a la función aleatoria $Z(x)$ que generó la realización del espacio de muestras desde donde se obtuvo el variograma. En el caso de la estadística multipuntos la función aleatoria (su función de distribución) debe ser generada a partir de una previa interpretación de la realidad: interpretación geológica, imágenes de entrenamiento, bases de datos de zonas análogas, etc. Guardiano y Srivastava (1993) proponen la simulación de valores a partir de la generación de la función aleatoria como inferencia probabilística al recorrer una imagen de entrenamiento a través de un patrón específico, no obstante en cada realización de la variable aleatoria en un punto en el espacio se tenía que recorrer la imagen de entrenamiento nuevamente. Strebelle (2000, 2002) implementa las ideas generadas por Guardiano y Srivastava en el algoritmo de simulación SNESIM (*single normal equation simulation*) que propone utilizar árboles de búsqueda con la finalidad de almacenar la información inferida de la imagen de entrenamiento disminuyendo el tiempo del proceso. El problema surge cuando la imagen de entrenamiento es de gran tamaño, ya que la demanda de almacenamiento aumenta considerablemente o bien la variable regionalizada es continua debiendo ser discretizada para su análisis.

Zhang, et. al. (2006) desarrollan una nueva técnica, denominada FILTERSIM (*filter-based simulation*) para almacenar la inferencia en el comportamiento de las estructuras a partir de las imágenes de entrenamiento. La técnica ocupada es el uso de filtros que recorren el patrón de búsqueda sobre la imagen de entrenamiento almacenando el valor para cada uno de los 3 primeros momentos en cada dirección del espacio (2D o 3D). De esta manera hay un ahorro sustancial en el almacenamiento de información, sin verse mayormente afectado por el tamaño de la imagen de entrenamiento. Su aplicación es posible tanto para variables discretas como continuas.

2.1.1 Algoritmo SNESIM

SNESIM (*Single Normal Equation Simulation*), fue desarrollado por Strebelle en el año 2002 y publicado en el paper: *Conditional Simulation of Complex Geological Structure Using Multiple-Point Statistics*. Se caracteriza por ser un algoritmo secuencial, es decir, trabaja sobre un secuencia de nodos predefinida. Tiene la capacidad de simular sobre variables tanto discretas como continuas, siempre y cuando estas últimas estén subdivididas en intervalos finitos. Dentro del proceso de simulación genera un árbol de condiciones, lo que lleva a Snesim a ser altamente demandante de memoria RAM y lento en su proceso global. Es posible caracterizar el algoritmo SNESIM tal como sigue:

- Simula sobre variables discretas o continuas subdivididas en K intervalos finitos.
- Simple de ser generado bajo algún lenguaje de programación.
- Presente en el software SGeMS.
- Demandante en recursos: mucha memoria RAM para almacenar las condiciones previa confección de la función distribución de probabilidad condicionada.
- Lento. Si la imagen de entrenamiento es grande entonces el proceso demanda mucho tiempo.
- Dificultad de incorporar la media local, ya que el algoritmo es rígido en estructura.

2.1.1.1 Terminología

Previo a una explicación detallada de la estructura del algoritmo, es necesario definir varios conceptos y terminologías [Strebelle, 2002]:

- Se denota por S una característica de la variable y por k los posibles estados de esta característica $\{S_k \text{ con } k = 1..K\}$, es decir, S_k es la característica S en el estado k .
- Se denota por d_n un evento condicionante de tamaño n centrado en el punto (nodo) u a ser simulado que considera:
 - Un sector geométrico (vecindad) de n direcciones desde el nodo central a un nodo vecino separado por una distancia h_α para $\alpha = 1 \dots n$. En otras palabras, se tiene el evento d_n cuando se consideran n vectores con origen en u y término en un dato existente a una distancia h_α . La figura 4 presenta un ejemplo ilustrativo de lo anterior.

- El valor a estimar $S(u)$ y los valores de los n datos que rodean a $S(u)$ son $S(u_\alpha)$ que se definen como $S(u_\alpha) = S(u+h_\alpha)$

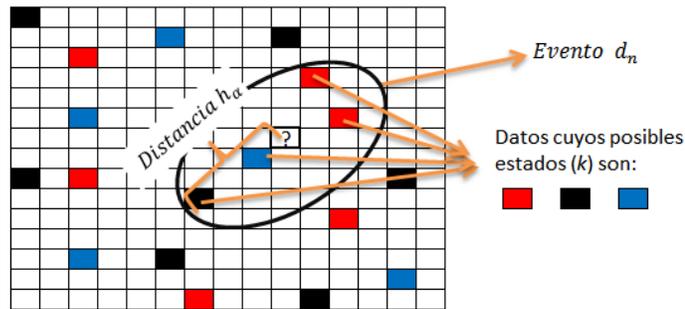


Figura 4 Ejemplo ilustrativo de un evento condicionante con 3 posibles estados y 4 datos condicionantes

2.1.1.2 Proceso de simulación

Utilizando la terminología anterior se describe a continuación el proceso que lleva a cabo SNESIM para desarrollar una realización.

1. Generación del árbol de búsqueda.
 - a. Se establece una vecindad móvil (que será la misma utilizada con los datos condicionantes más adelante), la cual recorrerá toda la imagen de entrenamiento caracterizando todos los nodos en su interior. La figura 5 presenta un ejemplo de confección de árbol de búsqueda en donde, en 4.a, una vecindad de 4 nodos recorre la imagen de entrenamiento para ir luego en 4.b almacenando la cantidad de veces que se presenta el valor Blanco (W) o Negro (B) en el centro de la vecindad. Los niveles de ramificación aumentan al ir incorporando la información de cada nodo. Con el árbol de condicionamiento generado es posible llevar la información a una función de distribución de probabilidad condicionada (*cpdf*, por sus siglas en inglés).

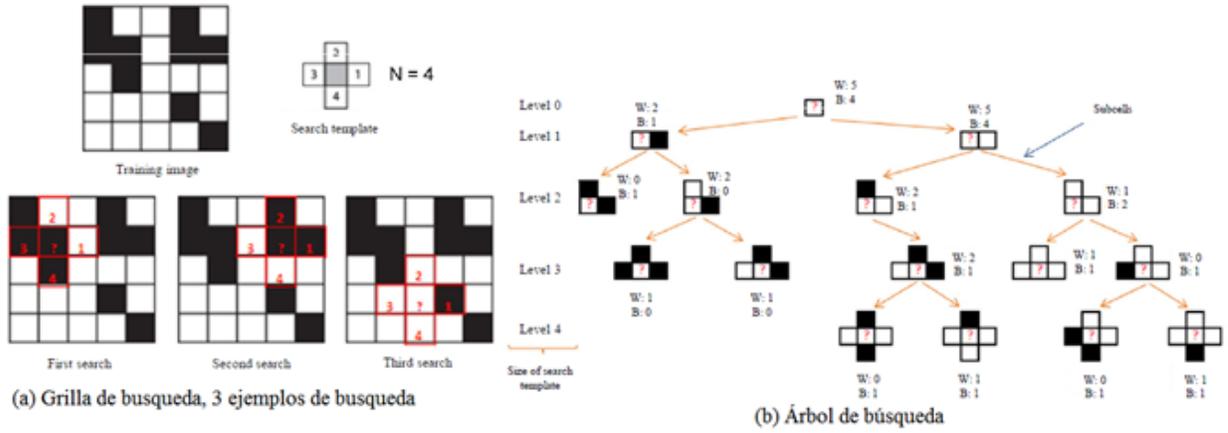


Figura 5 Ejemplo árbol de condicionamiento. Fuente: [Emery, 2012].

A continuación, utilizando la terminología antes expuesta se explica la realización de la *cpdf* a partir de la información entregada por la imagen de entrenamiento.

- b. A partir del árbol de búsqueda se establecen las frecuencias en que el nodo central toma el estado S_k dado un estado fijo para el resto de los nodos dentro de la vecindad. Las siguientes ecuaciones están basadas en la descripción realizada por Strebelle en 2002. La ecuación 1 presenta la analogía entre el concepto probabilístico de la ocurrencia del evento (izquierda) y la cuantificación desde el árbol de búsqueda (derecha)

$$Prob \{S(u) = s_k | d_n\} = f(u; k | d_n), \text{ con } k = 1, 2, \dots, K \quad (\text{Ec. 1})$$

SNESIM establece como indicadores binarios A_k y D tal como se aprecia en la ecuación 2 y 3 respectivamente:

$$A_k = \begin{cases} 1 & \text{si } S(u) = S_k \\ 0 & \text{si no} \end{cases} \quad (\text{Ec. 2})$$

A_k tomará el valor 1 sólo si el nodo central, ubicado en el punto u , está en el estado k y 0 para el resto.

$$D = \begin{cases} 1 & \text{si } S(u_\alpha) = S_{k_\alpha} \forall \alpha = 1 \dots n \\ 0 & \text{si no} \end{cases} \quad (\text{Ec. 3})$$

D tomará el valor de 1 cuando los nodos restantes en la vecindad tomen el estado k_a . Notar que no es necesario que todos los nodos restantes presenten el mismo estado, si no que cada nodo debe tener un solo estado dentro de los K posibles.

- c. Strebelle establece que para confeccionar la curva *cpdf* se tiene que las probabilidades de la ecuación 1 pueden ser obtenidas como en la ecuación 4:

$$f(u; k|d_n) = \text{Probabilidad}\{A_k = 1|D = 1\} \quad (\text{Ec. 4})$$

Más aun, esta última equivale a la relación probabilística expuesta en la ecuación 5 (Teorema de Bayes).

$$\frac{\text{Probabilidad}\{A_k = 1, D = 1\}}{\text{Probabilidad}\{D = 1\}} \quad (\text{Ec. 5})$$

Desde el árbol de búsqueda se cuantifica la cantidad de veces que los nodos (sin considerar el central) presentan una condición d_n en sus estados, denotándolo por $c(d_n)$. En la figura 5.b viene representado por la suma de W + B en cada extremo de ramificación. Sobre los casos anteriores, es decir, en cada extremo de una rama del árbol de búsqueda, se cuantifica la cantidad de veces que el nodo central presenta el estado k , denotándolo por $c_k(d_n)$.

Finalmente la probabilidad, para generar la *cpdf*, viene dada por:

$$f(u; k|d_n) = c_k(d_n)/c(d_n) \quad (\text{Ec. 6})$$

- d. Finalmente se obtiene la función *cpdf*, tal como en el siguiente ejemplo. A través de esta función y los datos duros condicionantes se procede a realizar el proceso de simulación asociado al valor desconocido en “u” el valor del estado “k” con mayor probabilidad dado el evento condicionante de los valores en el entorno.

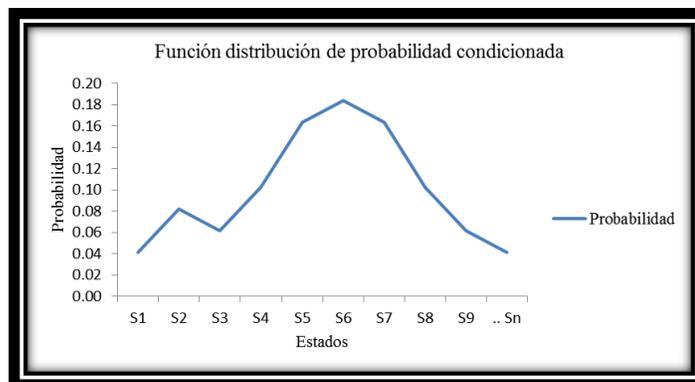


Figura 6 Ejemplo de función *cpdf*

2. Generación de una realización

- a. Se comienza por la generación de una grilla de simulación que contendrá en su interior un conjunto de datos condicionantes.
- b. Se asignan los datos condicionantes al nodo más cercano dentro de la grilla, tal como se aprecia en la figura 7.b.
- c. Se establece una ruta aleatoria que recorrerá todos los nodos sin información.
- d. Se establece una vecindad móvil, figura 7.a, que recorrerá la ruta previamente definida determinando los eventos condicionantes, figura 7.c. Si dentro de la vecindad no se encuentran nodos con información condicionantes se pasa el siguiente nodo de la ruta y luego se volverá a analizar el nodo anterior al término de la ruta.
- e. Al registrar un evento condicionante se utiliza la función *cpdf* para generar un evento aleatorio dentro de los posibles valores que el nodo central puede tener. El valor resultante es asignado al nodo que pasa de no tener información a tener un valor asignado. Este valor simulado es considerado un nodo condicionante para las siguientes búsquedas de eventos condicionantes. Ver ejemplo en figura 8.
- f. Se repite el proceso (e) sobre los nodos restantes de la ruta previamente definida. Si aún quedan nodos sin información, se reanalizan considerando los nodos con información simulada como información condicionante.
- g. Se realiza el proceso (f) hasta que toda la grilla de simulación queda completamente informada. La figura 9 es un ejemplo de realización completo. Importante recalcar que no se utilizó ninguna imagen de entrenamiento a pesar de que es un requisito indispensable para el proceso de simulación. Esto último dado que es un caso ficticio para la representación icónica de un proceso más complejo.

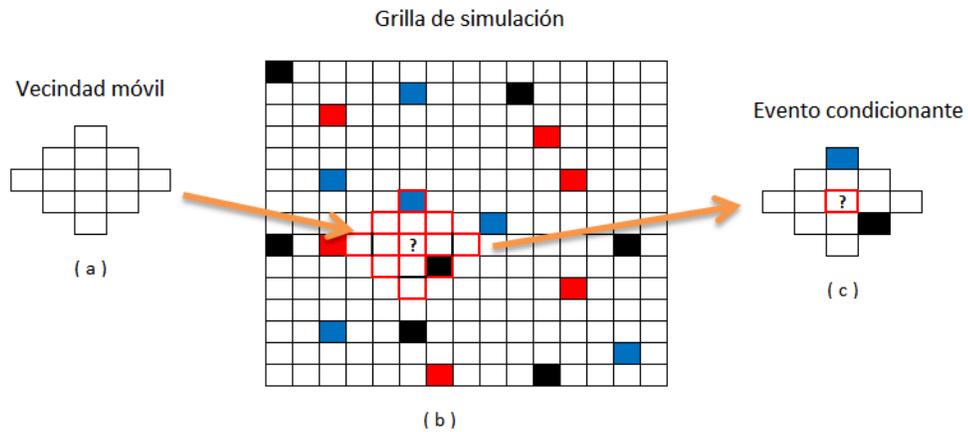


Figura 7 Ejemplo, determinación del evento condicionante. (a) Vecindad móvil, (b) grilla de simulación, (c) evento condicionante.

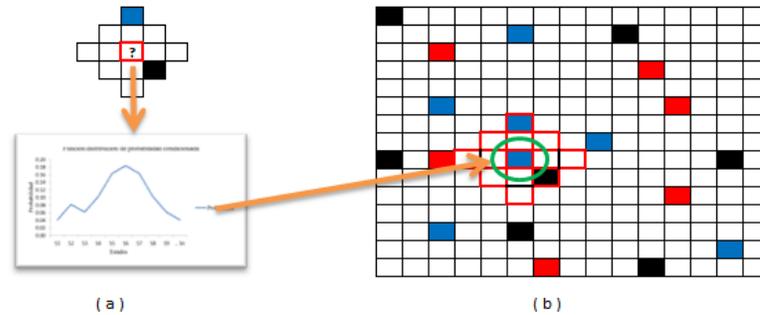


Figura 8 Determinación del valor asociado al nodo central. (a) Uso de la *pdf* dado el evento condicionante, (b) asignación del valor al nodo central.

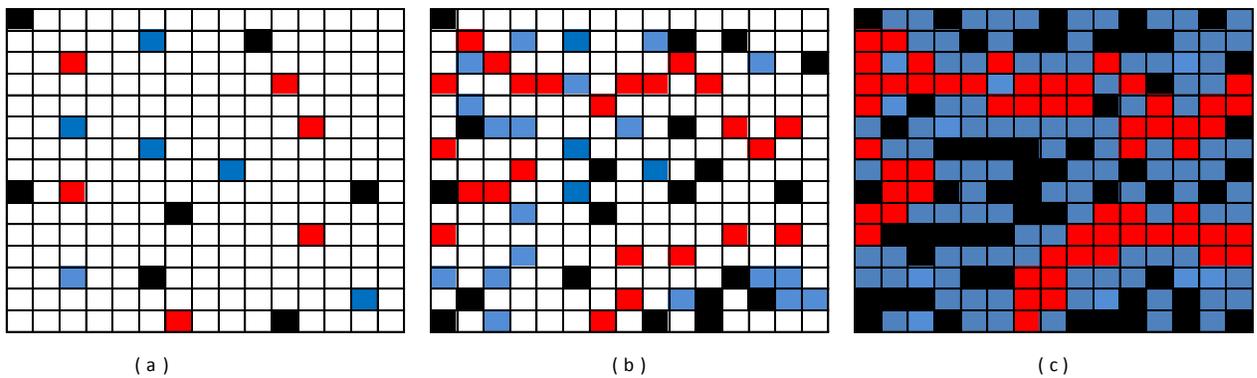


Figura 9 Ejemplo de una realización con SNESIM (caso ficticio). (a) Grilla a simular con datos condicionantes, (b) mitad del proceso de simulación, (c) grilla de simulación completa.

2.1.2 Algoritmo FILTERSIM

El algoritmo Filter-Based Simulation [Zhang et al., 2006], FILTERSIM fue desarrollado por Zhang en el año 2006 como tesis de doctorado y publicó el mismo año el paper *Filter-Based Classification of Training Image Patterns for Spatial Simulation*. FILTERSIM, como metodología para la simulación basada en estadísticas multipunto, presenta como características:

- Es aplicable tanto para variables categóricas (litologías, unidades geológicas, etc.) como continuas (leyes de mineralización, porosidades, densidad, concentraciones, etc.) por lo que permite su utilización en campos como la geología, evaluación de recursos mineros, petrología, etc.
- Reduce las dimensiones del problema. Si se considera, por ejemplo, la utilización de 9 filtros (3 filtros en las 3 direcciones principales) para una imagen de entrenamiento en 3 dimensiones bajo una grilla de 10x10x10 se reduce el patrón desde 1.000 valores (en cada nodo respectivo) a un vector de 9 valores correspondiente a cada filtro.
- Menor demanda de memoria RAM, relacionada directamente con el punto anterior al desarrollar el algoritmo de forma computacional.
- Proceso relativamente más veloz si se le compara con SNESIM sobre imágenes de entrenamiento de gran tamaño.

El proceso completo, de manera general, que es llevado a cabo al utilizar FILTERSIM es descrito a continuación:

1. FILTERSIM requiere de una imagen de entrenamiento (T.I.) que puede ser obtenida por una interpretación geológica, una fotografía de un sector o estructura similar, información geofísica, etc. lo suficientemente grande y completa para entregar los patrones de comportamiento estructural necesarios para una simulación de buena calidad.
2. Un conjunto de F filtros son aplicados sobre la T.I. bajo una grilla de búsqueda móvil, cuadrada en 2 dimensiones o cúbica en 3 dimensiones (se considerará la aplicación en 2 dimensiones, sin pérdida de generalidad para un caso en 3 dimensiones).
3. La aplicación de los F filtros sobre un mismo sector de la I.T llevan a caracterizar cada patrón local con un vector de F dimensiones. Ver figura 17.

4. El proceso anterior da lugar a una serie de F mapas de puntuación asociados cada uno a un filtro en específico, por lo tanto, se genera un espacio de puntuación (*score space*) de F -dimensiones.
5. Cada punto en el espacio de F -dimensiones antes mencionado representa a un patrón de entrenamiento específico. Cabe destacar que es posible encontrar puntos de igual valor que representen patrones de entrenamiento distintos. Se procede a reducir el espacio a través de un agrupamiento (*Clustering*) utilizando uno de los siguientes métodos: *K-mean partition* o *Cross partition*. La presente explicación seguirá utilizando el algoritmo *Cross partition*, no obstante, se detallan ambos métodos en la sección 2.2.3. *Cross partition* se basa en la subdivisión de los histogramas de frecuencia para cada dimensión de puntuación, en M intervalos ($2 \leq M \leq 10$) de igual frecuencia, independiente uno de otro.
6. Luego se obtiene un conjunto de M^F celdas, provenientes de la combinación de los M intervalos entre las F dimensiones, las cuales almacenan puntos del espacio de F dimensiones o bien pueden estar vacías.
7. Dentro de las celdas que no están vacías se procede a analizar la similitud de los patrones internos (representados por cada punto en el espacio dentro de la celda, o bien un vector de F dimensiones) además de cuantificar la cantidad de elementos en su interior con la finalidad de establecer el prototipo representativo de cada celda y los posibles prototipos hijos internos que representan a conjuntos de patrones de gran similitud. Ver sección 2.1.2.3 para más detalle.
8. Sobre la grilla a simular se genera sobre los nodos un camino aleatorio para el proceso de simulación. Sobre cada nodo en la ruta se establece el centro de la grilla de búsqueda (cuyo tamaño es igual a la utilizada con anterioridad en la generación de los patrones), se aplican los F filtros, y se obtiene un vector de F dimensiones (o un punto en el espacio de puntuación de F -dimensiones) denominado DEV_F . No es necesario contar con datos condicionantes ya que la primera simulación logra congelar datos simulados que irán completando y dando forma a las estructuras simuladas.
9. Se procede a buscar el prototipo más cercano a la condición presentada por DEV_F para luego elegir un patrón, dentro de la celda representada por el prototipo escogido, de manera aleatoria. Si se presentan subdivisiones, es decir, prototipos hijos dentro de un prototipo mayor se debe elegir el prototipo hijo más cercano y luego elegir el patrón de

forma aleatoria. Es el proceso aleatorio interno de escoger el patrón dentro del prototipo (o prototipo hijo) lo que le da el carácter estocástico al algoritmo.

10. Existe una grilla interna (*inner path*), centrada en el mismo punto que la grilla de búsqueda, que establece que valores del patrón escogido serán congelados dentro de la grilla a simular y utilizados como información secundaria hasta completar el proceso completo de simulación.

A continuación se explican con más detalles los procesos antes mencionados:

2.1.2.1 *Imagen de entrenamiento*

Por imagen de entrenamiento (T.I.) se entiende al conjunto de patrones representativos de una o más estructuras tanto categóricas como continuas en un espacio de dos o tres dimensiones. Los algoritmos de simulación multipuntos (más de 2 puntos en el espacio), en particular FILTERSIM, requieren de una T.I. desde la cual se obtendrán los conjuntos de patrones representativos del comportamiento espacial de las estructuras. Cabe destacar que la T.I. debe lograr caracterizar el comportamiento tanto local como global de las estructuras.

Las T.I. tienen distintos orígenes. Estas pueden ser creadas por la interpretación geológica de una zona, pueden ser la fotografía (interpretada) de un corte real en un ‘set estructural’, el modelo en 2D de un sector ya conocido para interpretar otro desconocido, los resultados de un mapeo geofísico en profundidad, etc. Tomando en consideración que el input necesario para la simulación con FILTERSIM es la T.I. se interpreta que las principales aplicaciones de este algoritmo están ligadas con la evaluación de recursos (mineros, forestales, etc.), la industria petrolífera, la geología estructural, etc. La figura 10 presenta 3 tipos de T.I., en 2 dimensiones, cuyas variables son binarias, categóricas y continuas.

Un patrón es un sector de la imagen de entrenamiento bajo una grilla de búsqueda específica. Sea una T.I. de 20 x 20 y una grilla de búsqueda móvil de 7 x 7 es posible encontrar 196 patrones al desplazar la grilla por toda la T.I. La figura 11 da cuenta de 2 patrones obtenidos para el ejemplo del caso anterior.

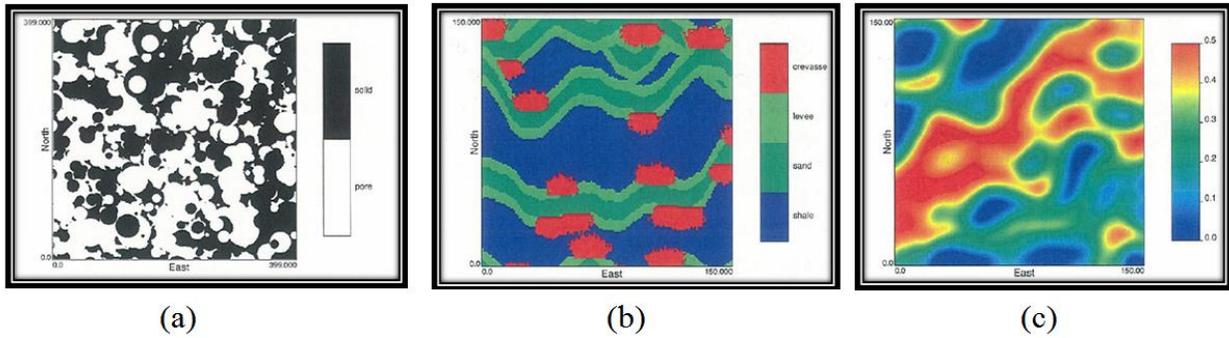


Figura 10 Imágenes de entrenamiento. (a) T.I. de una variable binaria – porosidad media. (b) T.I. de una variable categórica – Depositación fluvial con cuatro categorías. (c) T.I. de una variable continua – Distribución de la porosidad. Fuente: Filter-Based training pattern classification for spatial pattern simulation (Zhang et al., 2006)

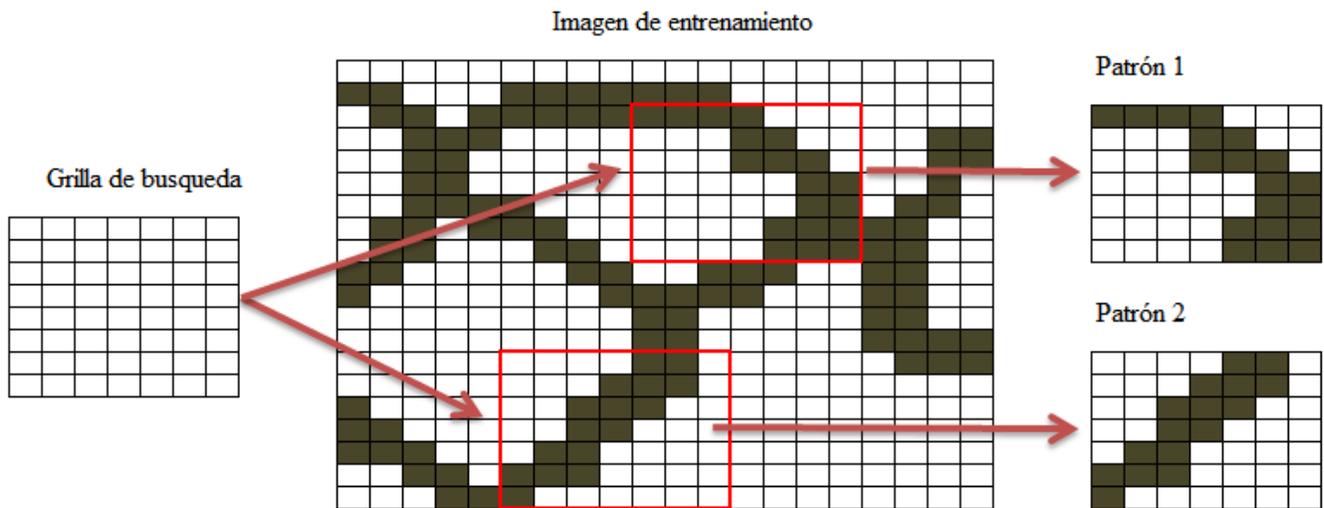


Figura 11 Obtención de patrones a partir de una grilla de búsqueda desde una imagen de entrenamiento

2.1.2.2 Clasificación de patrones

Tal como se mencionó anteriormente, para la obtención de los patrones desde la T.I. es necesario definir una grilla de búsqueda móvil que recorrerá toda la imagen de entrenamiento. Cabe destacar que los contornos de la T.I. no logran ser caracterizados ya que es necesario que toda la grilla de búsqueda esté dentro de los márgenes de la T.I., a la sección interna que logra ser analizada se le denomina imagen erodada, este concepto se analiza con más detalle en la sección 2.1.2.2.2.

2.1.2.2.1 Grilla de búsqueda

Por grilla de búsqueda se entiende al espacio, en dos o tres dimensiones, con que se irán seleccionando los patrones provenientes de la imagen de entrenamiento. Si se considera una grilla pequeña los patrones serán representativos de los bordes de las estructuras y los comportamientos a pequeña escala, mientras que de ser grande se logrará capturar el comportamiento a nivel global.

Es necesario que los lados de la grilla, independiente en la dimensión del problema que se esté trabajando, sean de tamaño impar ya que garantiza la existencia de un punto medio al interior de la grilla que caracterizará al patrón como punto en el espacio de puntuación (score space) (ver sección 2.2.2.3). En el ejemplo de la figura 11 se utilizó una grilla de 49 subdivisiones que provienen de lados $(2n + 1)$ de largo y $(2n + 1)$ de ancho, con $n = 3$.

2.1.2.2.2 Filtros

El algoritmo presenta 3 filtros predefinidos, con los cuales se aprecia la media, las gradientes direccionales y las curvaturas de los datos al interior de las grillas de búsqueda. Si los filtros son aplicados sobre un caso 2D ó 3D se obtienen 6 y 9 dimensiones respectivamente.

Las figuras 12, 13 y 14 (desarrolladas en el software U-Mine) presentan de forma gráfica el valor de cada nodo para cada filtro en cada dirección. Cabe destacar que los valores de los nodos sólo se modifican en una dirección mientras que para el resto (1 dirección restante en 2D o bien 2 direcciones restantes en 3D) se mantiene el mismo valor. Sean u , v y w los parámetros que recorren las direcciones x , y y z respectivamente, y una grilla de tamaño $(2n + 1) \times (2n + 1) \times (2n + 1)$, los filtros quedan definidos como:

$$\text{Primer filtro: promedio en dirección } N - S, \quad f_1: 1 - \frac{|v|}{n}, v = -n, \dots, n$$

$$\text{Segundo filtro: promedio en dirección } E - O, \quad f_2: 1 - \frac{|u|}{n}, u = -n, \dots, n$$

$$\text{Tercer filtro: promedio en la elevación,} \quad f_3: 1 - \frac{|w|}{n}, w = -n, \dots, n$$

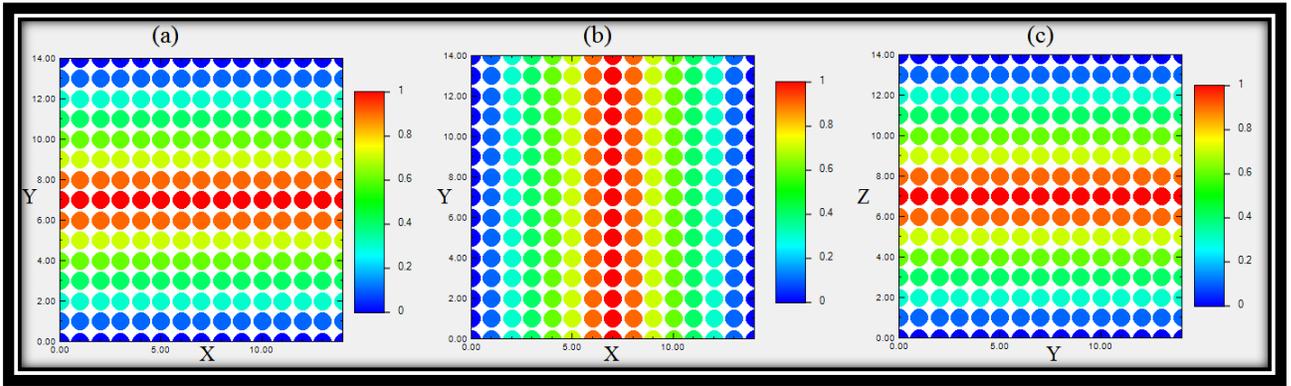


Figura 12 Filtros asociados a la media. (a) Norte-Sur, (b) Este-Oeste, (c) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

Cuarto filtro: gradiente direccional N – S, $f_4: \frac{v}{n}, v = -n, \dots, n$

Quinto filtro: gradiente direccional E – O, $f_5: \frac{u}{n}, u = -n, \dots, n$

Sexto filtro: gradiente direccional en la elevación, $f_6: \frac{w}{n}, w = -n, \dots, n$

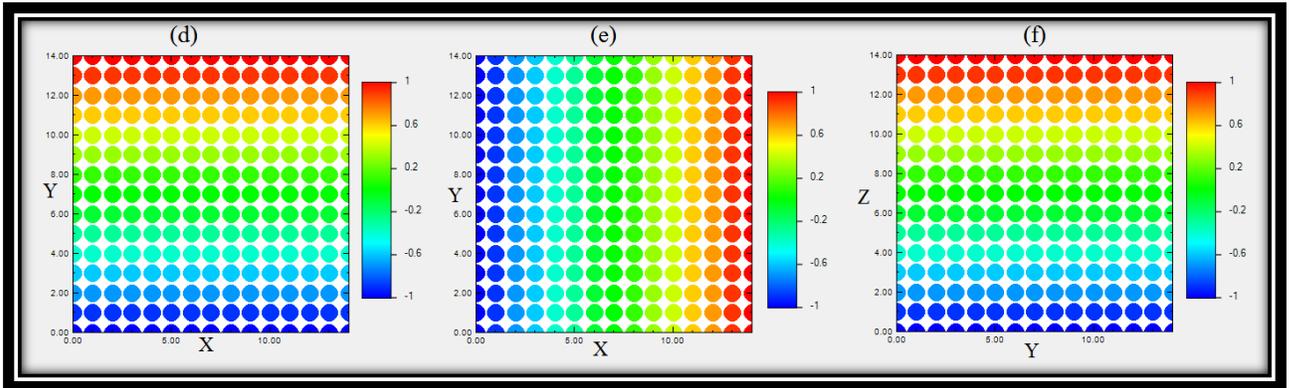


Figura 13 Filtros asociados al gradiente. (d) Norte-Sur, (e) Este-Oeste, (f) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

Septimo filtro: curvatura direccional N – S, $f_7: \frac{2|v|}{n} - 1, v = -n, \dots, n$

Octavo filtro: curvatura direccional E – O, $f_8: \frac{2|u|}{n} - 1, u = -n, \dots, n$

Noveno filtro: curvatura en la elevación, $f_9: \frac{2|w|}{n} - 1, w = -n, \dots, n$

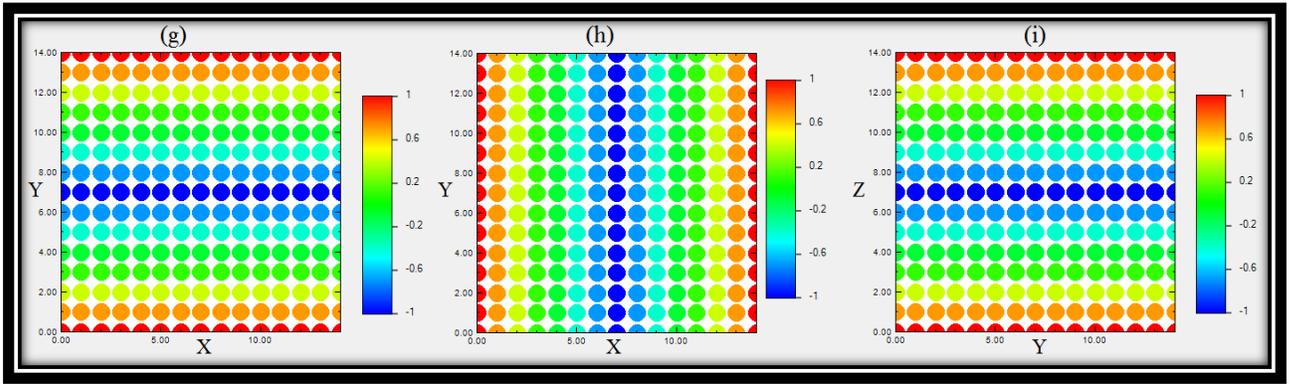


Figura 14 Filtros asociados a la curvatura. (g) Norte-Sur, (h) Este-Oeste, (i) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

A continuación se presenta la aplicación de los filtros dentro de una grilla de búsqueda móvil sobre una T.I. en 2 dimensiones [Zhang, 2006] sin pérdida de generalidad para el caso en 3 dimensiones. Sea $X(i, j)$ el valor del dato en la posición (i, j) dentro de la T.I., asumiendo que todos los puntos al interior de la T.I. están informados. Sea la grilla cuadrada de tamaño $(2n + 1) \times (2n + 1)$. Se tiene entonces que el puntaje $S_f(i, j)$ asociado al punto (i, j) bajo el filtro f viene dado por:

$$S_f(i, j) = \sum_{v=-n}^n \sum_{u=-n}^n f(u, v) * X(i + u, j + v) \quad (\text{Ec. 7})$$

Donde $f(u, v)$ representa uno de los 6 filtros predefinidos o algún otro filtro generado por el usuario. La figura 15 presenta como ejemplo la aplicación del segundo filtro sobre el patrón 2 que se obtuvo en la figura 11, en donde se obtiene un patrón binario (color blanco igual a 0 y negro igual a 1). El puntaje obtenido es de **0.15**. Este proceso se repite con los 5 filtros restantes sobre el mismo patrón lo que lleva a definir un punto en el espacio de 6 dimensiones que caracteriza al patrón 2. Cabe mencionar que distintos patrones pueden ubicarse en el mismo punto espacial ya que la ponderación entregada por la ecuación 7 puede ser idéntica para distintas combinaciones de datos.

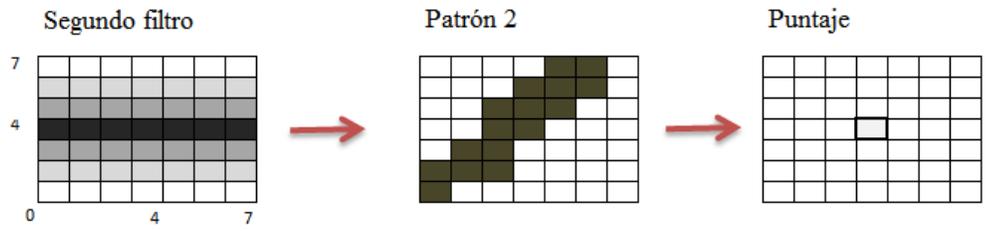


Figura 15 Caracterización de un patrón, asignación de puntaje, al aplicar un filtro en específico

Al recorrer toda la imagen de entrenamiento utilizando un mismo filtro, es posible obtener la puntuación $S_f(i, j)$ para todos los puntos que están en su interior, en los cuales al centrar la grilla de búsqueda ésta queda completamente al interior de la T.I. Esto último significa que si se tiene una grilla de búsqueda de $(2n + 1) \times (2n + 1)$ y una imagen de entrenamiento de tamaño $(p) \times (q)$ entonces el tamaño del mapa con los valores de la puntuación obtenida tendrá un tamaño de $(p - 2n) \times (q - 2n)$. Tal como se dijo en secciones anteriores, se le denomina imagen de entrenamiento erodada a la sección de $(p - 2n) \times (q - 2n)$ obtenida desde la imagen de entrenamiento siendo ambos centros coincidentes. La figura 16 presenta la imagen erodada del ejemplo expuesto en la figura 11.

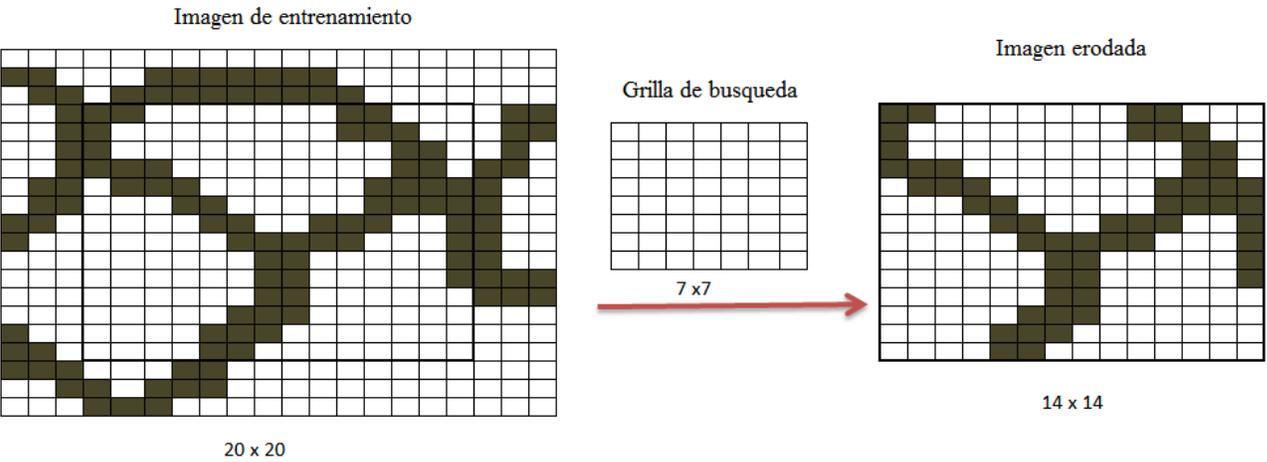


Figura 16 Ilustración del concepto de imagen erodada

Es posible encontrar un ejemplo de aplicación de filtros y mapas de puntuación en *Filter-Based Classification of Training Image Patterns for Spatial Simulation* [Zhang et al., 2006].

2.1.2.2.3 Generación del espacio de puntuación (score space)

Sea F la cantidad de filtros que se han de utilizar sobre una T.I. Se obtienen F mapas de puntuación en los cuales cada punto en su interior caracteriza a un cierto patrón bajo el filtro correspondiente. Cada mapa de puntuación presenta un rango en los cuales toman valores sus puntos internos (pueden ir de $[0,1]$, $[-1,1]$ o bien como lo defina el usuario). Por lo tanto, si se utiliza un eje para ubicar cada punto del mapa de puntuación y se repite el proceso para los F mapas distintos se genera un espacio de F dimensiones. Para facilitar la comprensión se utilizan vectores de $F \times 1$ en los cuales cada elemento representa el valor en el mapa de puntuación. La figura 17 presenta un esquema de lo antes mencionado.

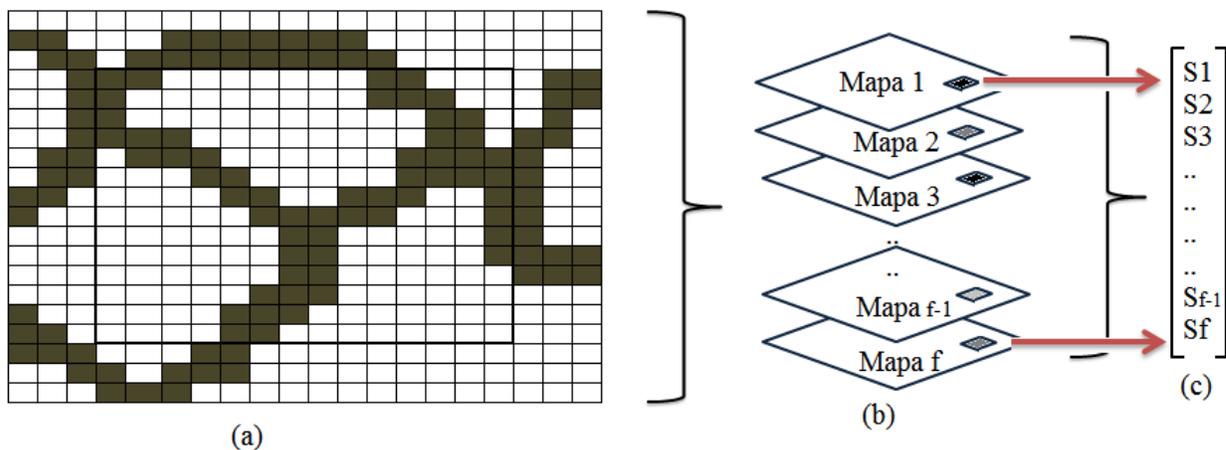


Figura 17 Ubicación de un patrón en el espacio de puntuación de F dimensiones. (a) Imagen de entrenamiento, (b) F mapas de puntuación, (c) vector representativo de la ubicación de la puntuación del patrón en el espacio de F dimensiones.

2.1.2.3 División del espacio de puntuación

Al obtener todos los posibles patrones, varios suelen ser muy similares de aspecto lo que lleva a que tengan puntuaciones similares, es decir, que se ubiquen muy cercanamente en el espacio de puntuación. Resulta provechoso generar un agrupamiento (*Clustering*, en inglés) de los patrones similares con la finalidad de que el proceso de simulación sea veloz. Para ello se proponen 2 herramientas para de generar el *Clustering*: a través de *cross partition* [Zhang et al., 2006] o bien *k-mean partition* [Hartigan, 1975]. A continuación se detallan ambos procesos de agrupamiento.

2.1.2.3.1 Cross partition

Este algoritmo de agrupamiento se basa en la generación de celdas al dividir en espacios de igual frecuencia las curvas de distribución para cada mapa de puntuación (o eje en el espacio de F dimensiones). Consta de las siguientes 3 etapas [Zhang et al., 2006]:

1. Cada mapa de puntuación presenta un rango de valores asociados a las puntuaciones capturadas sobre cada patrón. Sea M un valor entero en el rango $2 \leq M \leq 10$, se divide cada rango de valores en M intervalos de igual frecuencia. Sea F la cantidad de filtros utilizados, esta división se realiza para cada rango de valores de manera independiente unos de otros. Al realizar este proceso, se logra dividir el espacio de F dimensiones en M^F celdas de las cuales las que están vacías se descartan y no son consideradas en los siguientes procesos, mientras las que presentan patrones en su interior son denominadas clases de puntuación. Cada clase almacena uno o más patrones en su interior, luego es posible identificar la clase a través de un “prototipo” el cual corresponda al promedio punto a punto de todos los patrones al interior de la clase. La figura 18 presenta un ejemplo ilustrativo de 6 patrones similares pertenecientes a la misma clase y el prototipo representativo e identificador de la misma.

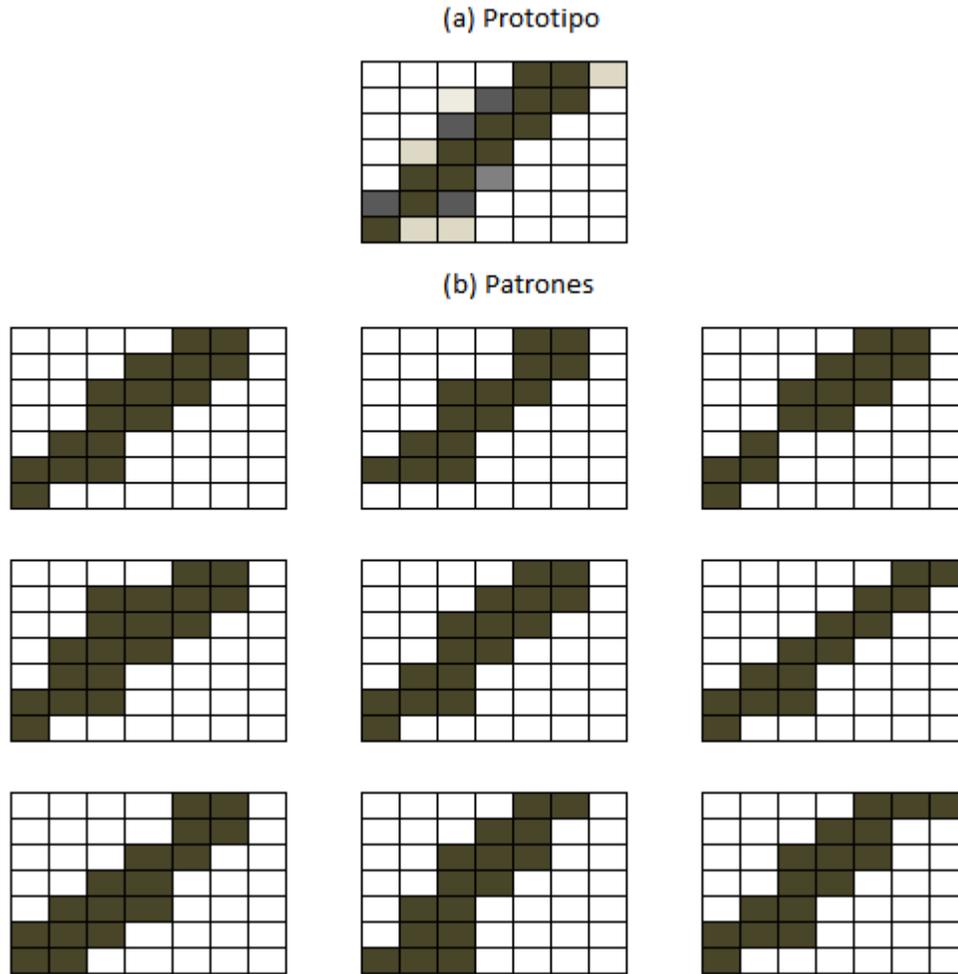


Figura 18 Ejemplo de una clase. (a) Prototipo que identifica a la clase, (b) patrones similares dentro de una misma clase.

2. Luego de obtener todas las clases no vacías al dividir el espacio de puntuación es necesario cuantificar la cantidad y la calidad en similitud de los patrones internos de cada clase. Para ello, Zhang propone el índice de ajuste SI (por sus siglas en inglés de *Sharpness Index*) [Zhang et al., 2006] para las variables continuas:

$$SI = 1 - \frac{\left(\frac{1}{n} \sum_{j=1}^n \sigma_j^2\right)}{\sigma^2} \quad (Ec. 8)$$

En donde, n es el tamaño de los patrones (o equivalentemente, el tamaño de la grilla de búsqueda), σ_j^2 es la varianza del valor en la posición j de la grilla obtenida entre todos los patrones dentro de la clase, y σ^2 es la varianza entre todos los patrones y todas las ubicaciones al interior de cada uno de ellos.

La expresión para las variables categóricas o binarias no son descritas en esta sección ya que es similar y se prioriza la explicación detallada de las variables continuas dado que el caso de estudio expuesto en las próximas secciones presenta variables continuas. No obstante, para más detalle en casos categóricos o binarios ver *Filter-based training pattern classification for spatial pattern simulation*. Zhang et al., 2006. Pág, 16.

El índice SI , ecuación 8, puede tomar valores en el intervalo $[0,1]$ considerando que al ser 1 todos los patrones internos son idénticos entre sí, y al ser 0 se presenta el caso más desigual entre todos los patrones. Para restringir la cantidad de patrones al interior de cada clase se fijan los límites C_{max} y C_{min} como las cantidades máximas y mínimas de patrones al interior de una clase.

3. El proceso anterior se debe repetir hasta que cada clase tiene asignado su respectivo índice SI . Es posible que el SI sea bastante bajo o bien que la clase presente un gran número de patrones ($\gg C_{max}$). Cuando esto ocurre se debe subdividir la clase en “clases hijas” las cuales serán representadas por un “patrón hijo” que vendrá siendo el promedio de los patrones dentro de su respectiva “clase hija”. Para ello se debe proceder como sigue:
 - a. Si una clase presenta una cantidad de patrones inferior a C_{max} , entonces no se subdivide.
 - b. Si una clase presenta una cantidad de patrones superior a C_{max} y un $SI \geq 0.95$, entonces no se subdivide.
 - c. En caso contrario se subdivide la clase en 2 de la siguiente forma:
 - i. Para los rangos que entrega cada filtro, se propone dividir este por la mitad utilizando la mediana de los valores.
 - ii. Calcular el índice SI para ambas subclases.
 - iii. Mantener la subdivisión que entregue los más altos índices de SI para ambas subclases.

Los últimos procesos, de la (a) a la (c) deben ser aplicados sobre todas las clases hasta que no sea necesaria su subdivisión.

2.1.2.3.2 K-mean partition

El algoritmo de agrupamiento *K-mean*, propuesto por Hartigan en el año 1975 se puede aplicar en la generación de clases de patrones en FILTERSIM. El método consiste en que, dado un número fijo de grupos en el espacio de puntuación de F dimensiones, este entrega el centroide de cada grupo optimizando el siguiente criterio: cada elemento (patrón en este caso) debe pertenecer al grupo cuyo centroide presenta la menor distancia relativa respecto a los demás. La figura 19 es un ejemplo en el cual se cuenta con 8 patrones representados por círculos naranjos y 3 centroides (C1, C2 y C3); las líneas continuas simbolizan al conjunto de puntos equidistantes de cada par de centroides más cercanos.

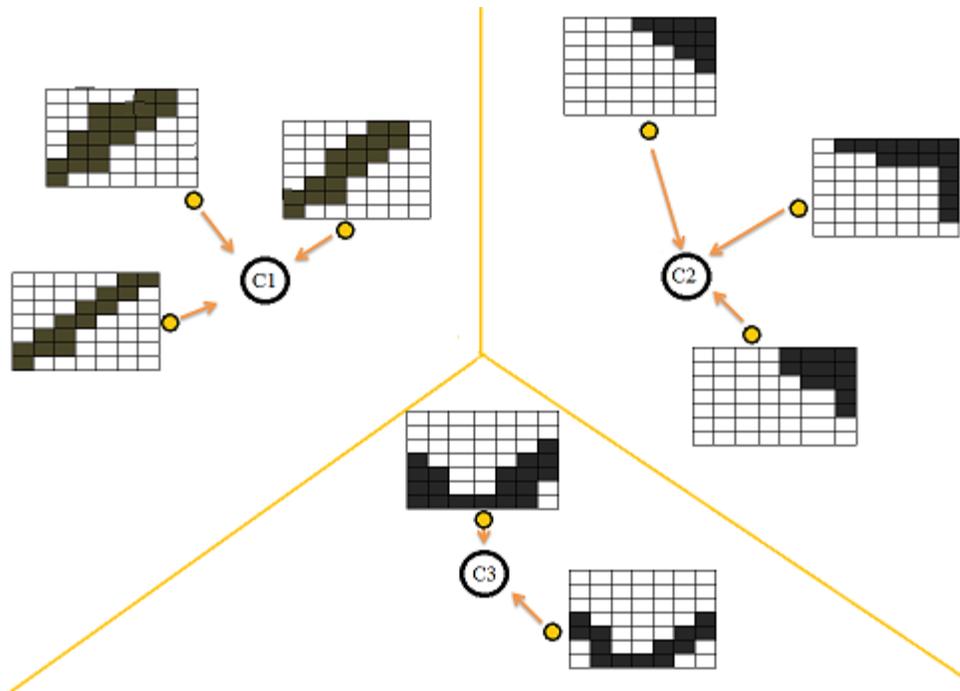


Figura 19 Ejemplo del proceso de agrupamiento utilizando el método *K-mean*. Se observan 3 centroides C1, C2 y C3 que llevan a generar las 3 líneas continuas que establecen los puntos equidistantes a los centroides más cercanos.

Con esto se logra obtener grupos cuya cantidad de elementos es relativamente similar. Cabe destacar que no existe una única solución (forma de agrupar los elementos) sino más bien distintos óptimos locales. A continuación se explica brevemente los pasos que realiza el algoritmo [Izenman, 2008]:

1. Sean K la cantidad de grupos a generar y X_i la posición en el espacio para n elementos a ser agrupados.
2. Se debe iniciar el proceso realizando uno de los siguientes pasos:
 - a. Ubicar en el espacio los K centroides, \bar{x}_k con $k = 1, \dots, K$.
 - b. Aleatoriamente ubicar todos los elementos del espacio en K grupos distintos y calcular cuál es el centroide de cada grupo, \bar{x}_k con $k = 1, \dots, K$.
3. Teniendo definido los centroides se calcula el cuadrado de la distancia euclidiana entre cada elemento del grupo y su centroide, entre todos los grupos. A esta métrica se le denomina ESS (*Squared-Euclidean Distance*) y viene dada por la ecuación 9:

$$ESS = \sum_{k=1}^K \sum_{c(i)} (x_i - \bar{x}_k)^2 \quad (Ec. 9)$$

Donde $c(i)$ representa el grupo i .

4. Calcular el valor actual de ESS, reubicar elementos de un grupo en otro con la finalidad de reducir el valor de ESS. Realizar esta reasignación de elementos hasta que no sea posible minimizar más el índice ESS.

Analizando las ventajas y desventajas de ambos algoritmos [Austin, 2007] se tiene el resumen de la tabla 1:

Tabla 1 Ventajas y desventajas de los algoritmos de agrupamiento Cross y K-mean [Austin, 2007]

Algoritmo de agrupamiento	Ventajas	Desventajas
Cross partition	Sencillo y muy eficiente desde un punto de vista de velocidad (CPU) y almacenaje (RAM)	La cantidad de elementos al interior de cada clase no logra ser bien unificada, sin considerar que muchas de las celdas quedan vacías
K-mean partition	Mejor metodología de agrupación de elementos y se logran obtener clases cuyas cardinalidad (cantidad de elementos) es menos variable	Altamente demandante de velocidad (CPU)

2.1.2.4 Proceso de simulación

Una vez obtenida la clasificación de los patrones inferidos desde la imagen de entrenamiento, es posible dar inicio al proceso de simulación. Sea L la cantidad de prototipos obtenidos en la clasificación, cada uno se representa como $prot^{(l)}$, con $l = 1, \dots, L$. El proceso consiste en [Zhang et al., 2006]:

1. Generación de una grilla de simulación. En ella se ubican los datos duros (muestras) condicionantes, y la información es almacenada en el nodo más cercano.
2. Un camino sobre todos los nodos de la grilla es generado de manera aleatoria.
3. Sobre cada nodo del camino se extrae el evento condicionante, d_τ , a través de una grilla de búsqueda, τ , de tamaño n . Cabe destacar que el tamaño n debe ser idéntico al utilizando para la inferencia desde la imagen de entrenamiento.
4. Se calcula n' que representa la cantidad de nodos informados dentro del evento condicionante d_τ .
5. Comenzando en el primer nodo del camino aleatorio se verifica que el nodo central, u_o , no sea un dato duro condicionante o bien un dato previamente simulado.
6. Si no es así se sigue con el paso 7, de lo contrario se pasa al siguiente nodo y se vuelve a verificar.
7. Se verifica el valor de n' , es decir, la cantidad de nodos informados a lo largo de toda la grilla de simulación.
8. Si se obtiene que el valor de n' es igual a cero a lo largo de toda la grilla de simulación, es decir, que no existen datos condicionantes entonces se elige un patrón de manera aleatoria dentro de todos los prototipos y se centra en la grilla de búsqueda congelando los valores internos como datos simulados.
9. Si se tiene que $n' > 0$ para al menos una grilla de búsqueda, es decir, que existe al menos un evento condicionante d_τ , entonces se debe encontrar el prototipo más cercano al evento condicionante a través de una función distancia (*template pixels* o *filter scores*). Estas funciones de distancia consideran solamente los nodos informados dentro de la grilla de búsqueda.
10. Cuando el prototipo cercano es establecido, se elige de manera aleatoria un patrón representativo de la clase.

11. El patrón elegido es centrado en la grilla de búsqueda y los valores que están dentro de la grilla interna de congelamiento son congelados sobre los nodos de la grilla de simulación. Cabe destacar que si dentro de la grilla interna de congelamiento hay presencia de nodos con datos duros condicionantes o datos previamente simulados estos no son reemplazados por los nuevos valores. Finalmente, la grilla de búsqueda se desplaza a otro nodo a simular.
12. Repetir los pasos del 5 al 11 hasta que los nodos en la grilla de simulación estén completamente informados.

Las funciones distancia para el cálculo del prototipo más cercano son *template pixels* o *filter scores*. Ambas son disyuntivas dentro del algoritmo FILTERSIM, es decir, se debe optar por una u otra. A continuación se presenta la metodología implementada por cada una:

2.1.2.4.1 Método para el cálculo de distancia: *template pixels*

La distancia es calculada como la suma de las diferencias pixel a pixel entre los datos del evento condicionante y del prototipo de cada clase. Sean:

- L la cantidad de prototipos obtenidos en la clasificación, cada uno se representa como $prot^{(l)}$, con $l = 1, \dots, L$.
- Tipos de datos según información: $d = 1, 2$ y 3 , para los datos duros condicionantes, datos previamente simulados y nodos con información suave (información condicional secundaria) respectivamente.
- u_o el nodo central de la grilla de búsqueda.
- $w(d)$ el peso asociado al dato del tipo d . Se debe cumplir que: $w(1) \geq w(2) \geq w(3)$ y $w(1) + w(2) + w(3) = 1$ con la finalidad de dar más relevancia a la información proveniente de los datos duros condicionantes (muestras).
- u , es el vector distancia entre el nodo central, u_o , y el nodo con información dentro de la grilla de búsqueda.
- dat_d , es el valor del nodo en el evento condicionante, cuyo dato es del tipo d .

Con lo anterior, la función distancia queda definida en la ecuación 10:

$$D_l(u_o) = \sum_{d=1}^3 w(d) \text{Average} | \text{dat}_d(u_o + u) - \text{prot}^{(l)}(u) | \quad (\text{Ec. 10})$$

2.1.2.4.2 Método para el cálculo de distancia: *filter scores*

La distancia es calculada como la suma de las diferencias entre los puntajes, asignados por cada uno de los filtros, entre el prototipo y el evento condicionante. La función distancia viene dada por la ecuación 11:

$$D_l(u_o) = \sum_{i=1}^F | S_{\text{dat}}^{(i)}(u_o) - S_{\text{prot}^{(l)}}^{(i)}(u) | \quad (\text{Ec. 11})$$

Dónde:

- F , es el número de filtros.
- $S_{\text{dat}}^{(i)}$ es el valor del filtro i sobre el evento condicionante.
- $S_{\text{prot}^{(l)}}^{(i)}$ es el valor del filtro i sobre el prototipo l .
- u_o y u representan el nodo central del evento condicionante y del prototipo respectivamente.

Cabe destacar, para ambos métodos, que no todos los nodos en el evento condicionante están informados, por lo tanto, los filtros sobre el evento condicionante sólo ponderarán los nodos con información, o dicho de otra forma, los nodos sin información valdrán cero. Para no tener sesgo en la determinación del patrón más cercano, los filtros sobre los prototipos deben considerar solo los nodos en el prototipo que se ubican en la misma posición que los nodos con información en el evento condicionante

2.1.2.5 Grillas múltiples

Para capturar los comportamientos a gran escala de las estructuras se utiliza el concepto de multigrillas [Zhang et al., 2006] dentro del proceso de simulación. Se asume una grilla de búsqueda, τ , de tamaño n la cual será re-escalada una cantidad G de veces. Cada grilla de búsqueda re-escalada, tendrá una cantidad n nodos, no obstante la ubicación espacial de cada uno de ellos será distinta. Sean:

- u_0 la ubicación del nodo central de la grilla de búsqueda.
- h_j con $j = 1 \dots n$ el vector distancia entre el nodo central y el j -ésimo nodo dentro de la grilla de búsqueda.

Se tiene entonces que la g -ésima grilla de búsqueda, τ^g , tiene sus n nodos ubicados en las posiciones $u_j = u_0 + 2^{(g-1)} \cdot h_j$, es decir, cada aumento en el tamaño implica que las distancias entre nodos se van duplicando. La figura 20 ilustra 3 grillas múltiples anidadas de tamaño 5×5 .

La grilla de menor tamaño se alcanza cuando $g = 1$ mientras que la de mayor tamaño cuando $g = G$. Cabe mencionar que el número de nodos es igual en ambos casos, solo cambia la distancia entre ellos. Los procesos de simulación descritos en la sección 2.1.2.4. son aplicados para cada una de las g -ésimas grillas anidadas, incluyendo con esto el proceso de *clustering* de los patrones inferidos desde la imagen de entrenamiento utilizando exactamente la misma grilla de búsqueda utilizada en la g -ésima grilla anidada. Se debe comenzar el proceso de simulación desde la grilla más gruesa a la más fina. Los nodos simulados sobre la g -ésima grilla son considerados datos duros para la $(g-1)$ -ésima grilla, con la salvedad que estos nodos deben ser re-simulados.

Cabe destacar la posibilidad de utilizar distintas imágenes de entrenamiento y/o filtros para cada una de las grillas anidadas, ya que estas deben generar el proceso de *clustering* de patrones cada vez que cambia la grilla de búsqueda.

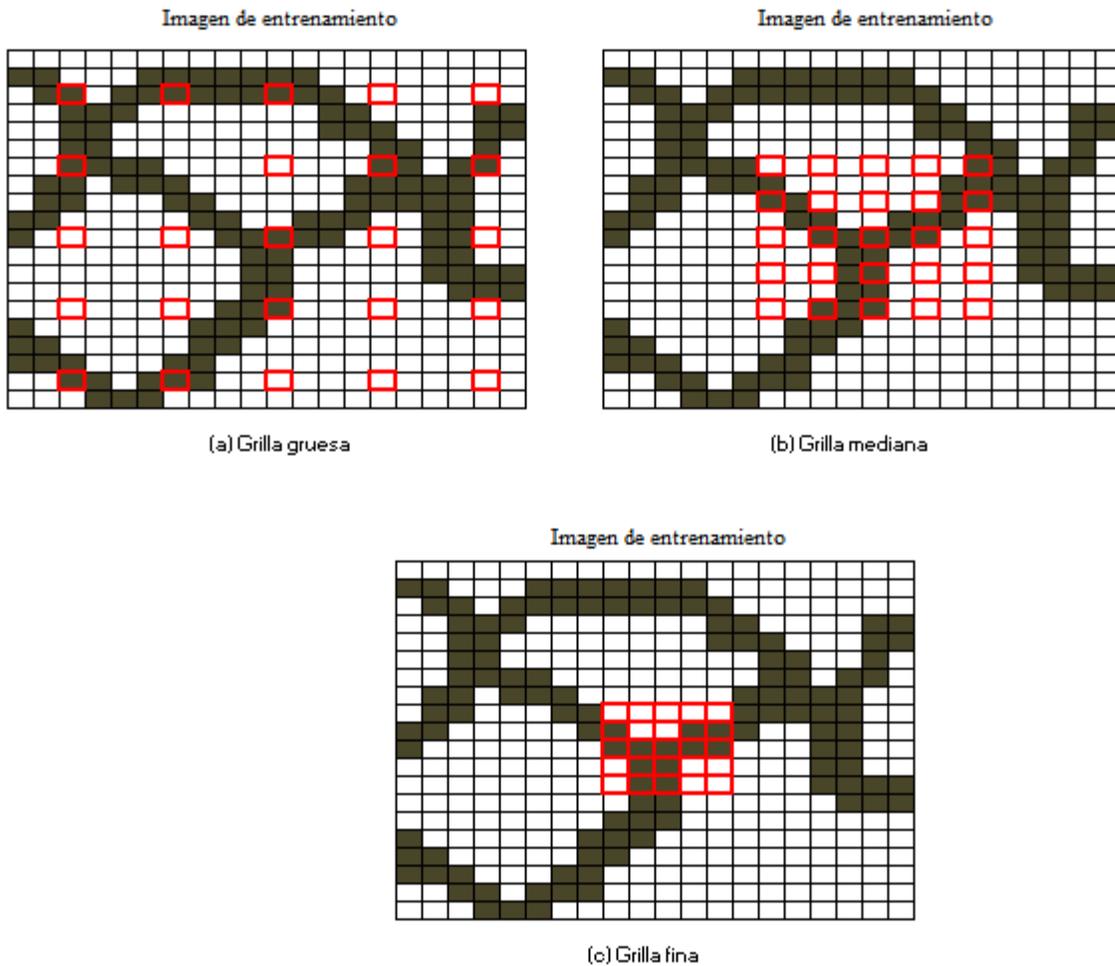


Figura 20 Ilustración de 3 multigrillas de tamaño 5x5 sobre una imagen de entrenamiento

2.1.3 Elección de algoritmo a utilizar

Finalmente, se opta por utilizar el algoritmo FILTERSIM sobre el algoritmo SNESIM principalmente por las siguientes razones:

1. FILTERSIM simula sobre variables continuas (para el desarrollo del caso de estudio real) sin necesidad de generar subdivisiones en intervalos como SNESIM.
2. FILTERSIM trabaja de manera más veloz al utilizar una T.I. de gran tamaño.
3. FILTERSIM reduce la dimensión del problema ahorrando memoria RAM.

Por lo anterior se decide utilizar FILTERSIM y proponen alternativas de incorporación de la media local.

2.2 Stanford Geostatistical Modeling Software (SGeMS)

El software SGeMS, *Stanford Geostatistical Modeling Software* [Remy et al, 2007], fue desarrollado en la Universidad de Stanford, California, y presenta una gran cantidad de algoritmos para el desarrollo de estimaciones y simulaciones de variables regionalizadas en el espacio. Se destaca por ser un software libre con sus respectivos códigos fuentes abiertos para todo usuario y por incorporar un visualizador 3D.

Se opta por utilizar el software SGeMS para el desarrollo de la presente memoria debido a: los códigos fuentes de los algoritmos son libres y abiertos (desarrollados en lenguaje C++); es posible intervenir los algoritmos actuales o bien agregar nuevos a través del uso de plug-ins; la importación y exportación de datos acepta diversos formatos además de la capacidad de visualización en 3D y la generación de estadísticas básicas entre otras utilidades.

2.2.1 Interfaz, estructura y utilidades

La interfaz de SGeMS se aprecia en la siguiente figura 21 y consiste en:

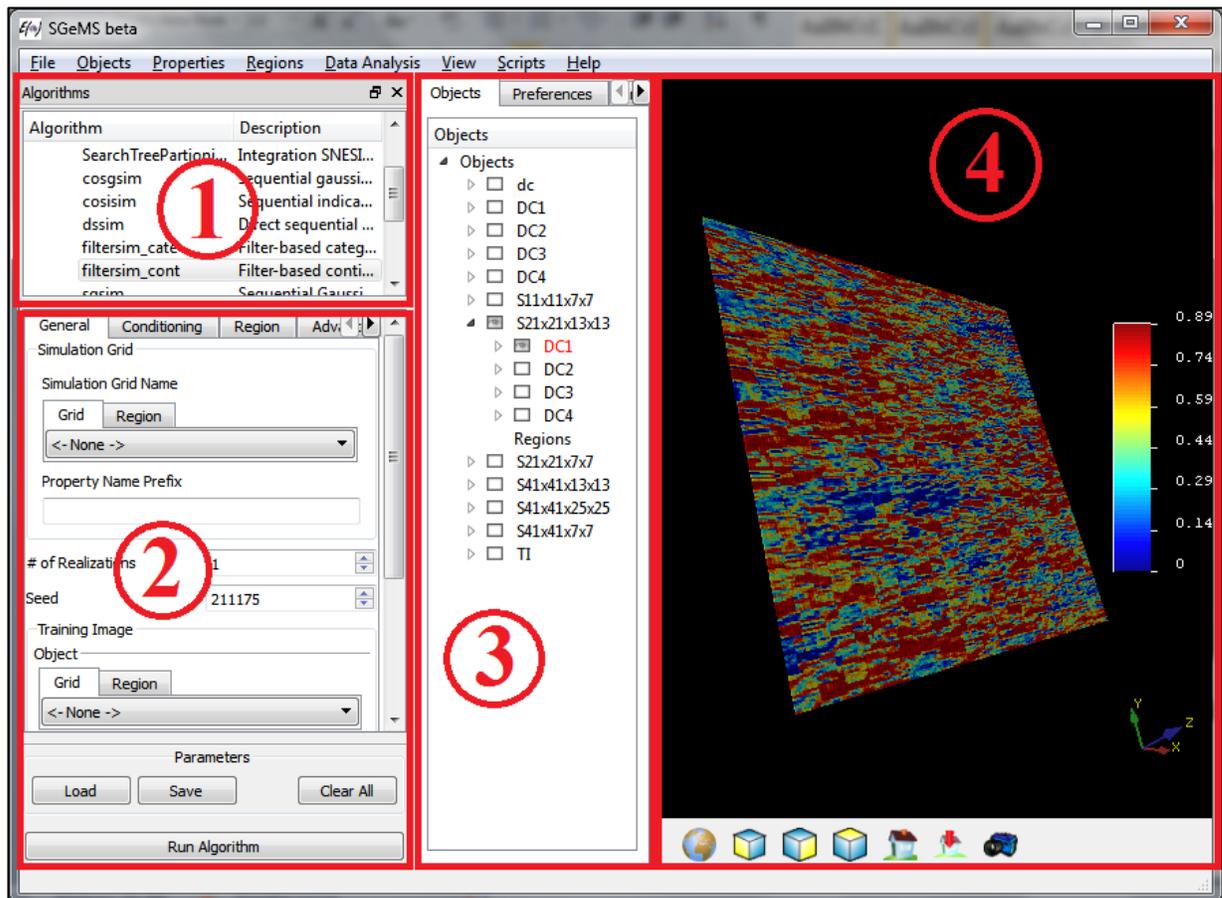


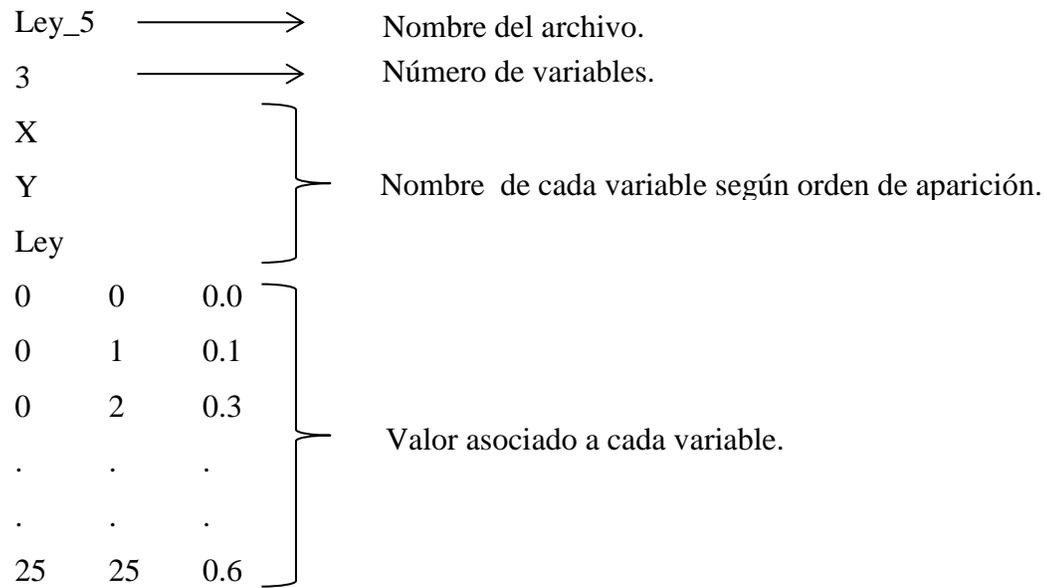
Figura 21 Interfaz SGeMS Beta 2.1

1. Pestaña de los algoritmos de estimación/simulación y de las utilidades varias que vienen incorporadas en SGeMS. De agregar un algoritmo a través del uso de plug-ins este aparecerá en esta pestaña.
2. Pestaña para todos los parámetros de entrada necesarios para el correcto funcionamiento de los algoritmos y herramientas que corresponden directamente al proceso seleccionado en la pestaña 1.
3. Lista de todos los objetos pertenecientes al proyecto que se está llevando a cabo. Están tanto los objetos importados como los resultados. Se aprecia que cada objeto presenta subdivisiones, esto se debe a que se crea el entorno del objeto (grilla 2D o 3D, conjunto de puntos, etc.) de manera vacía y luego se le van incorporando los valores correspondientes a las variables que se vayan generando para tal objeto.
4. Visualizador 3D de los datos. Escala de colores para las variables determinadas por el usuario en la pestaña 3 subpestaña “preferences”. Es posible generar archivos de imágenes correspondientes a lo que en el momento se esté visualizando.

La importación de datos es posible realizarla como grilla regular o como conjunto de puntos en el espacio y se realiza a través de archivos de textos con encabezado según la convención GSLIB (*Geostatistical Software Library, [Deutsch and Journel, 1998]*), es decir:

- Nombre del archivo: ubicado en la primera línea del archivo de texto
- Número de variables: se ubica en la segunda línea del archivo. Debe ser un número entero que especifica la cantidad de variables presentes (**n**).
- Nombre de las variables: las próximas (**n**) filas representan el nombre de cada variable.
- Valor de las variables: desde la fila (**n+3**) hasta el final representan los valores asignados a un punto y deben contener (**n**) valores numéricos a lo largo de toda la fila. Se asocia el primer valor de la fila con el primer nombre de la variable y así sucesivamente.

A continuación se presenta un ejemplo en donde el nombre del archivo es Ley_5. Se tienen 3 variables cuyos nombres son X,Y y Ley, y finalmente el valor para cada una de las 3 variables.



La estructura de SGeMS consiste en la importación de los datos y su posible visualización y estadística básica, luego se selecciona la herramienta a utilizar fijando todos los parámetros necesarios, se realiza la estimación/simulación/otros y es posible exportar los resultados, analizarlos y visualizarlos.

Dentro de las utilidades destaca la posibilidad de realizar:

- Post procesamiento de Kriging.
- Transformación de histogramas.
- Simulación y co-simulación secuencial gaussiana y de indicadores.
- Simulación basada en filtros para variables continuas y categóricas, FILTERSIM.
- *Single normal equation simulation*, SNESIM.
- Estimación por Kriging: simple, ordinario, de bloques, de indicadores, etc.

2.2.2 FILTERSIM

La implementación del algoritmo FILTERSIM dentro de SGeMS está subdividida en 2 partes: para variables continuas y para variables categóricas. Para efectos de la presente memoria se analiza y utiliza la versión diseñada para variables continuas [Wu et al., 2008].

A continuación se presentan cada uno de los parámetros de entrada para dar comienzo a la simulación, siendo algunos de carácter necesario y otros optativos.

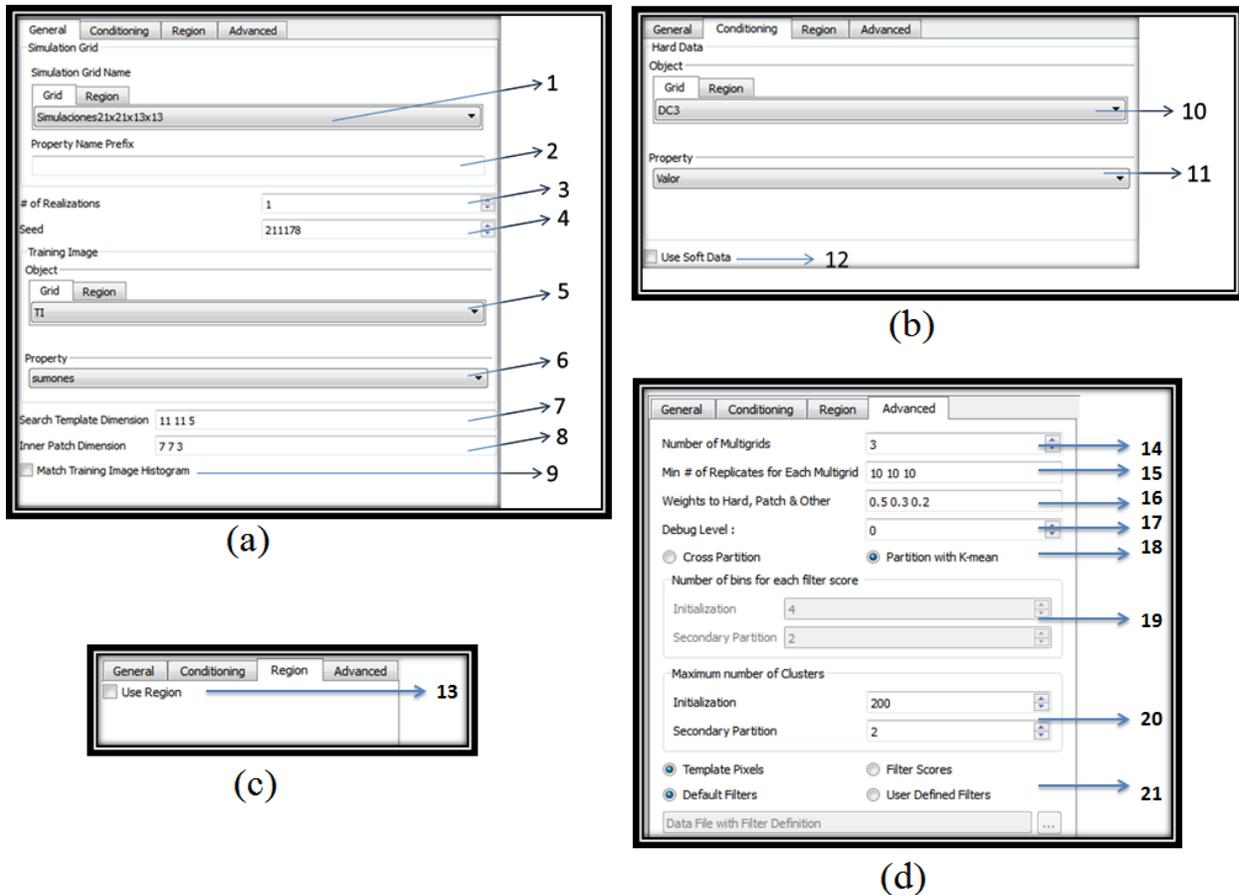


Figura 22 FILTERSIM en SGeMS. Se muestran las pestañas: (a) General, (b) Condicionamiento, (c) Región, (d) Avanzado.

1. Nombre de la grilla [Obligatorio]. Se debe haber creado previamente la grilla sobre la cual se va a realizar la simulación.
2. Nombre de prefijo para la propiedad de la variable que va a ser simulada [Obligatorio].
3. Número de realizaciones [Obligatorio]. Cantidad de veces que se va a realizar la simulación bajo las mismas condiciones.

4. Semilla [Obligatorio]. Número necesario para realizar la secuencia de números pseudo-aleatorios.
5. Imagen de entrenamiento desde donde se obtendrán los patrones condicionantes [Obligatorio].
6. Propiedad a analizar desde la imagen de entrenamiento como valor condicionante [Obligatorio].
7. Grilla de búsqueda móvil [Obligatorio] (*search template*). Tamaño de la grilla que genera los patrones condicionantes a partir de la imagen de entrenamiento. Es la misma grilla que se utiliza para el análisis de la condición proveniente de los datos duros condicionantes.
8. Dimensión de la grilla interior como datos secundarios fijos [Obligatorio] (*Inner patch*). Define el tamaño del sector que fijará, a la grilla simulada, los datos recién simulados. Debe ser de tamaño menor o igual a la grilla de búsqueda móvil.
9. Ajuste con el histograma de la imagen de entrenamiento [Optativo]. Da la posibilidad de que la estadística, histograma, de los datos simulados coincidan con los de la imagen de entrenamiento.
10. Datos duros condicionantes [Optativo]. Define el conjunto de datos que caracterizarán los eventos condicionantes.
11. Propiedad del patrón estructural, desde los datos duros condicionantes [Obligatorio si se opta por usar datos condicionantes]. Es la propiedad a ser simulada.
12. Uso de datos suaves [Optativo]. Capacidad del algoritmo de integrar información secundaria. Cuando un nodo no presenta información condicionante es posible utilizar el valor obtenido en una realización anterior.
13. Uso de regiones [Optativo]. Su análisis está fuera de los alcances de la presente memoria.
14. Número de multigrillas [Obligatorio]. Cantidad de grillas a utilizar en el proceso de simulación para cada realización. Ver sección 2.1.2.5
15. Número mínimo de datos por cada grilla [Obligatorio]. La cantidad de números mínimo debe ser igual al número de multigrillas.
16. Pesos [Obligatorio]. Se asocian pesos para los datos duros condicionantes, suaves y otros. La suma de estos debe ser igual a 1 y todos mayores o iguales que cero.
17. Resultados de la simulación [Obligatorio]. Establece cuales van a ser los resultados deseados al simular. Se debe optar por un número entero entre [0,1,2,3] , los cuales representan:

- 0: solo los resultados finales de la simulación son entregados.
- 1: los mapas de puntuación asociados a la grilla más fina son entregados junto con los resultados de las simulaciones.
- 2: además de entregar la información correspondiente a las opciones 0 y 1, entrega los resultados sobre las grillas gruesas.
- 3: agrega una propiedad a la grilla simulada, el mapa de indicadores para los prototipos.
18. Método para generar el agrupamiento de patrones [Obligatorio]: *cross partition* o *k-mean partition*
19. Parámetros de entrada si se optó por utilizar Cross partition [Obligatorio si se optó por cross partition]. El número de inicialización es la cantidad de divisiones de igual frecuencia a los rangos de puntaje. El número de partición secundaria es la cantidad posible de subdivisiones de los grupos de patrones.
20. Parámetros de entrada si se optó por utilizar Filter Score [Obligatorio si se optó por filter score]. El número de inicialización establece la cantidad de nodos a generar en el espacio. El número de partición secundaria es la cantidad máxima que se puede subdividir un grupo de patrones.
21. Método para el cálculo del patrón más cercano a la condición de evento [Obligatorio] y los filtros a utilizar [Obligatorio]. De los métodos de distancia se debe elegir entre *Template pixels* o *Filter score*. El uso de filtros puede ser: por defecto o generados por el usuario (más detalles en el manual de SGeMS: “*Applied Geostatistics with SGeMS: A Users’ Guide*” capítulo 8.2 [Remy, Boucher and Wu, (2007)]).

3 METODOLOGÍA

Entendiendo por metodología al ordenamiento de los procesos para llevar a cabo los objetivos, se plantea el siguiente procedimiento:

1. Estudio de los algoritmos SNESIM y FILTERSIM con la finalidad de exponer las razones de utilizar FILTERSIM a lo largo de la memoria.
2. Analizar la sensibilidad de FILTERSIM a sus parámetros de entrada desarrollando un caso de estudio sintético.
 - a. Se analiza tanto el tiempo de cómputo como la calidad en los resultados.
3. Utilizar los parámetros óptimos, provenientes del caso de estudio sintético, en el desarrollo de un caso de estudio real (yacimientos cupríferos).
4. Comparar los resultados simulados contra los resultados al estimar mediante kriging ordinario, en el caso de estudio real.
 - a. Se analizan mapas de probabilidad, varianzas condicionales versus varianzas de kriging, y las realizaciones por separado y en conjunto.
5. Proponer alternativas para la incorporación de la media local en el proceso de simulación utilizando FILTERSIM.

Se propone como posible metodología para una continuación del trabajo de la presente memoria:

1. Generar un nuevo filtro que determine la media local dentro de la grilla de búsqueda al realizar la generación de patrones a partir de la imagen de entrenamiento.
2. Agregar como nueva componente al vector de filtros la media local, o bien modificar la función distancia al patrón más cercano agregando la media local como parámetro de entrada.

3. Realizar el proceso de agrupamiento (Clustering) de los patrones obtenidos a partir de la imagen de entrenamiento.
4. Generar el prototipo asociado para cada conjunto de patrones agrupados por similitud.
5. Agregar como nuevo parámetro de entrada la media local esperada por el usuario en la grilla (2D o 3D) a simular.
6. Agregar la media local al vector de filtros proveniente de los eventos condicionantes, o bien seguir con el proceso normal y modificar la función distancia.
7. Modificar la función distancia que determina que prototipo, y por ende qué patrón, es más similar al evento condicionante, incorporando tanto los datos duros como la media local esperada por el usuario.
8. Compilación del algoritmo modificado en su versión desarrollada para SGeMS utilizando el lenguaje de programación C++.
9. Desarrollo de un caso de estudio sintético para estudiar el FILTERSIM actual y cómo repercuten los parámetros de entrada en la calidad de los resultados con la finalidad de establecer un criterio de comparación entre el algoritmo modificado y el actual.

4 APLICACIÓN FILTERSIM CONVENCIONAL: CASO DE ESTUDIO SINTÉTICO

Esta sección tiene por objetivo aplicar de manera directa el algoritmo FILTERSIM a un caso de estudio sintético con la finalidad de comprender cómo afectan las distintas variables, que el usuario puede modificar dentro del algoritmo, en la calidad de los resultados finales de la simulación. Entender cómo afectan estas variables servirá para determinar en qué rango o bien en qué valor específico fijarlas para realizar luego la comparación entre el algoritmo modificado y el no modificado, con conocimiento previo de que los resultados de las simulaciones serán lo más parecido posible a la realidad. Entendiendo por caso de estudio sintético al hecho de que los datos condicionantes presentes en la grilla de simulación provienen directamente de la imagen de entrenamiento y la calidad en la simulación se mide en cuan semejante son los resultados respecto a la T.I.

4.1 Metodología para el desarrollo del caso de estudio

La metodología de desarrollo es:

1. Presentación del caso de estudio: imagen de entrenamiento y datos condicionantes.
2. Definición de las variables a considerar: tamaño de la grilla de búsqueda, de la grilla interna de congelamiento, de la cantidad de datos condicionantes y de la metodología para la determinación del patrón más cercano al evento condicionante.
3. Se establecen los criterios para evaluar los resultados de las simulaciones: tiempo total de simulación y calidad de la simulación. En este último se mide la probabilidad de ocurrencia sobre ciertos umbrales, estadísticas básicas y análisis variográfico. Entendiendo el análisis variográfico como una manera de verificar la reproducción de las estadísticas de la función generadora de cada realización y no parte de la simulación.
4. Presentación de los resultados: visualización del promedio de las realizaciones (35 en total) para cada configuración de parámetros en análisis (48 en total), estadísticas básicas (media y varianza) sobre cada una de las 35 realizaciones generadas en cada conjunto de parámetros (ver sección 4.3.), mapas probabilísticos para visualizar la probabilidad de continuidad espacial, variogramas experimentales sobre una de las 35 realizaciones (elegida de manera aleatoria) y su comparación al variograma experimental de la T.I.
5. Análisis crítico de la repercusión que tiene el modificar cada variable en el proceso de simulación.

4.2 Presentación del caso de estudio

Realizar al menos un caso de estudio toma relevancia cuando se desea realizar modificaciones en procesos ya establecidos, ya que es una buena forma de cuantificar el efecto que tiene la modificación de una o más variables en el resultado final.

El caso se lleva a cabo de la siguiente manera:

1. Se elige una imagen de entrenamiento (T.I.), figura 24.a, representativa de una zona geológica asociada a depositación de sedimentos, con variables continuas en un intervalo espacial de 250 por 240 píxeles pertenecientes a un corte vertical de la zona antes mencionada (2 dimensiones).
2. Al ser un caso de estudio sintético, los datos duros son extraídos de manera aleatoria desde la T.I.
3. Utilizando el software de simulación SGeMS, específicamente la herramienta “Algoritmo de simulación Filtersim para variables continuas”, se generan 35 realizaciones para cada configuración de parámetros variando:
 - a. Grilla de búsqueda móvil sobre la T.I. (*search template*)
 - b. Dimensión de la grilla interior como datos simulados a congelar (*inner patch dimension*).
 - c. Cantidad de datos duros (muestras) provenientes de la T.I.
 - d. Metodología para el cálculo de la distancia entre el evento condicionante y el prototipo más cercano: *template pixel* o *filter score*.
4. Se realiza un proceso de post-simulación que consiste en determinar la conducta estructural promedio que presentan las realizaciones dentro de cada configuración de parámetros, así como la probabilidad de ocurrencia de la variable en el espacio sobre cierto umbral.
5. Se presentan los resultados y se realiza un análisis crítico de estos.

Cabe destacar que todas las simulaciones se realizan en un laptop con procesador Intel® Core™ i3 M330 @ 2.13 GHz, memoria RAM de 4 GB y sistema operativo (S.O.) Windows 7 Ultimate 64 bits. Es relevante esta mención ya que uno de los parámetros a evaluar es la velocidad del proceso de simulación, el cual puede variar entre una condición de hardware/software y otra.

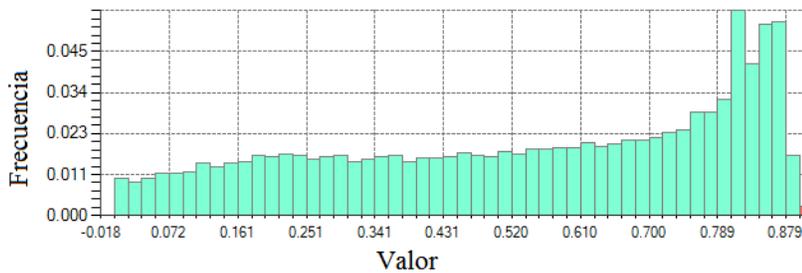
4.2.1 Imagen de entrenamiento (T.I.)

La imagen de entrenamiento, figura 24.a, representa un corte vertical de una zona geológica de depositación sedimentaria, a través de variables continuas en el rango $[0, 0.897]$ (tabla 3) cuyo histograma de frecuencia se muestra en la figura 23. Estas variables pueden ser representativas de: leyes, densidades, electro-conductividad, permeabilidad, entre otras opciones geometalúrgicas. Las dimensiones espaciales se encuentran en la tabla 2:

Tabla 2 Dimensiones de la Imagen de Entrenamiento

Dimensión	N° de pixeles
Eje X	250
Eje Y	240
Eje Z	1

Tabla 3 Estadística imagen de entrenamiento



N° de datos	60,000
Promedio	0.544
Varianza	0.068
Máximo	0.897
Mínimo	0
Mediana	0.591

Figura 23 Distribución de los valores en la imagen de entrenamiento

A pesar de que el algoritmo FILTERSIM no utiliza en su proceso de simulación el variograma como herramienta de simulación, se considera que el variograma es una herramienta adecuada para comparar los resultados de las simulaciones respecto a la realidad que se desea simular. Es por esto que además de las estadísticas básicas, tabla 3, se aplica un análisis de correlación espacial a la imagen de entrenamiento entre sus valores, tal como muestran las figuras 24.b y 24.c utilizando como variables de entrada los datos en la tabla 4.

Tabla 4 Variables de entrada - Mapa variográfico

Variables de entrada - Mapa Variográfico	
Variograma	Semi-variograma tradicional
N° pasos en X	125
N° pasos en Y	120
Largo del paso en X	1.0
Largo del paso en Y	1.0
Altura en Z	1.0
Tipo de grilla	Regular

La figura 24 presenta la imagen de entrenamiento (24.a), su respectivo mapa variográfico (24.b) y los variogramas experimentales en las direcciones principales de anisotropía:

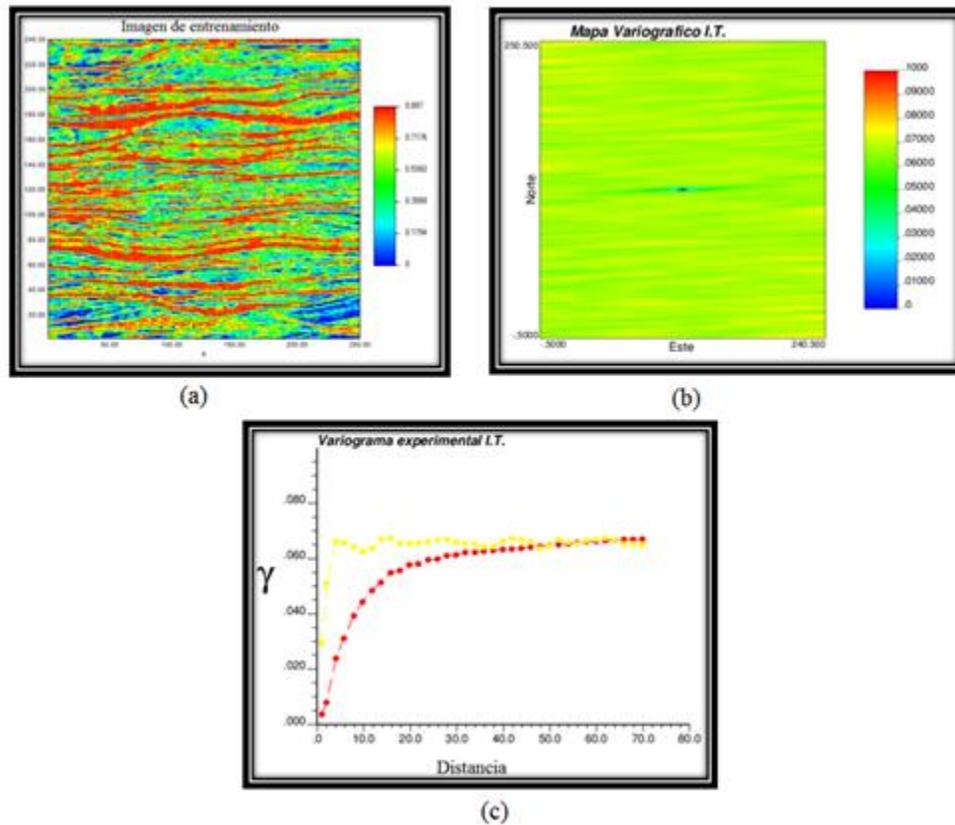


Figura 24 (a) Imagen de entrenamiento, (b) Mapa variográfico sobre la T.I., (c) variogramas experimentales en las direcciones de principales de anisotropía

De la figura 24.b se observan claramente bandas en dirección semi-horizontal N85E, reflejo de la existencia de anisotropía. Es por esto que se ha decidido utilizar como direcciones para el análisis del variograma experimental N85°E y N5°O. Realizando un variograma experimental en ambas direcciones (N85°E en color rojo y N5°O en color amarillo) con 35 pasos de 2 unidades cada uno y tolerancia de 1 en el paso se obtienen los variogramas experimentales en la figura 24.c.

4.2.2 Datos condicionantes

Entendiendo por evento condicionante a la disposición espacial y el valor asociado a cada elemento de un conjunto de datos ubicados dentro de un sector específico, se propone extraer, de manera aleatoria, el 0.1%, 0.5%, 1% y 5% de los datos presentes en la imagen de entrenamiento. Cabe destacar que FILTERSIM tiene la capacidad de generar simulaciones sin la necesidad de datos duros condicionantes.

En la figura 25 se observa la distribución espacial de los datos condicionantes mientras que en la tabla 5 se presentan sus estadísticas básicas.

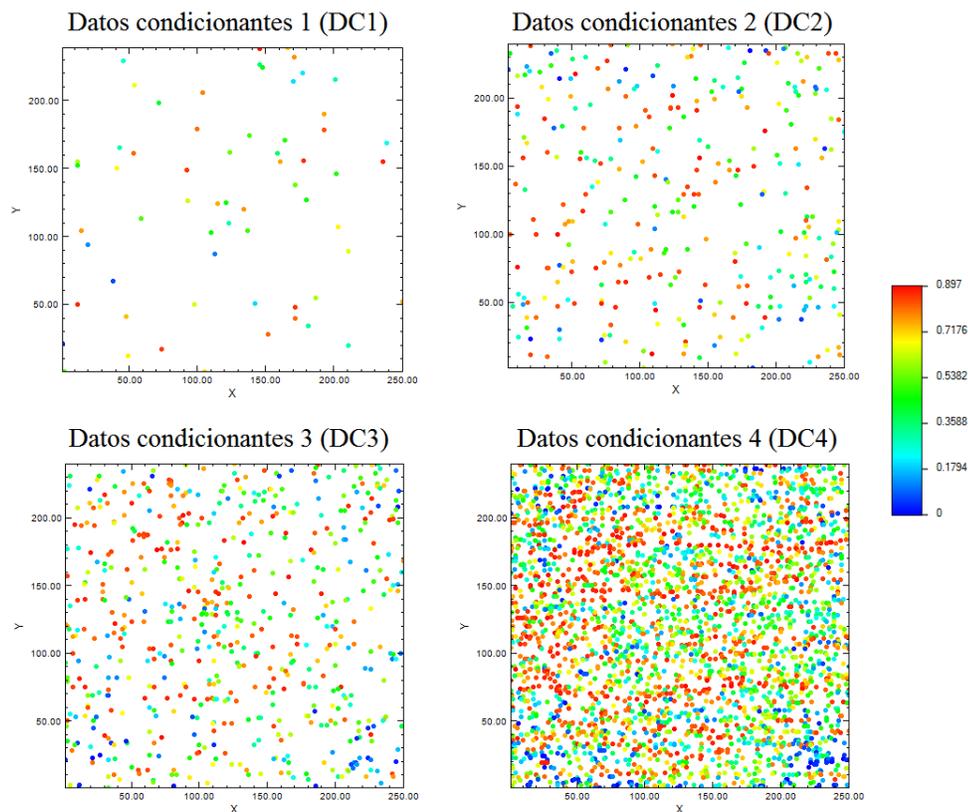


Figura 25 Distribución espacial de los datos condicionantes DC1, DC2, DC3 y DC4.

Tabla 5 Estadística básica de los datos condicionantes

	DC 1	DC 2	DC 3	DC 4	T.I.
N° de datos	62	308	591	2,983	60,000
Cantidad extraída de la T.I. (%)	0.1	0.5	1.0	5.0	100.0
Promedio	0.555	0.55	0.538	0.546	0.544
Varianza	0.056	0.068	0.068	0.068	0.068
Máximo	0.876	0.897	0.897	0.897	0.897
Mínimo	0.035	0	0	0	0
Mediana	0.594	0.603	0.568	0.603	0.591

Se observa en la tabla 5 que los conjuntos de datos condicionantes representan el 0.1%, 0.5%, 1.0% y 5.0% del total (60,000) de datos presentes en la imagen de entrenamiento.

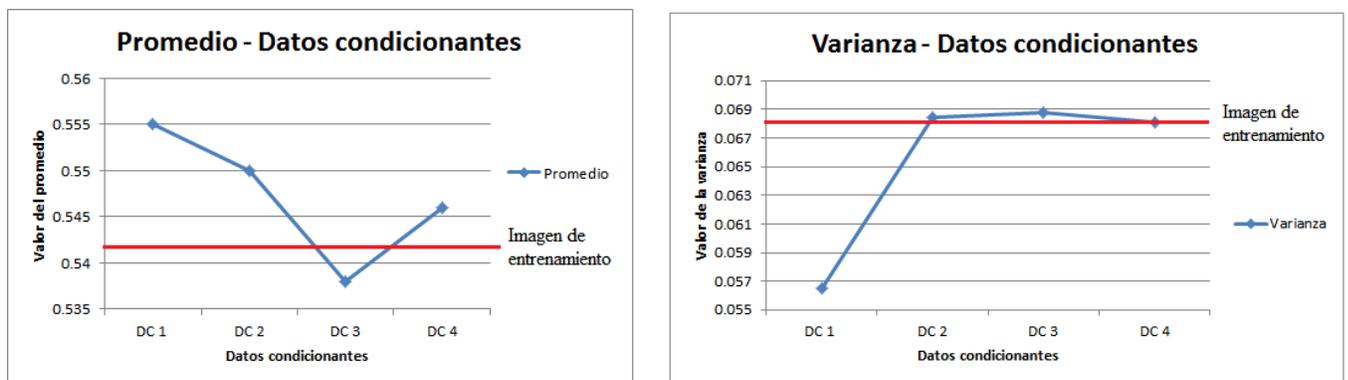


Figura 26 Promedio y varianza de los datos condicionantes

La figura 26 presenta el promedio y la varianza para cada conjunto de datos condicionantes en comparación a la imagen de entrenamiento. Estos indicadores toman relevancia al momento de compararlos respecto a los resultados basados en un conjunto en específico.

Los histogramas y estadísticas básicas para cada base de datos condicionantes están en la sección de anexos: Anexo A.

4.3 Variables a considerar

Tal como se describió en la sección 2.3.2 muchos son los posibles parámetros de entrada a considerar en una sola simulación de FILTERSIM, dentro de los cuales se encuentran:

- Tamaño de la grilla de simulación.
- Número de realizaciones asociadas a cada configuración de parámetros.
- Número semilla para la generación de números pseudo-aleatorios.
- Imagen de entrenamiento.
- Tamaño de la grilla móvil de búsqueda en la imagen de entrenamiento
- Tamaño de la grilla interna. Establece el cuadrante de los datos que luego de ser simulados serán conservados. Los valores fuera de este cuadrante no se almacenan en la grilla simulada. Los valores internos a esta grilla se consideran luego como datos secundarios.
- Ajustar el histograma de los datos simulados al histograma de los datos de la T.I.
- Conjunto de datos condicionantes (muestras) y la propiedad asociada a cada elemento que será tomada en consideración.
- Incorporación de información secundaria (datos suaves) y propiedad de los elementos a considerar.
- El número de multigrillas utilizadas en la simulación.
- Número mínimo de elementos al interior de un conjunto de patrones representados por un solo prototipo. Para cada multigrilla se debe asignar un número mínimo de elementos.
- Peso asociado a datos provenientes de un evento condicionantes (muestras), a un sector de información secundaria, u otro tipo de información.
- Metodología para la agrupación de los patrones provenientes de la imagen de entrenamiento al ser recorrida por completo utilizando la grilla de búsqueda. Dos son las alternativas de agrupamiento: *Cross partition* o *K-mean partition*. Ello acarrea definir los números de inicialización y partición secundaria los cuales dependerán del método elegido.
- Metodología para la determinación del prototipo más cercano a un evento condicionante: *Template pixels* o *Filter score*.
- Definir si se utilizan filtros personalizados o bien los que el software presenta por defecto.

Se estudia el impacto que tiene, sobre la calidad de las simulaciones, la variación de los siguientes parámetros:

- Tamaño de la grilla de búsqueda móvil (*search template*).
- Tamaño de la grilla interior (*inner patch*) de congelamiento de datos simulados.
- Cantidad de datos condicionantes.
- Metodología para determinar el prototipo más cercano a la condición de evento.

Los valores de estos parámetros que darán lugar a una serie de configuraciones se detallan en la tabla 6.

Tabla 6 Variables a analizar y sus valores, en la simulación

Variable	Valores	Razón
Tamaño de la grilla de búsqueda móvil (<i>search template</i>)	11x11 - 21x21 - 41x41	Se duplica el ancho y alto de la grilla en 2D para analizar su influencia.
Dimensión grilla interior (<i>inner patch</i>)	7x7 - 13x13 - 25x25	Se duplica el ancho y alto de la grilla en 2D para analizar su influencia.
Cantidad de datos condicionantes	0.1% - 0.5% - 1% - 5% de los datos en la imagen de entrenamiento, extraídos de manera aleatoria.	Ver sección 4.2.2
Metodología para determinar el prototipo más cercano a la condición de evento	Template pixels y Filter scores	Ya que no se han realizados estudios para cuantificar la repercusión de estos métodos en las simulaciones FILTERSIM.

Al considerar que la grilla de búsqueda siempre debe ser de tamaño mayor o igual a la grilla interna, se obtiene entonces un total de 48 combinaciones de parámetros (2 métodos, 4 conjuntos de datos condicionantes, 6 combinaciones de grillas). Por lo tanto se tendrán 48 escenarios distintos, con 35 realizaciones por cada una de ellas.

El resto de los parámetros deben permanecer constantes en los 48 escenarios con la finalidad de que no intervengan en los resultados. Dujardin, en 2007, desarrolló como tema de tesis de doctorado la repercusión en el histograma que tiene el variar los parámetros que para el desarrollo de la presente tesis se mantendrán constantes. Basados en ello se asumen estáticos los siguientes valores de los parámetros:

- Número de realizaciones: 35. Se decide optar por 35 realizaciones con la finalidad de poder capturar la dispersión de la media y desviación estándar de las realizaciones para cada simulación.
- Semilla: 211175. Es el valor por defecto, y cumple con ser un número impar y de gran tamaño con la finalidad de generar los números pseudo-aleatorios.
- Ajustar histograma de las realizaciones respecto a la imagen de entrenamiento: No. Se decide no utilizar esta opción, con la finalidad de dejar los resultados arrojados por FILTERSIM con la menor cantidad de intervención previa.
- Uso de datos suaves: No. Se decide no trabajar con datos condicionantes secundarios.
- Número de multigrillas: 3. Dujardin establece que 3 es un número adecuado. Aumentar el número (>3) no toma relevancia ya que el tamaño de las multigrillas está condicionando al tamaño de la imagen de entrenamiento. Por el contrario, disminuir el valor (<3) no logra caracterizar bien la continuidad espacial de las estructuras.
- Mínimo número de elementos por grilla: 10 en cada una de las 3 grillas. Dujardin establece que considerar 10 elementos como mínimo logra generar particiones que llevan a las simulaciones en donde las estructuras son más continuas. Aumentar o disminuir este valor no varía el tiempo de simulación y afecta muy poco a la cantidad de prototipos almacenados.
- Pesos para los datos condicionantes: 0.5; datos simulados congelados: 0.3; otros datos suaves: 0.2. En este caso, al no incorporar datos suaves se considera: datos condicionantes 0.71 (0.5/0.7) y datos simulados congelados 0.29 (0.2/0.7).

- Resultados de la simulación: 1. Se decide sólo analizar los resultados de las simulaciones sobre la grilla fina, sin considerar los mapas de filtros (puntajes) ni las simulaciones en las grillas gruesas y medianas.
- Tipo de partición (agrupamiento): se considera Cross partition en desmedro de K-mean partition ya que la primera logra obtener resultados similares a la segunda con una velocidad de simulación 35 veces menor, en otras palabras, es posible llegar a resultados similares 35 veces más rápido [Dujardin, 2007].
 - Número de inicialización: 4. Se decide por dividir en 4 intervalos de igual frecuencia cada dimensión del espacio de puntuación.
 - División secundaria: 2. De ser necesaria una subdivisión en los grupos de patrones se opta por una partición dual. Ver sección 2.2.3.1.
- Filtros: se utilizan los filtros que vienen por defecto en el SGeMS. Ver sección 2.2.2.2.

4.4 Criterios de evaluación

Se consideran 48 sets de parámetros, todos distintos en al menos un parámetro, con 35 realizaciones cada uno. Los instrumentos e índices para evaluar la calidad de las simulaciones son de vital importancia, es por ello que se establecen los siguientes criterios de evaluación:

- Tiempo de simulación.

En los procesos de evaluación de recursos y simulaciones geológicas no tan sólo toman relevancia la calidad de los resultados sino también el tiempo que llevó obtenerlos. Por ejemplo, si un resultado demora 10 veces más obteniendo sólo un 0.05% más de precisión, se optará por perder ese porcentaje de precisión a favor de optimizar los tiempos de simulación.

Para cuantificar los tiempos de simulación se captura el tiempo, en milisegundo, que demora cada simulación. Luego, a través de tablas y gráficos es posible realizar los análisis pertinentes.
- Calidad de la simulación.

FILTERSIM tiene la capacidad de generar simulaciones sin la necesidad de tener datos condicionantes, tan sólo con la imagen de entrenamiento de donde obtener los patrones de conducta de la estructura. Es por esto, que se decide no realizar una comparación directa entre la imagen de entrenamiento y los resultados de las

simulaciones, a pesar de que se trata de un caso de estudio sintético que llevará a resultados semejantes a la imagen de entrenamiento (ya que los datos condicionantes provienen de ésta). Por lo anterior, se medirá la calidad de las simulaciones en cuanto a la continuidad espacial de las estructuras y en cuán rápido estas se desestructuran. Para realizar esto se utilizan las siguientes herramientas/indicadores:

- Umbrales de probabilidad.

Se realizan mapas de probabilidad con la finalidad de capturar la probabilidad de que un punto en el espacio tenga un valor sobre cierto valor umbral, basando el cálculo de probabilidad sobre las 35 realizaciones asociadas a cada configuración de parámetros. Luego, si el valor umbral caracteriza la presencia de la estructura, los mapas de probabilidad entregan la información necesaria para interpretar cuán probable es que la estructura sea continua a lo largo de la grilla simulada. De la figura 24.a se interpreta que la estructura en color rojo tiene valores en el rango $[0.7, 0.897]$, con esto se establecen los umbrales 0.7, 0.8 y 0.85 para visualizar y capturar la probabilidad de continuidad espacial de las estructuras.

- Estadísticas básicas.

Se estudia la dispersión en el promedio y desviación estándar entre las realizaciones de una misma simulación. Se establece el mínimo, el máximo y el promedio para ambos indicadores (promedio y varianza). Para visualizar los resultados se presentan gráficos Valor del promedio (o valor de la desviación estándar) versus Configuración asociada a la simulación. La finalidad es poder apreciar si las simulaciones reproducen o no las estadísticas de los datos condicionantes.

- Análisis variográfico.

Se estudia la correlación espacial que tienen los datos simulados. Se elige una realización por cada configuración, de manera aleatoria, con la finalidad de comprobar si los variogramas de las realizaciones se asemejan al de la imagen de entrenamiento. Estos deberían tender a ser similares ya que una realización de la función aleatoria debe respetar los primeros 2 momentos (media-varianza y correlación espacial).

4.5 Resultado de las simulaciones

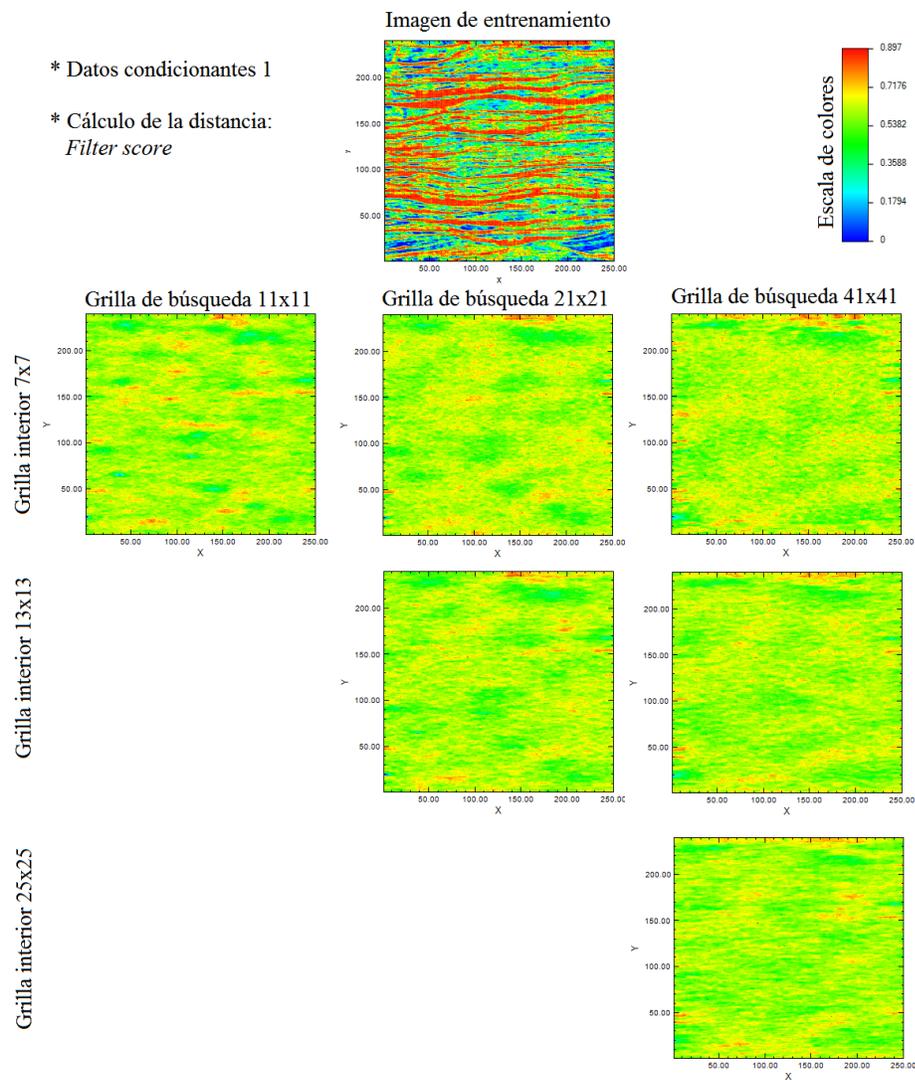
A lo largo de la presente sección se utilizará la abreviación $ExE - IxI$ (o $ExExI - IxIxI$) en donde E e I representan las dimensiones de la grilla cuadrada de búsqueda móvil e interna de congelado respectivamente.

4.5.1 Promedio entre las 35 realizaciones

Se presentan los resultados enunciados como:

Método para el cálculo de la distancia: conjunto de datos condicionantes

Filter Score: Datos condicionantes 1



Filter Score: Datos condicionantes 2

* Datos condicionantes 2

* Cálculo de la distancia:
Filter score

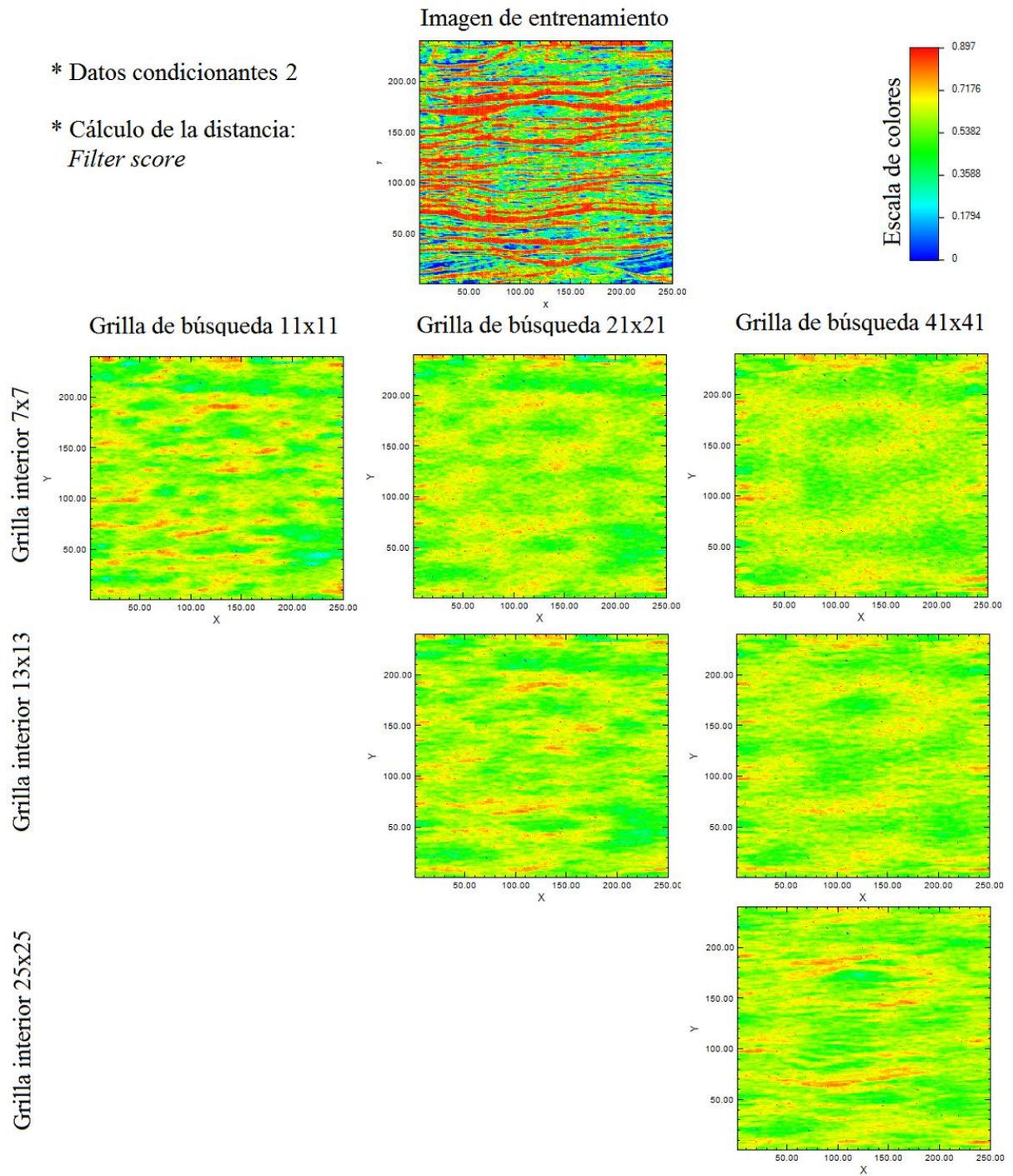


Figura 28 Promedio de 35 realizaciones. Método distancia: filter score. DC 2

Filter Score: Datos condicionantes 3

* Datos condicionantes 3

* Cálculo de la distancia:
Filter score

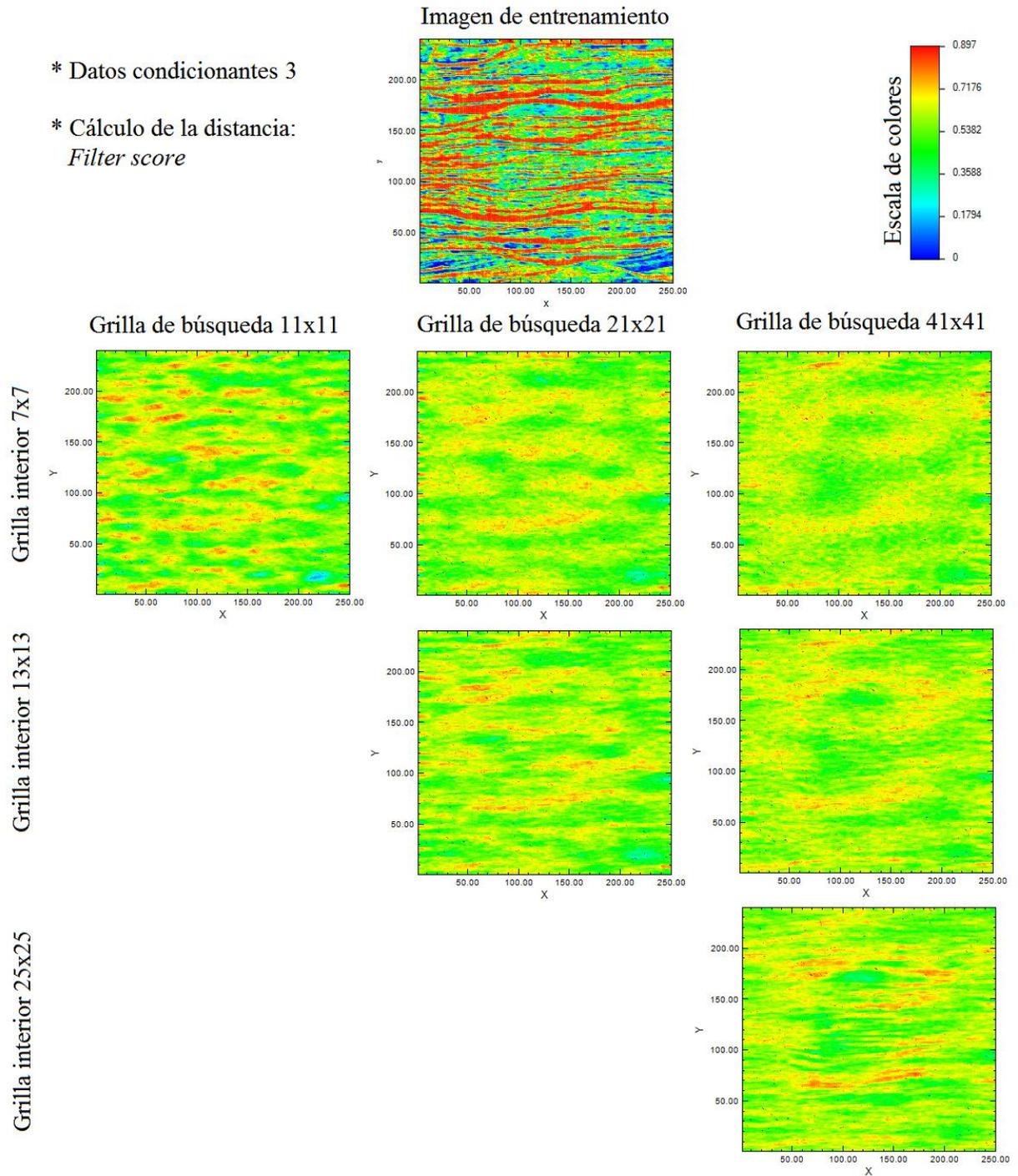


Figura 29 Promedio de 35 realizaciones. Método distancia: filter score. DC 3

Filter Score: Datos condicionantes 4

* Datos condicionantes 4

* Cálculo de la distancia:
Filter score

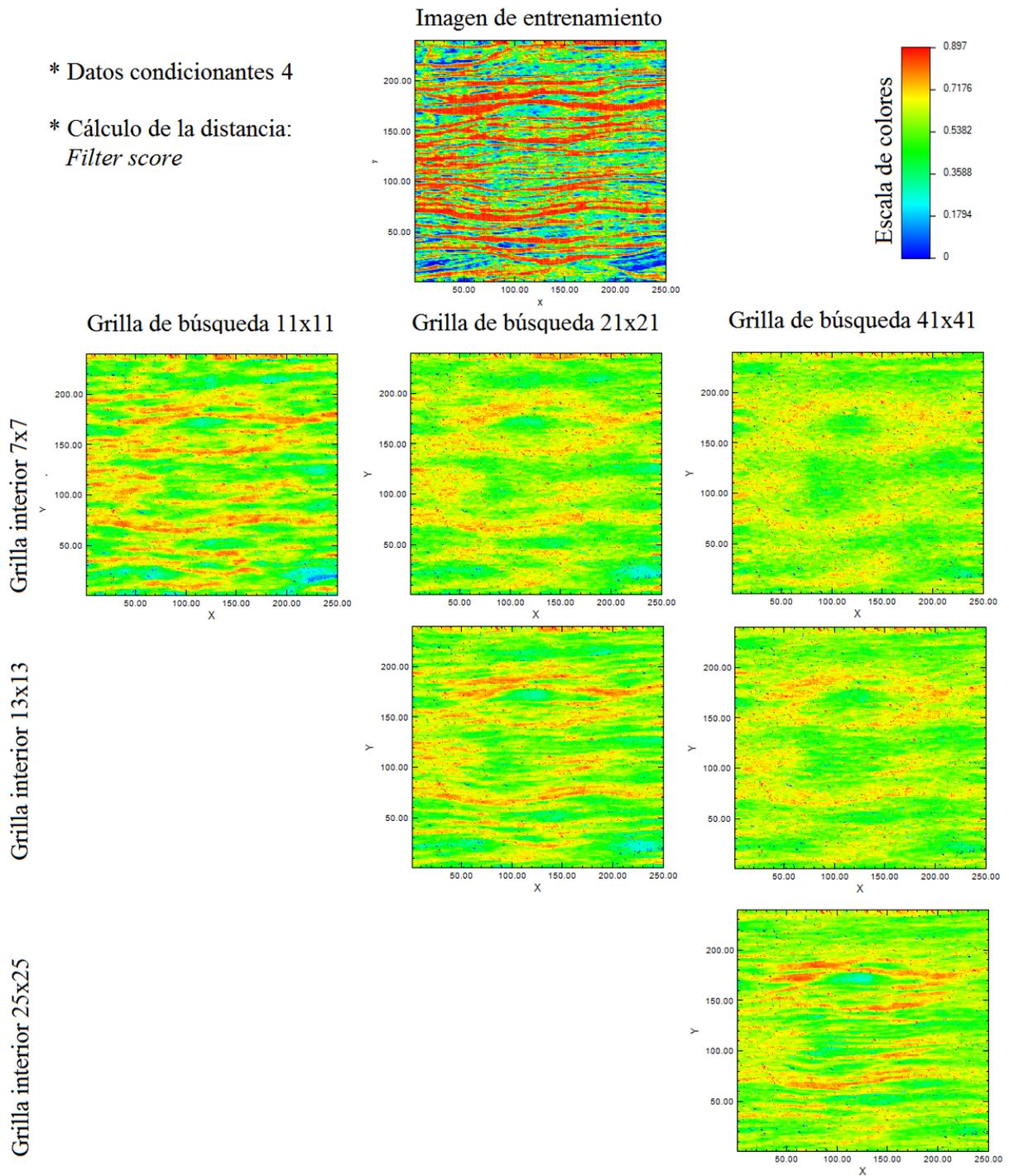


Figura 30 Promedio de 35 realizaciones. Método distancia: filter score. DC 4

Template pixels: Datos condicionantes 1

* Datos condicionantes 1

* Cálculo de la distancia:

Template Pixels

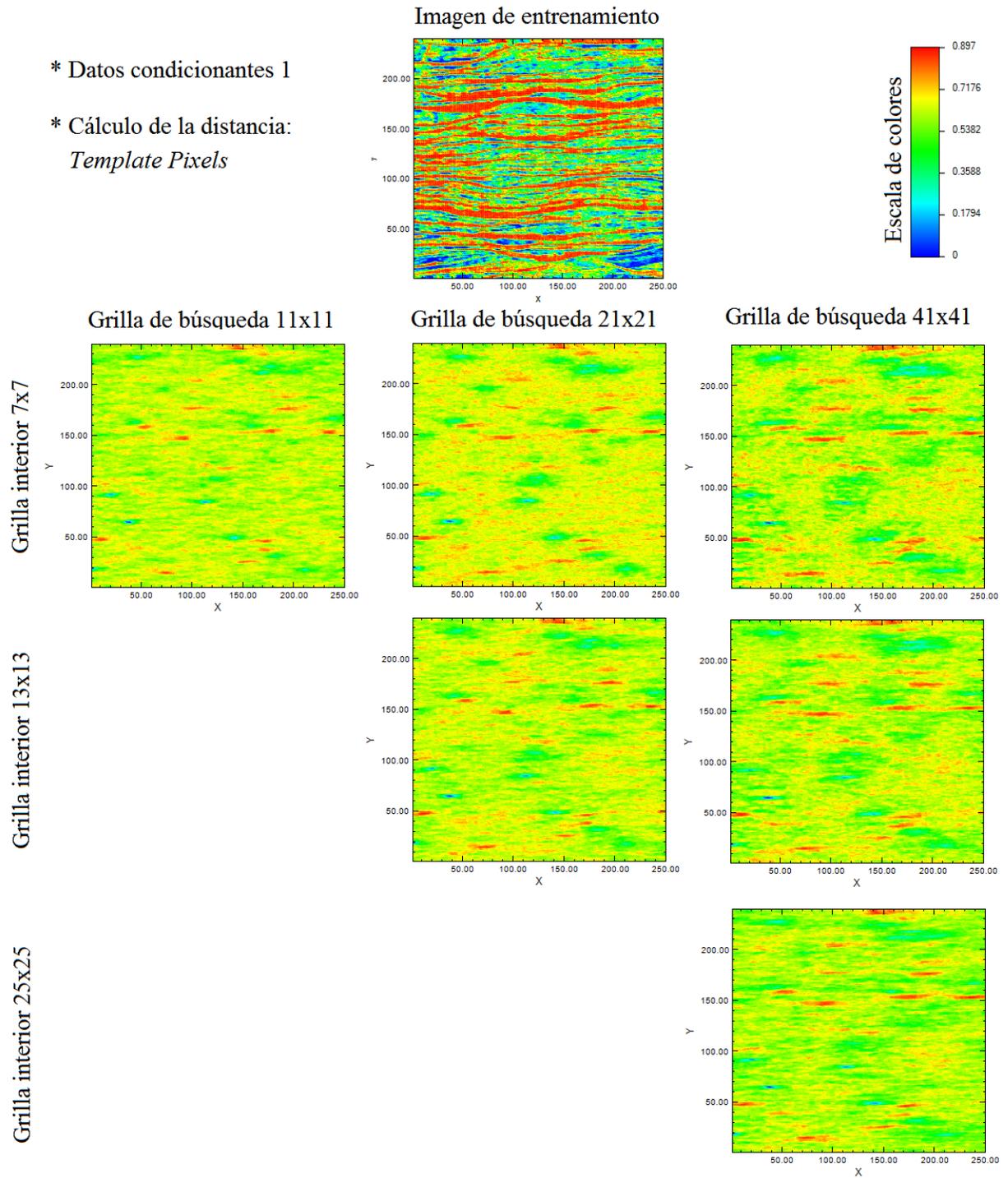


Figura 31 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 1

Template pixels: Datos condicionantes 2

* Datos condicionantes 2

* Cálculo de la distancia:
Template Pixels

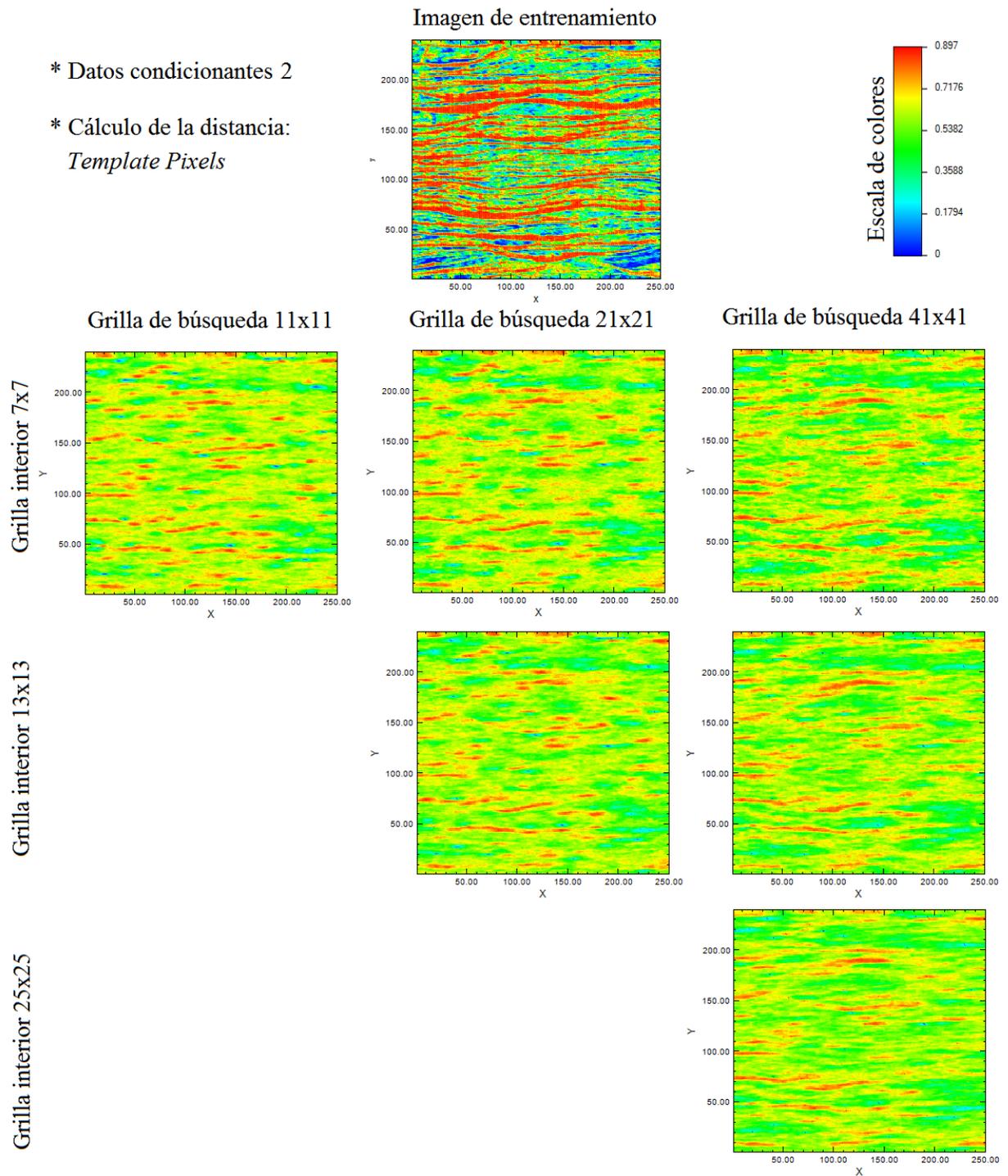


Figura 32 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 2

Template pixels: Datos condicionantes 3

* Datos condicionantes 3

* Cálculo de la distancia:
Template Pixels

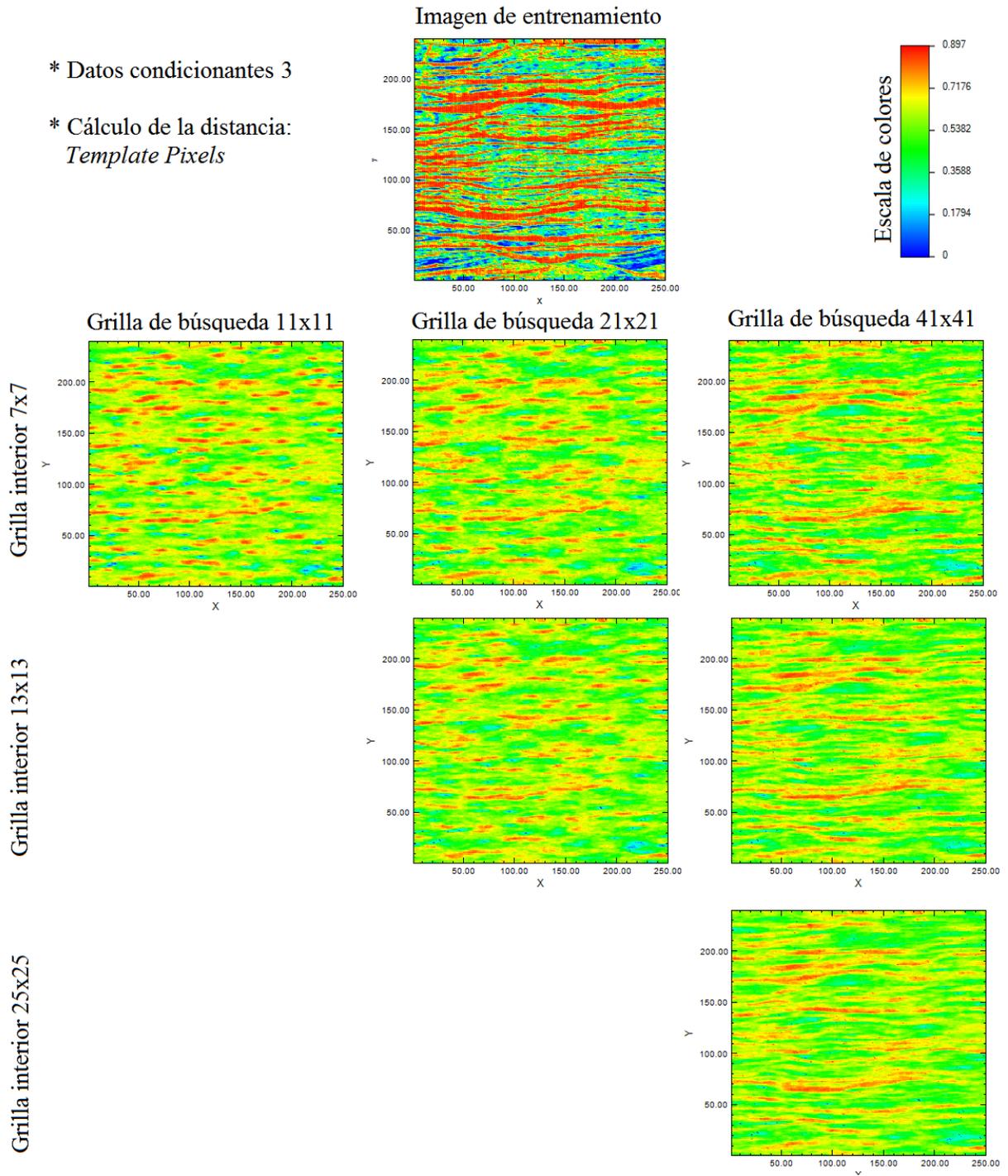


Figura 33 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 3

Template pixels: Datos condicionantes 4

* Datos condicionantes 4

* Cálculo de la distancia:
Template Pixels

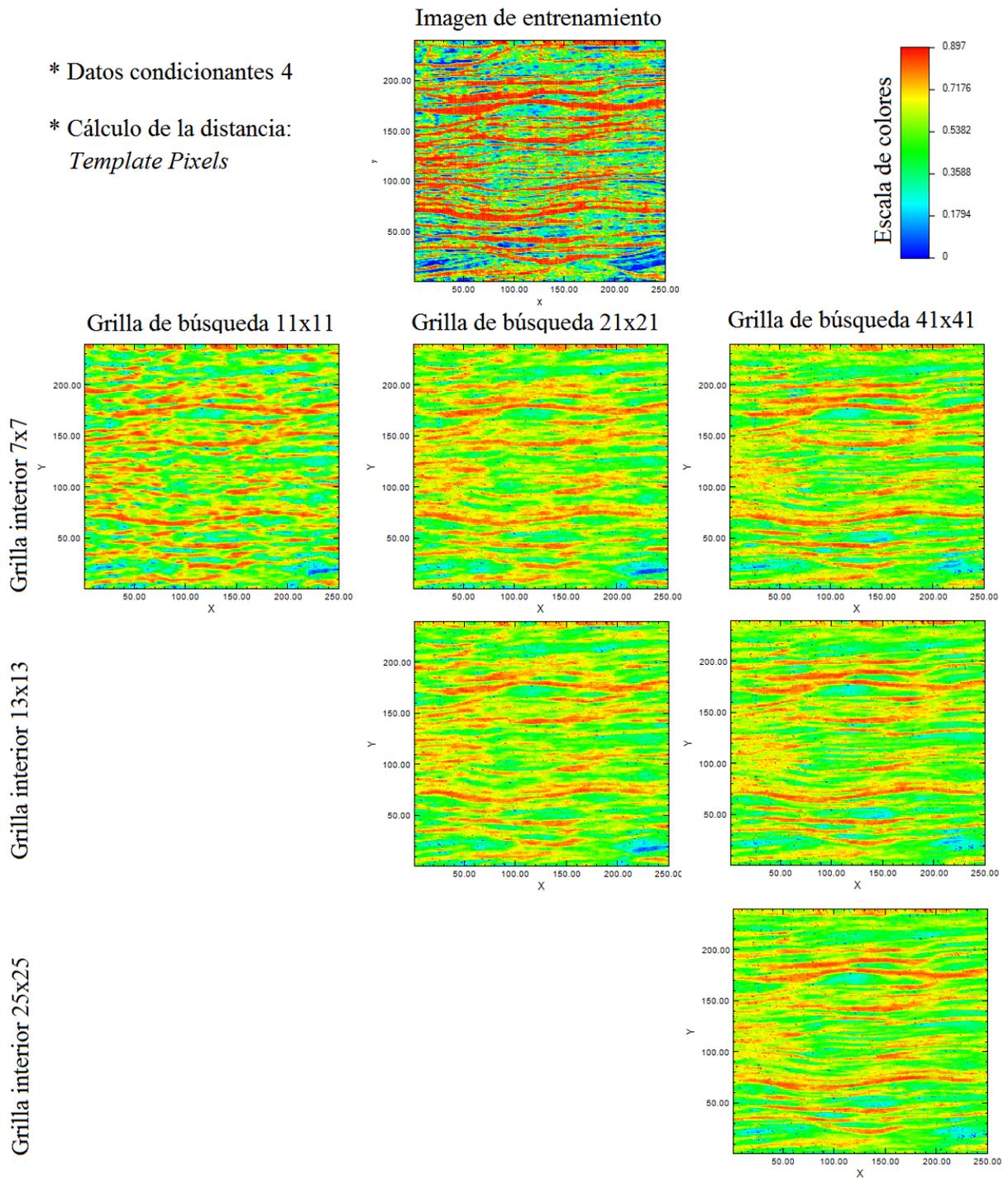


Figura 34 Promedio de 35 realizaciones. Método patrón más cercano: template pixels. DC 4

4.5.2 Tiempo de cómputo

La tabla 7 contiene el tiempo que tardó la realización de cada escenario, considerando 35 realizaciones por cada una, además de las condiciones de hardware (ver sección 4.2).

Tabla 7 Tiempo simulación, considerando 35 realizaciones. Software: SGeMS.

Configuración	Método patrón más cercano			Relación: [TP]/[FS]
	Datos condicionantes	Template pixels [TP] (ms)	Filter scores [FS] (ms)	
11x11 - 7x7	DC1	433,836	143,161	3.03
	DC2	441,849	139,299	3.17
	DC3	435,150	140,221	3.10
	DC4	485,559	140,781	3.45
21x21 - 7x7	DC1	1,466,359	395,513	3.71
	DC2	1,471,165	412,047	3.57
	DC3	1,480,419	409,092	3.62
	DC4	1,481,725	401,010	3.69
21x21 - 13x13	DC1	466,299	138,368	3.37
	DC2	465,929	138,888	3.35
	DC3	467,499	142,227	3.29
	DC4	510,040	143,881	3.54
41x41 - 7x7	DC1	4,557,372	1,050,987	4.34
	DC2	4,742,290	1,019,437	4.65
	DC3	4,951,876	1,032,867	4.79
	DC4	4,601,776	1,018,605	4.52
41x41 - 13x13	DC1	1,450,191	343,917	4.22
	DC2	1,469,410	350,850	4.19
	DC3	1,469,739	349,204	4.21
	DC4	1,457,255	357,923	4.07
41x41 - 25x25	DC1	412,312	134,488	3.07
	DC2	413,999	134,181	3.09
	DC3	419,526	134,108	3.13
	DC4	418,008	143,529	2.91

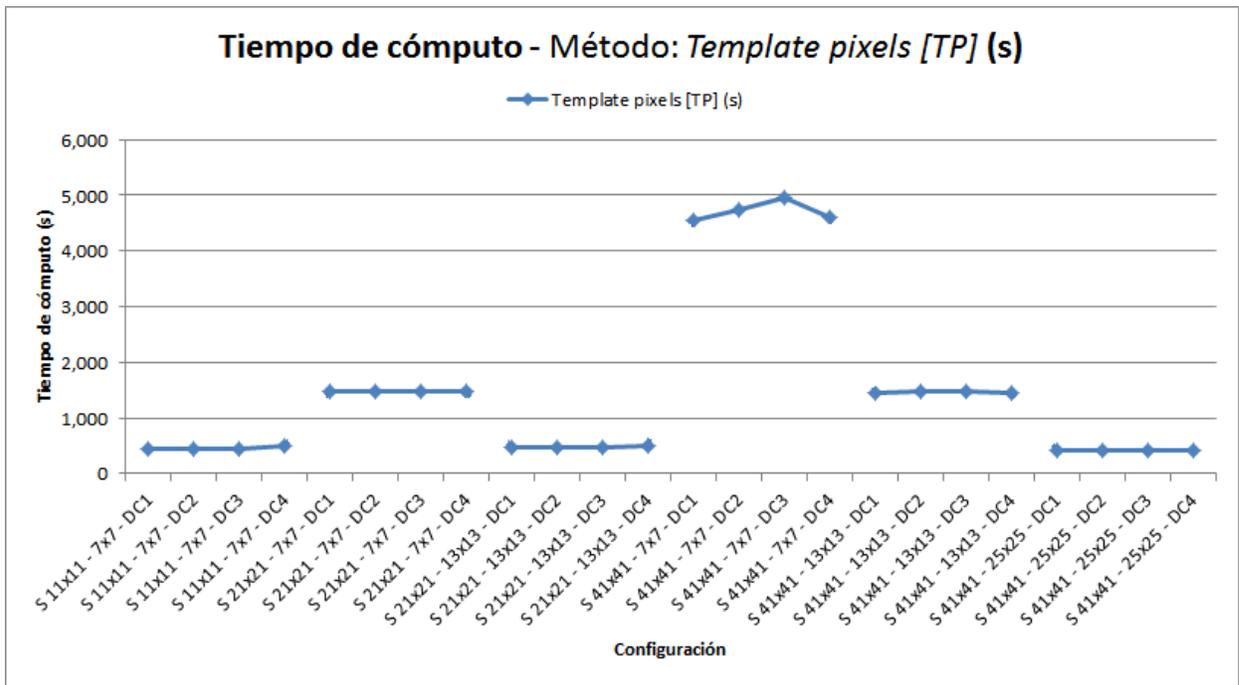


Figura 35 Tiempo de cómputo al simular considerando 35 realizaciones y utilizando el método *template pixels*

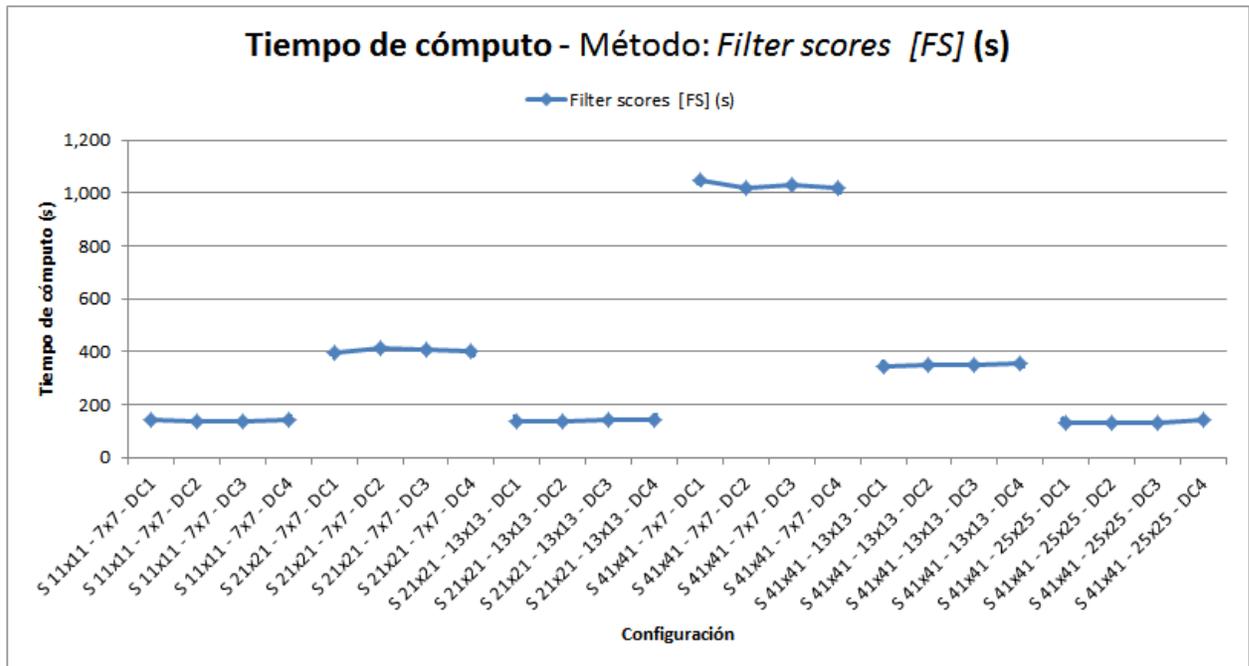


Figura 36 Tiempo de cómputo al simular considerando 35 realizaciones y utilizando el método *Filter scores*

4.5.3 Dispersión en el promedio y la desviación estándar, entre realizaciones

Se observa en la figura 37 que todas las realizaciones presentan un promedio de sus valores por sobre el promedio tanto de sus datos condicionantes como de la imagen de entrenamiento. Además cabe notar que entre más datos condicionantes se utilizan la diferencia entre el mínimo y el máximo valor promedio tiene a disminuir considerablemente y se acerca más al promedio de la imagen de entrenamiento.

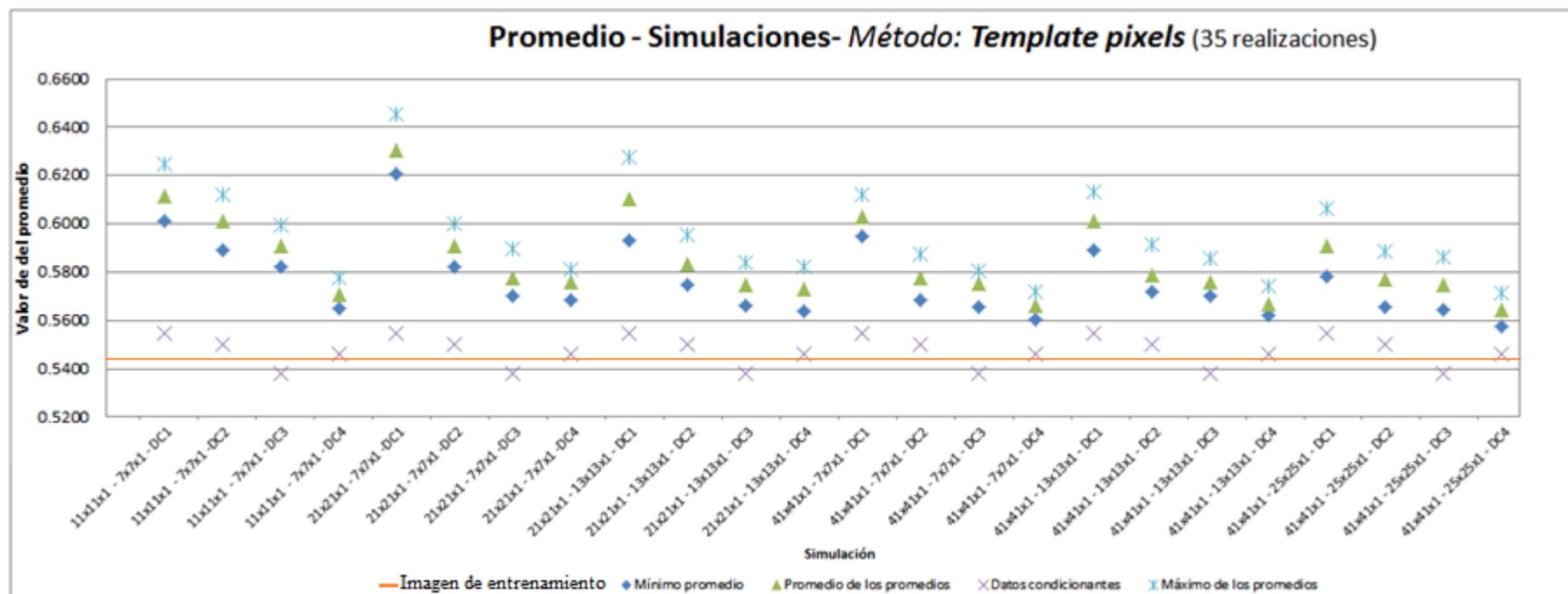


Figura 37 Promedio de los valores internos de las simulaciones – Método: *Template pixels*

Se observa en la figura 38 que la dispersión en el promedio de las varianzas disminuye considerablemente al considerar más datos condicionantes y utilizar el método *template pixels*. Además, al incrementar el número de datos condicionantes la varianza se acerca a la varianza de kriging.

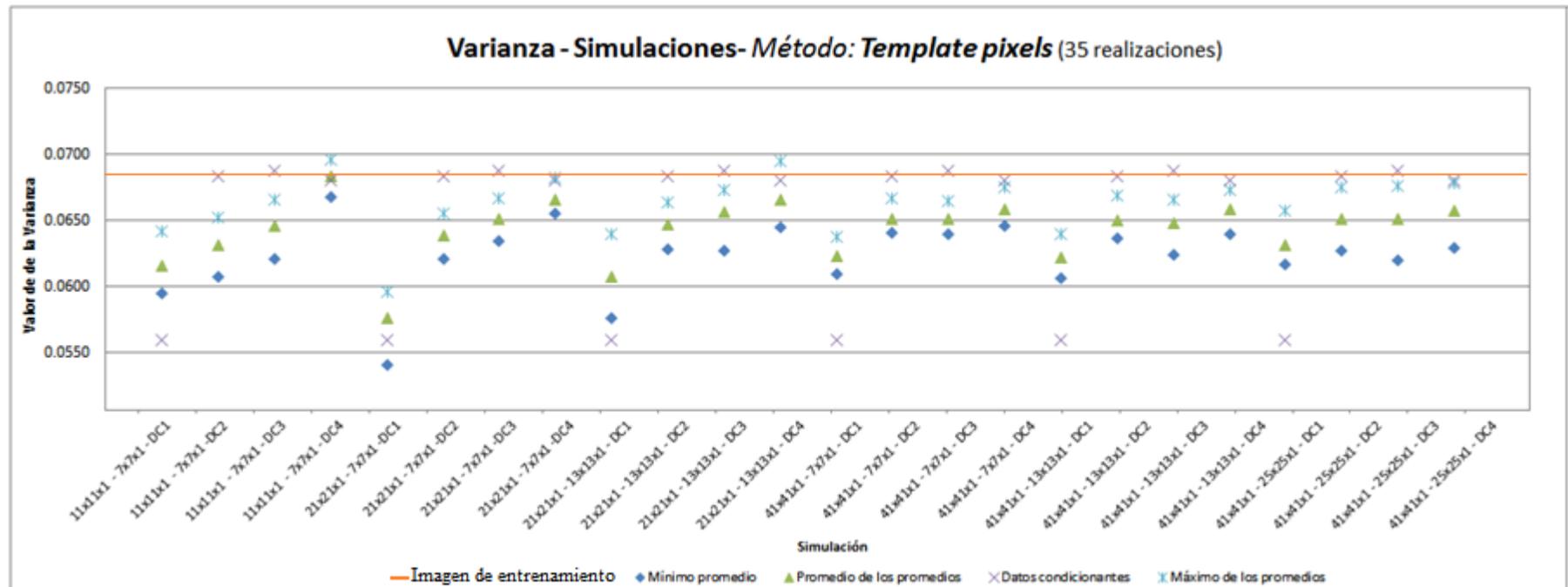


Figura 38 Varianza de los valores internos de las simulaciones – Método: *Template pixels*

Se observa un comportamiento similar al descrito en la figura 37. Todas las realizaciones se encuentran por sobre los datos condicionantes y la imagen de entrenamiento. No obstante, al utilizar *filter scores* por sobre *template pixel* no se percibe un impacto en la diferencia entre el máximo y el mínimo a la hora de aumentar los datos condicionantes.

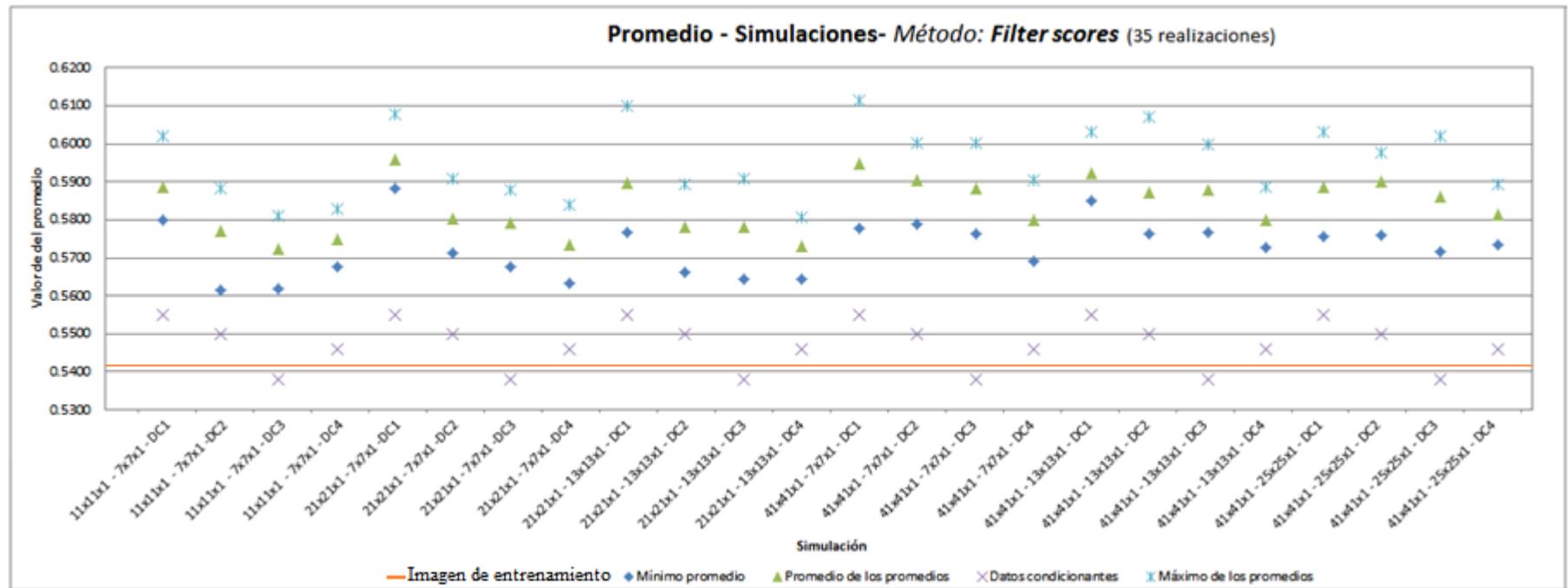


Figura 39 Promedio de los valores internos de las simulaciones – Método: *Filter scores*

A diferencia de lo observado en la figura 38, ahora en la figura 40 la cual considera como método *filter scores*, al considerar una mayor cantidad de datos condicionantes estos no afectan considerablemente el rango de dispersión entre el mínimo y el máximo valor promedio de la varianza. Esto se aprecia en la configuración $41 \times 41 \times 1 - 7 \times 7 \times 1$ y $41 \times 41 \times 1 - 25 \times 25 \times 1$.

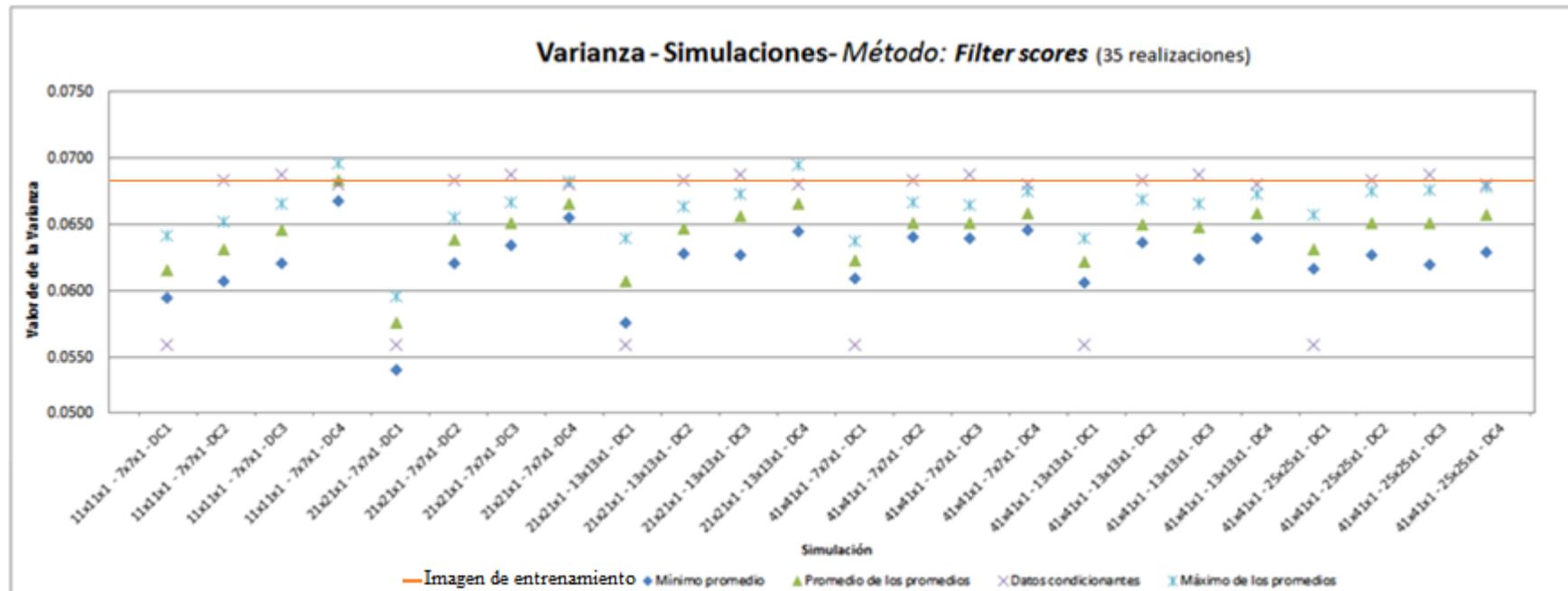


Figura 40 Varianza de los valores internos de las simulaciones – Método: *Filter scores*

4.5.4 Promedio y varianza sobre el promedio de las realizaciones en cada simulación

Los valores se encuentran detallados en las tablas del Anexo B.

4.5.4.1 Promedio

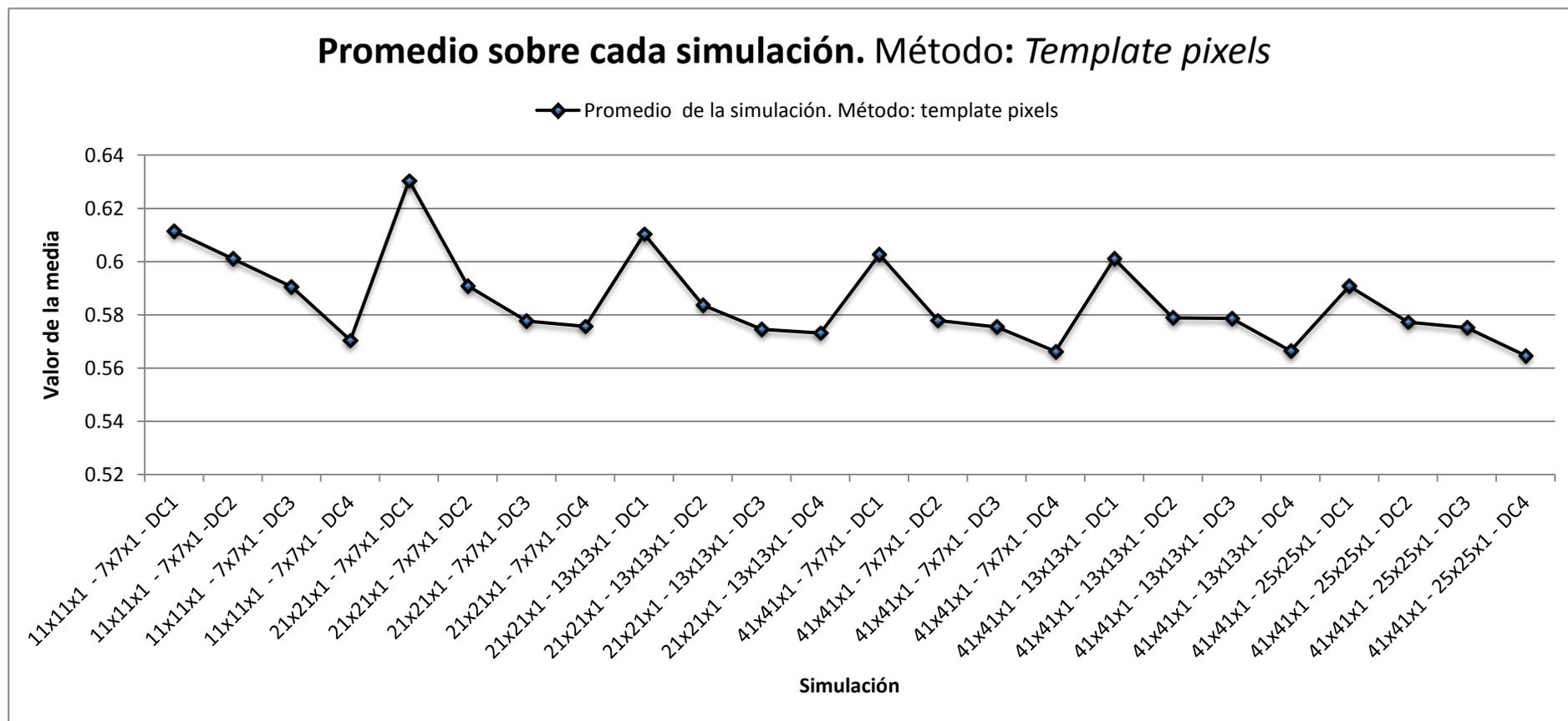


Figura 41 Promedio sobre cada simulación. Método: template pixels

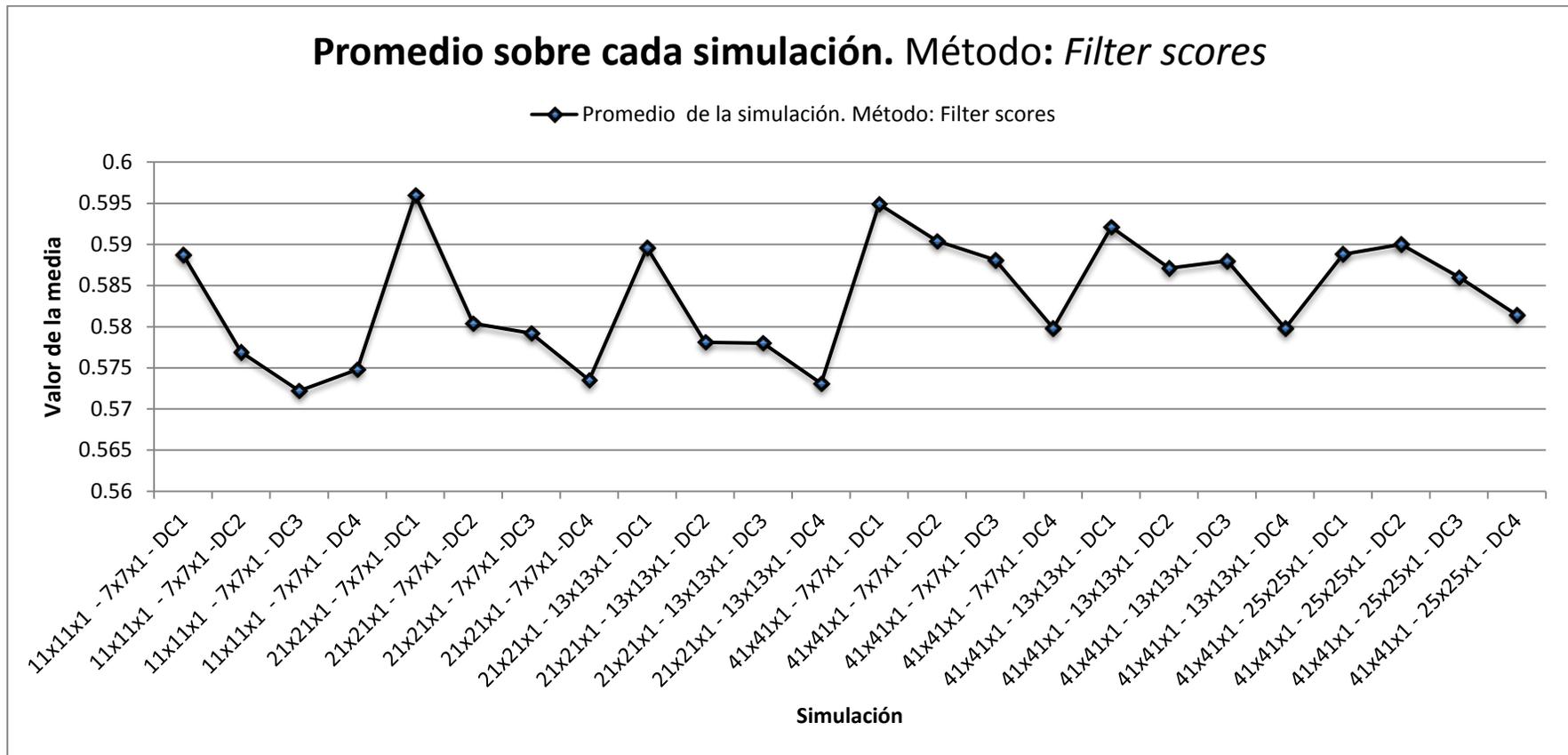


Figura 42 Promedio sobre cada simulación. Método: filter scores

Tanto en las figuras 41 como en la 42 se observa que al aumentar el tamaño de la grilla de búsqueda y la grilla interna se disminuye considerablemente el intervalo de los promedios sobre las simulaciones al considerar los cuatro conjuntos de datos condicionantes.

4.5.4.2 Varianza

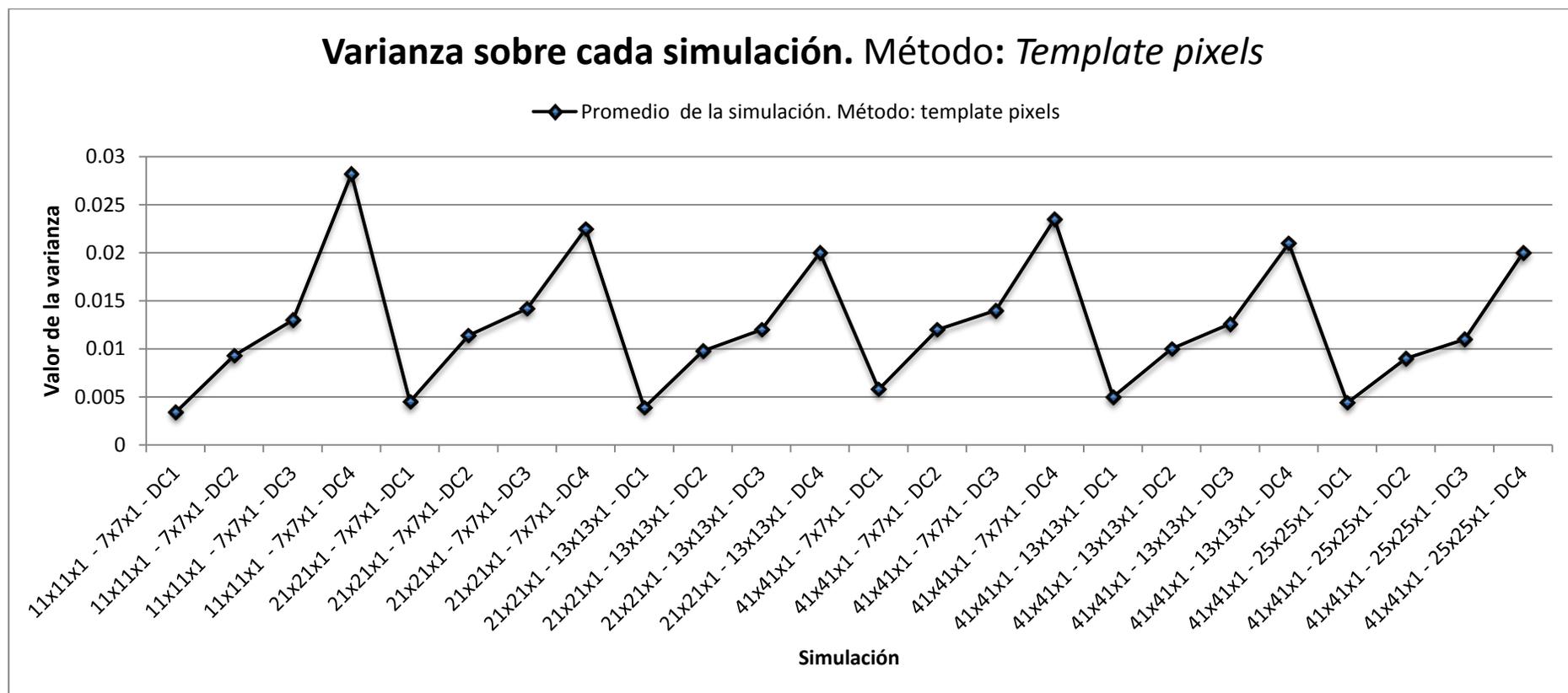


Figura 43 Varianza sobre cada simulación. Método: template pixels

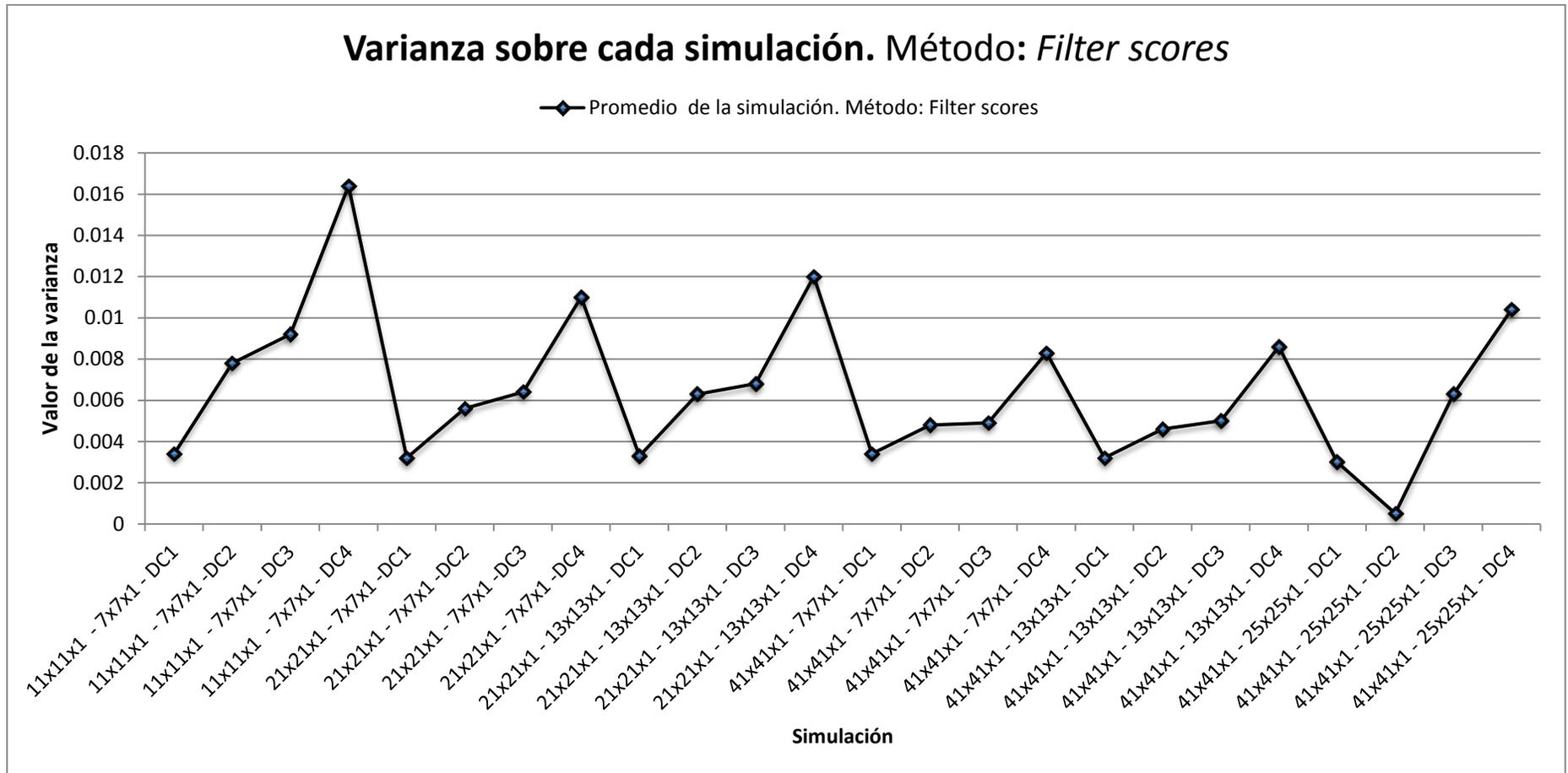


Figura 44 Varianza sobre cada simulación. Método: *filter scores*

Tanto en la figura 43 como en la 44 se observa que un aumento en la cantidad de datos condicionates lleva a un aumento en la varianza. Además, al utilizar *template pixels* el rango de varianza es más amplio que al utilizar *filter scores*.

4.5.5 Mapas probabilísticos

4.5.5.1 Tamaño grilla de búsqueda 11x11 y grilla interna de 7x7

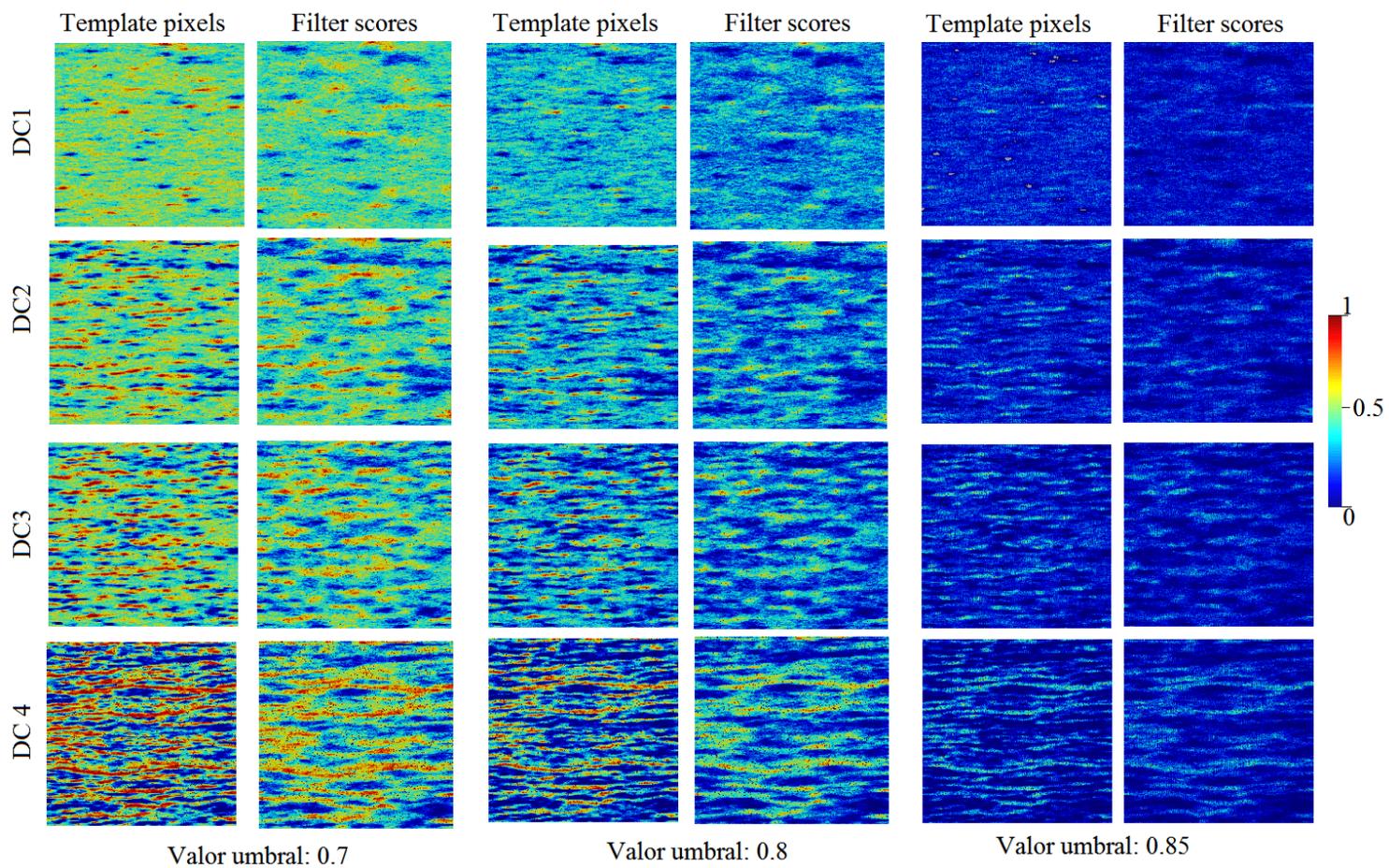


Figura 45 Mapas de probabilidad. Configuración 11x11 – 7x7

4.5.5.2 *Tamaño grilla de búsqueda 21x21 y grilla interna de 7x7*

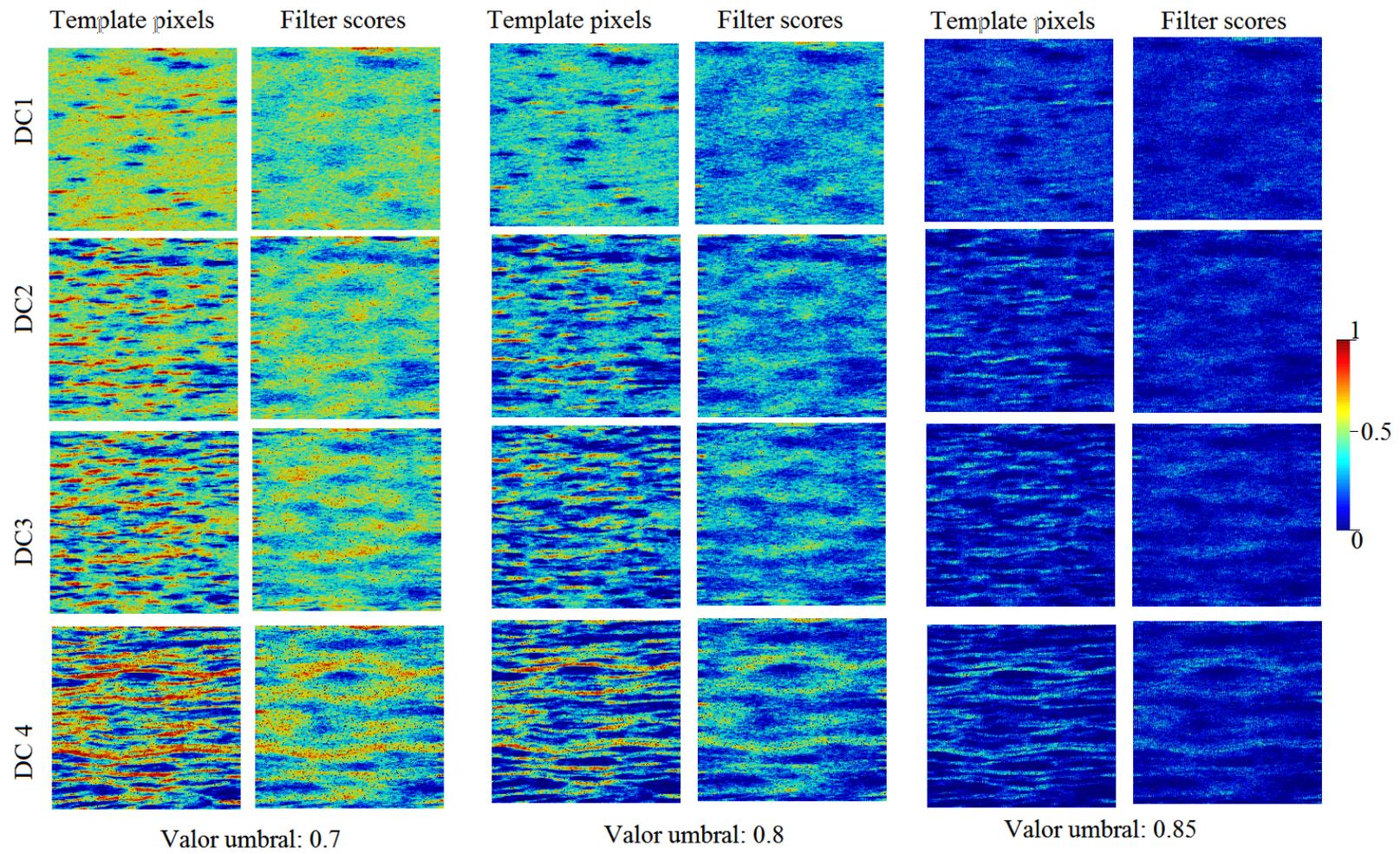


Figura 46 Mapas de probabilidad. Configuración 21x21 – 7x7

4.5.5.3 *Tamaño grilla de búsqueda 21x21 y grilla interna de 13x13*

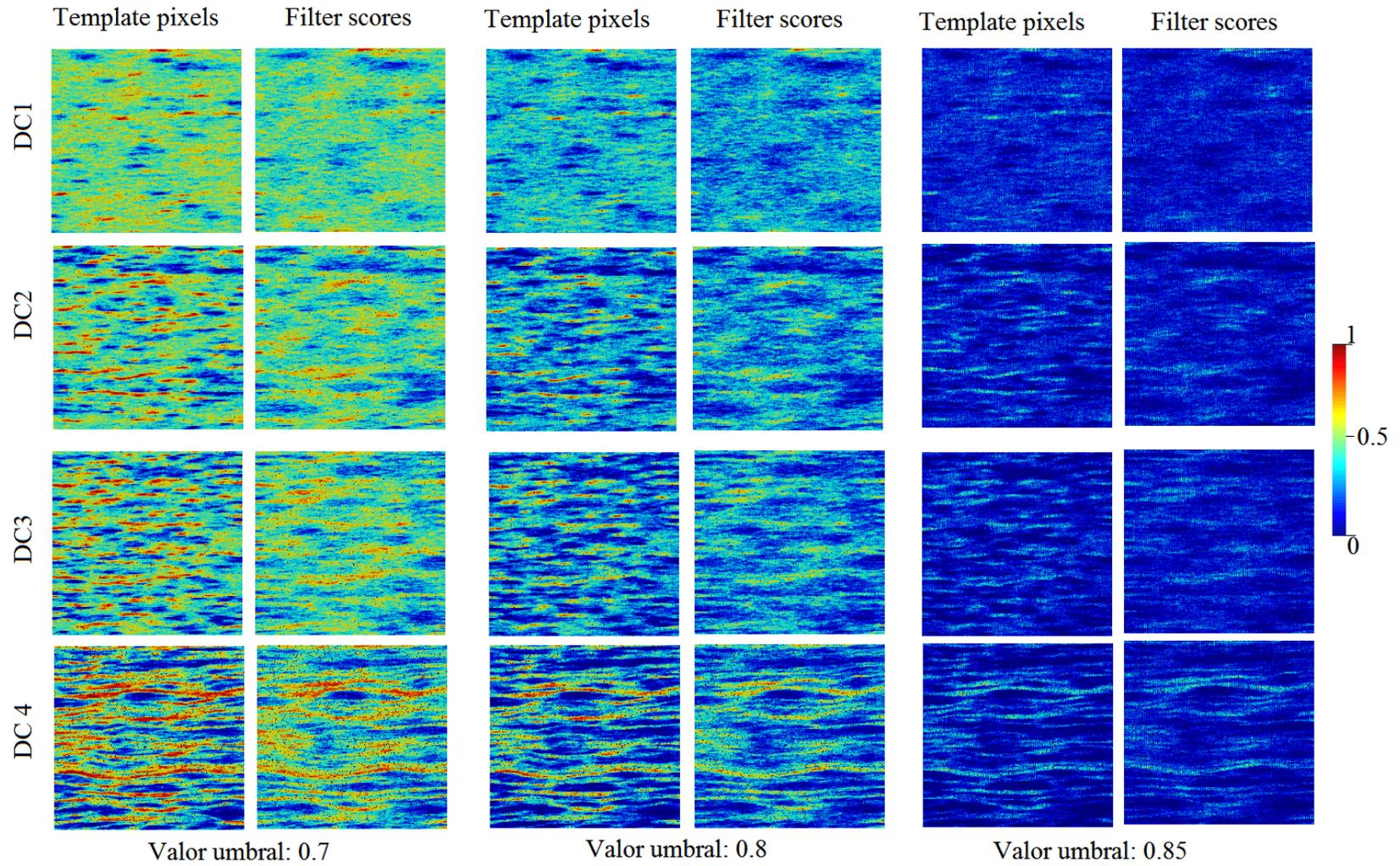


Figura 47 Mapas de probabilidad. Configuración 21x21 – 13x13

4.5.5.4 Tamaño grilla de búsqueda 41x41 y grilla interna de 7x7

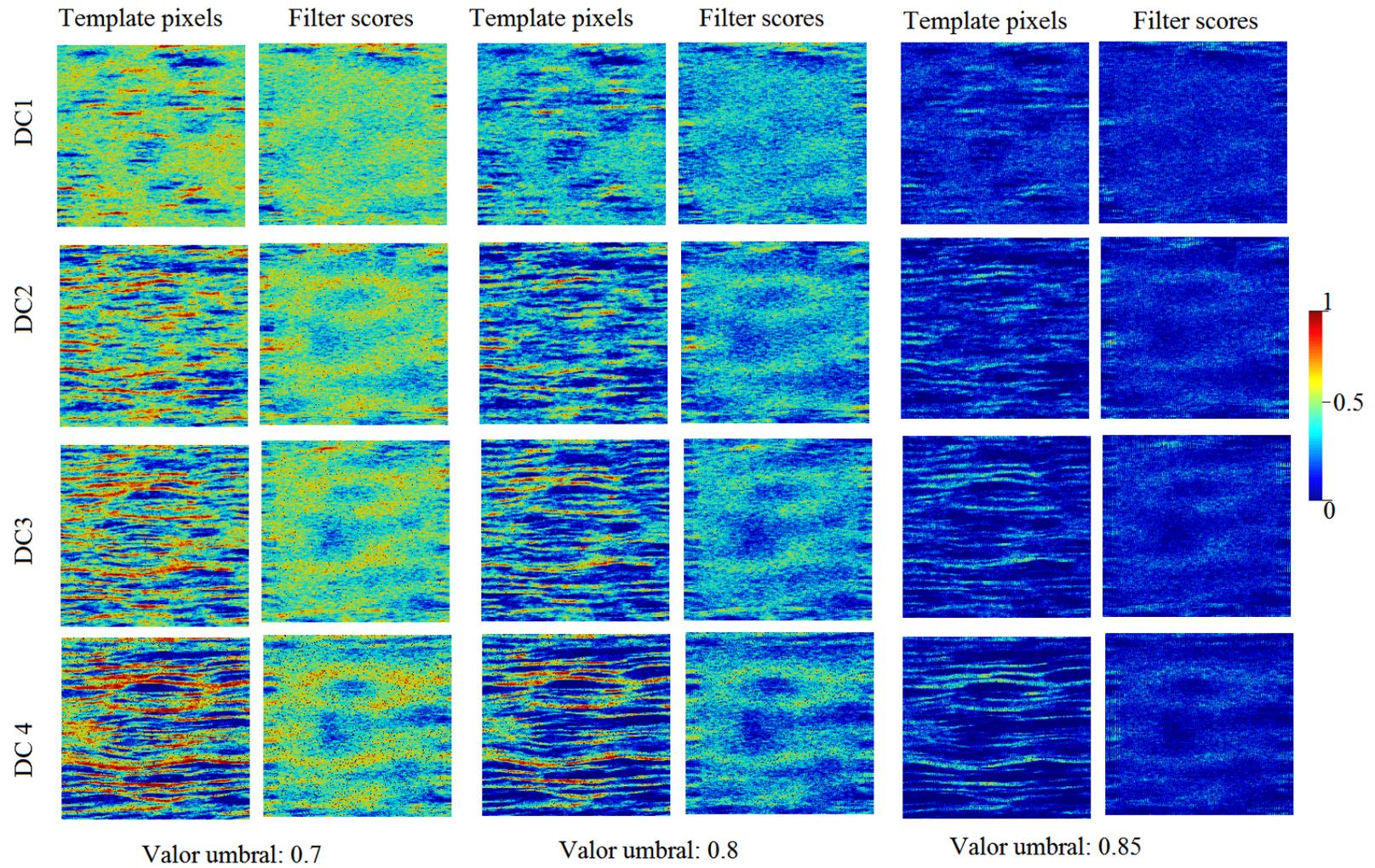


Figura 48 Mapas de probabilidad. Configuración 41x41 – 7x7

4.5.5.5 *Tamaño grilla de búsqueda 41x41 y grilla interna de 13x13*

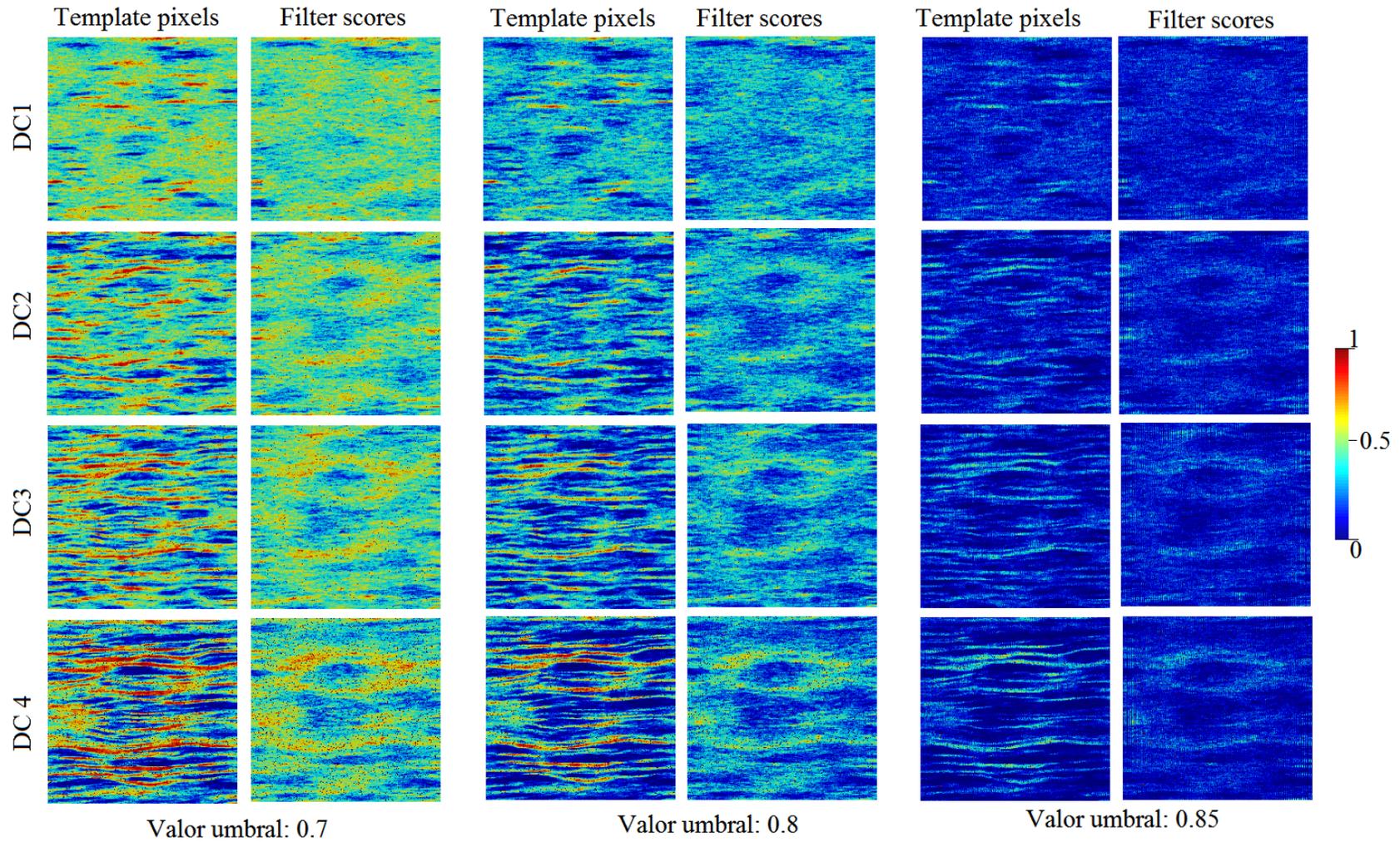


Figura 49 Mapas de probabilidad. Configuración 41x41 – 13x13

4.5.5.6 *Tamaño grilla de búsqueda 41x41 y grilla interna de 25x25*

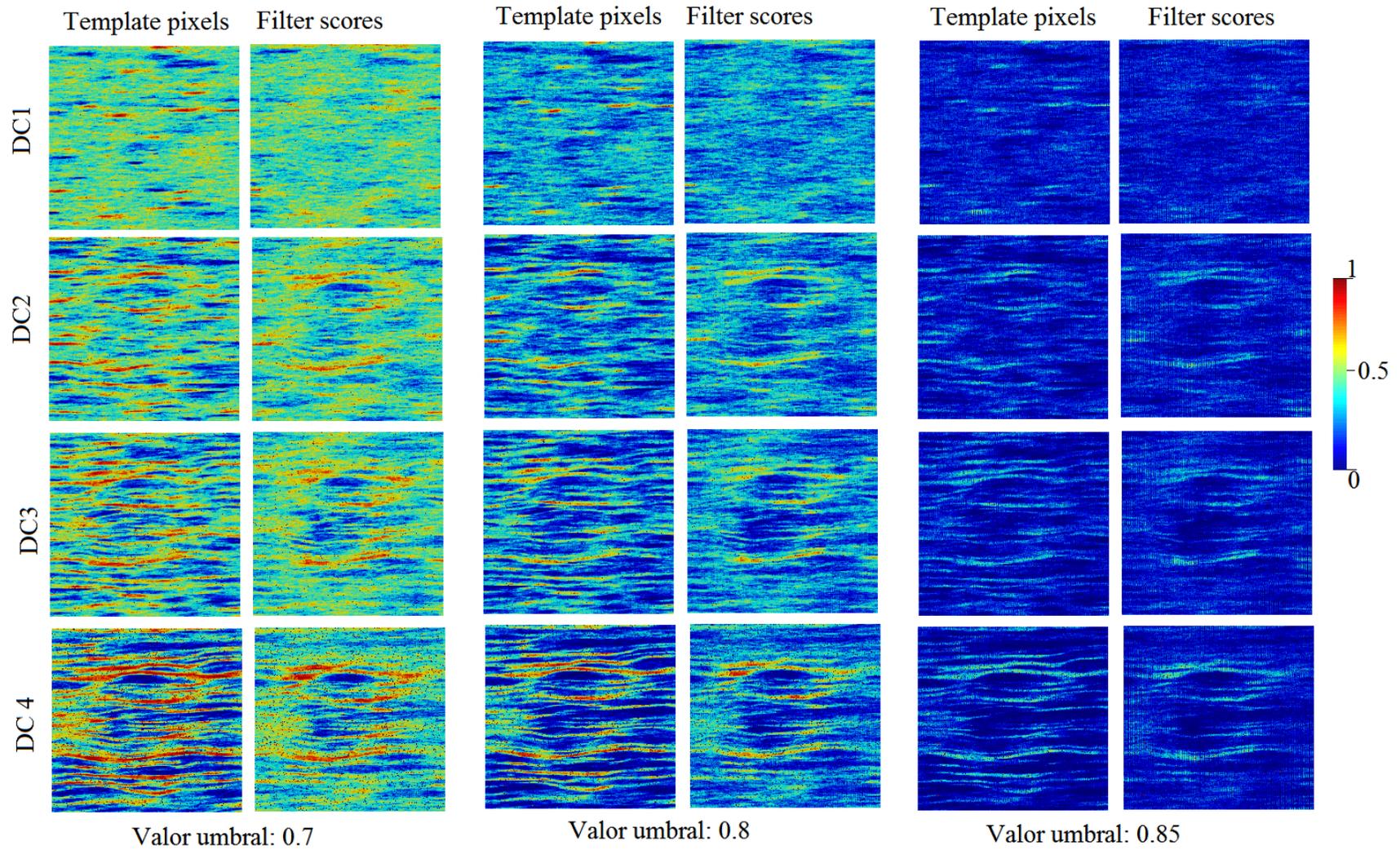


Figura 50 Mapas de probabilidad. Configuración 41x41 – 25x25

4.5.6 Análisis variográfico

Para validar los resultados de las simulaciones, es necesario comprobar que cada una de las realizaciones respete tanto la media-varianza como la correlación espacial de la función aleatoria generadora. En el caso de FILTERSIM, cada una de las realizaciones debe cumplir que el variograma de los datos sea lo más similar posible al variograma de la imagen de entrenamiento. Esto se debe a que el comportamiento de las variables regionalizadas es inferido del comportamiento de las variables en la imagen de entrenamiento. La figura 51 presenta los variogramas, para los 4 conjuntos de datos condicionantes, en las 2 principales direcciones de anisotropía (N5°O y N85°E) para 2 configuraciones (elegidas al azar): 11x11/7x7 y 41x41/25x25 y una sola realización de las 35. En la sección ANEXO C se presentan los variogramas para 1 realización por configuración.

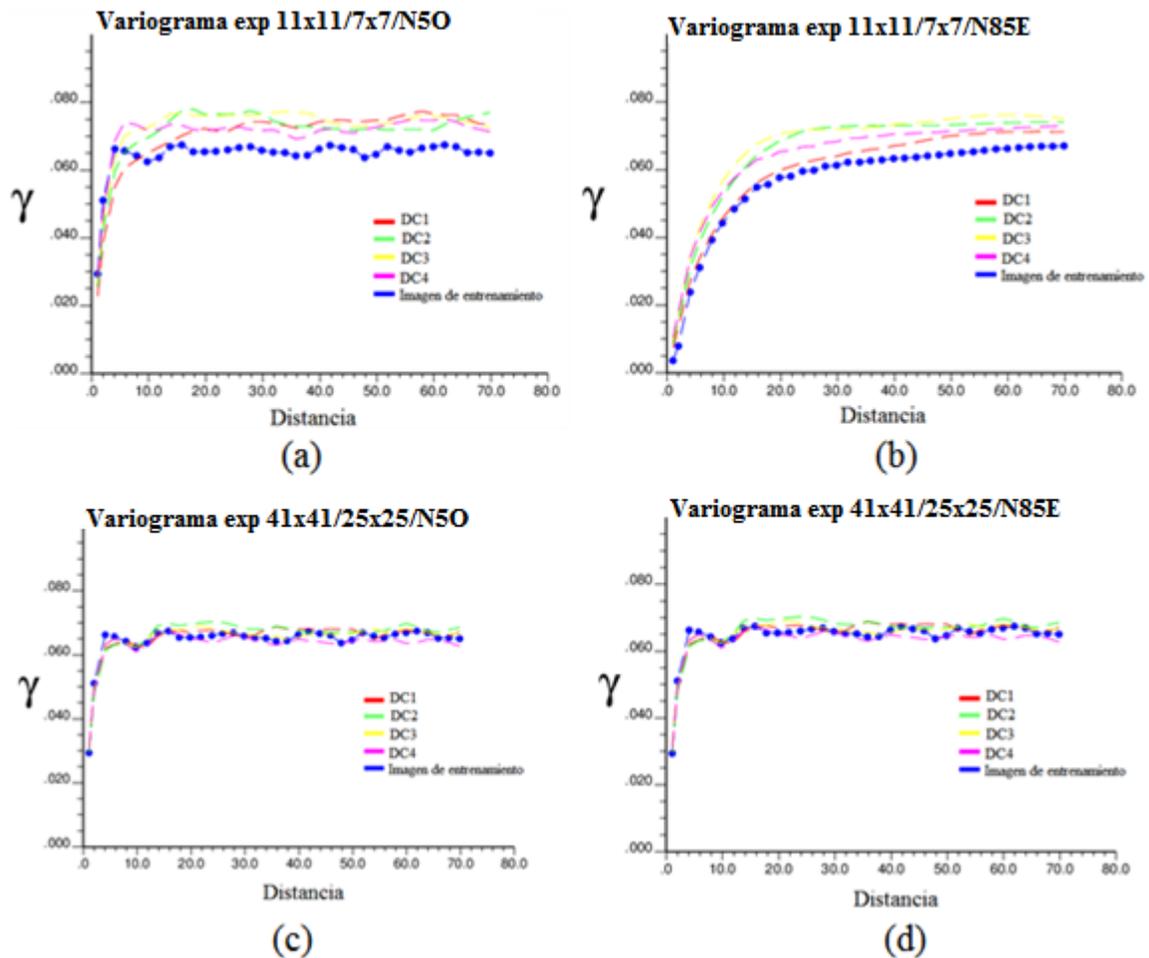


Figura 51 Variogramas experimentales sobre una sola realización. Configuraciones: 11x11/7x7 y 41x41/25x25.

4.6 Análisis de resultados

Los resultados demuestran que una de las 35 realizaciones dentro de cada configuración de parámetros es un caso posible de la realidad, ya que estas presentan una varianza en torno a la varianza de la imagen de entrenamiento y variogramas experimentales relativamente similares entre estos y la imagen de entrenamiento (alcance del variograma en 35 y meseta en 0.68 con un comportamiento de desestructuración de la variable similar).

Al analizar cada realización por separado, de manera visual, no es posible apreciar una continuidad espacial en la estructura, sin embargo, al realizar los distintos mapas de probabilidad se aprecia claramente la existencia de canales cuya probabilidad es relativamente alta respecto al entorno en donde el valor de la variable regionalizada sobrepasa los umbrales 0.7, 0.8 y 0.85. Más aun, es posible interpretar que el uso de la función *template pixels* para el cálculo del patrón más cercano al evento condicionante entrega canales de probabilidad mejor definidos a lo largo de la grilla de simulación en comparación al método *filter score* que presenta menores porcentajes de probabilidad y de manera más diseminada.

Respecto al tiempo de cómputo no se aprecia un cambio sustancial al incorporar una mayor o menor cantidad de datos condicionantes a la simulación. Sí se aprecia que el aumento en el tamaño de la grilla de búsqueda (dejando fijo el tamaño de la grilla interna) aumenta considerablemente el tiempo de cómputo, esto debido posiblemente al mayor tiempo que demora el proceso de *clustering* de los patrones inferidos de la imagen de entrenamiento. Por otro lado, al dejar fijo el tamaño de la grilla de búsqueda y aumentar el tamaño de la grilla interna el tiempo disminuye considerablemente y con una proporcionalidad inversa que establece: si el área de la grilla interna se cuadruplica (como en los casos de estudio) el tiempo de cómputo es cuatro veces menor, esto debido a que una mayor cantidad de nodos pasan de no tener información a tener información simulada, acortando la ruta de nodos a simular. Más importante aún, es la determinación del método para el cálculo del patrón más cercano ya que utilizar *template pixel* demanda entre 3 a 4 veces más tiempo que al utilizar *filter scores* bajo una misma configuración de parámetros.

Respecto al promedio de las 35 realizaciones: en un análisis visual es posible apreciar que fijando todos los parámetros y variando sólo el tamaño de la grilla de búsqueda se aprecia que su aumento lleva a una pérdida de información a pequeña escala del comportamiento de las

estructuras. Mientras que al fijar todos los parámetros de entrada y variar sólo la grilla interna se observa que su aumento de tamaño lleva a perder calidad en los comportamientos a pequeña escala.

Desde un análisis estadístico, el método *filter scores* presenta una menor dispersión en los valores de los promedios en comparación a *template pixels*. Además, *template pixels* disminuye el valor de la media al aumentar la cantidad de datos condicionantes bajo una misma configuración (ver figura 37) mientras que *filter scores* respeta la relación entre las medias de los datos condicionantes en sus resultados (ver figura 39).

Se decide no comparar los resultados directamente con la imagen de entrenamiento, ya que, a pesar de ser un caso sintético el algoritmo solo extrapola el comportamiento (continuidad) de la variable. Por lo tanto, si al simular con FILTERSIM no hay datos condicionantes o bien el 20% de la grilla está condicionada con algún valor, la calidad de los resultados se analiza desde un punto de vista de continuidad espacial de la variable y no de la exacta similitud a la imagen de entrenamiento.

Se observa en el promedio de las 35 realizaciones que la variable está considerablemente suavizada. Esto se debe principalmente a que el promedio de las simulaciones tiende a la estimación mediante kriging, la cual suaviza las leyes sobreestimando las bajas leyes y subestimando las altas.

Finalmente, se recomienda utilizar siempre 3 multigrillas por lo tanto se debe considerar que el largo de la grilla de búsqueda multiplicado por 2^2 debe ser igual o menor a la mitad del largo de la imagen de entrenamiento (y preferentemente de la grilla a simular también). Respecto a los parámetros, se prefiere *template pixel* aunque tome cuatro veces más tiempo si se desea obtener una estimación con continuidad espacial más marcada. Respecto al método de agrupamiento ambos métodos son aconsejados, la diferencia se genera en la velocidad de cómputo siendo considerablemente más veloz *cross partition* en desmedro de *k-mean partition*.

5 APLICACIÓN FILTERSIM CONVECCIONAL: CASO DE ESTUDIO REAL

El presente capítulo tiene por finalidad llevar las lecciones aprendidas en el capítulo anterior sobre que rango es más óptimo para ciertos parámetros, que funciones de distancia utilizar, etc. directamente a un caso de estudio real (asumiendo como *real* el hecho de trabajar con bases de datos provenientes de un yacimiento real). Se presenta como caso de estudio la simulación de bancos (en profundidad), de una mina a cielo abierto, condicionados a sondajes utilizando como imágenes de entrenamiento sus bancos anteriores (más superficiales) cuya distribución de leyes se asume representativa.

5.1 Metodología

El caso de estudio tiene como componentes:

1. Yacimiento cuprífero interpretado banco a banco a partir del muestreo de pozos de tronadura.
2. Muestras provenientes de sondajes como datos de condicionamiento para la estimación de las leyes banco a banco o del sector completo.

El análisis se lleva a cabo paso a paso tal como sigue:

1. Se detalla el yacimiento, utilizando la información desde los pozos de tronadura banco a banco.
2. Se asume que la información proveniente de los bancos superficiales es representativa del comportamiento de la mineralización de los bancos inferiores. Bajo este criterio se analizan las opciones de imágenes de entrenamiento (T.I.s) extraíbles.
3. Se establecen 3 T.I.s a utilizar en el proceso de simulación. Dado que la información de los pozos de tronadura están distribuidos de manera irregular en cada banco, una grilla regular de 10 x 10 es superpuesta y los datos son migrados a cada nodo para generar cada una de las T.I.s. Estas estarán completamente informadas y con suficientes datos para inferir los comportamientos de la mineralización.
4. Se detalla la información contenida en los sondajes. Las leyes de los sondajes son consideradas como datos condicionantes para el proceso de simulación.

5. Los sondeos asociados a los bancos inferiores son analizados para identificar 3 bancos a ser simulados. Los criterios son: número razonable de sondeos y una distribución espacial amplia.
6. Se asocian, de menor a mayor profundidad, las T.I.s con los bancos a simular. En otras palabras, cada T.I. está asociada a un determinado banco a simular.
7. Se realiza el proceso de simulación utilizando FILTERSIM, generando un total de 1,000 realizaciones para cada banco simulado.
8. Se analizan los resultados bajo los siguientes puntos de vista:
 - a. Mapas de probabilidad de obtener leyes sobre umbrales: 1.0 %, 1.5 %, y 2.0 %.
 - b. Se ilustran 6 realizaciones al azar con la finalidad de presentar posibles distribuciones espaciales de las leyes simuladas.
 - c. Se obtiene la ley promedio sobre cada escenario (caso posible de la realidad) para luego obtener el histograma de las leyes media sobre las 1,000 realizaciones.
9. Comparación de los resultados respecto a la estimación mediante Kriging. Se debe establecer el variograma experimental, luego modelar para generar una estimación mediante kriging ordinario.
 - a. Comparación entre la varianza de kriging y la varianza de cada punto sobre el total de escenarios simulados.
 - b. Comparación entre la media estimada en el banco y el histograma de medias de los escenarios.

5.2 Descripción del yacimiento

El yacimiento consiste en un cuerpo cuprífero cuyos márgenes son: 203 metros de profundidad, 837 metros de largo y 558 metros de ancho. La reconstrucción visual de la figura 52 se realiza con el apoyo de las muestras obtenidas a través de pozos de tronadura cuya distribución espacial es irregular. Se caracterizan 15 bancos que van desde lo más superficial a lo más profundo. (4024, 4000, 3988, 3964, 3952, 3940, 3928, 3916, 3900, 3884, 3868, 3852, 3850, 3836, y 3820).

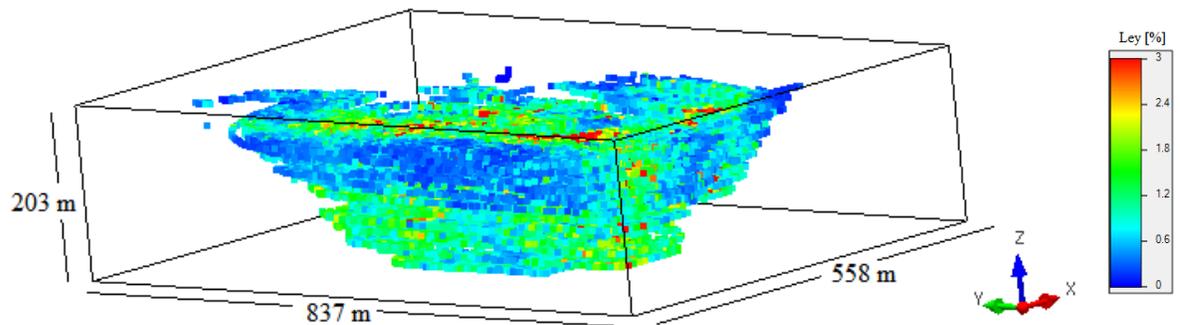


Figura 52 Visualización en 3 dimensiones del yacimiento completo

La figura 53 es una representación longitudinal del yacimiento completo (no un corte) en dirección YZ. Se observa los sectores poco mineralizados en superficie (azules) y medianamente mineralizados (verdes) en profundidad y hacia el sector norte.

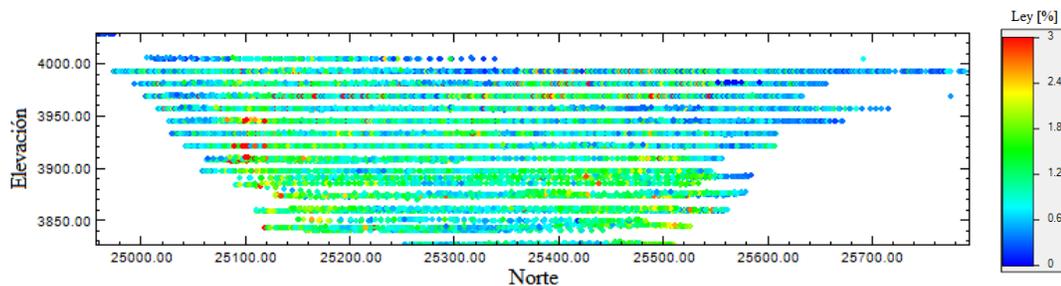


Figura 53 Vista YZ del yacimiento completo

Por su parte, la figura 54 es una representación transversal (no un corte) en dirección XZ. Nuevamente, se observa que en superficie hay presencia de bajas leyes (azul) mientras que en el profundidad se observan leyes medias (verdes). En este caso, las leyes medias en profundidad no tienen una preferencia específica a lo largo de la dirección Este.

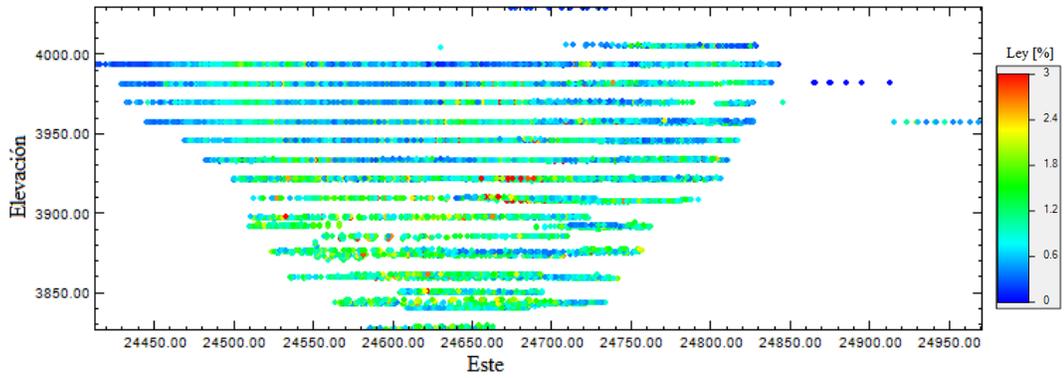


Figura 54 Vista XZ del yacimiento completo

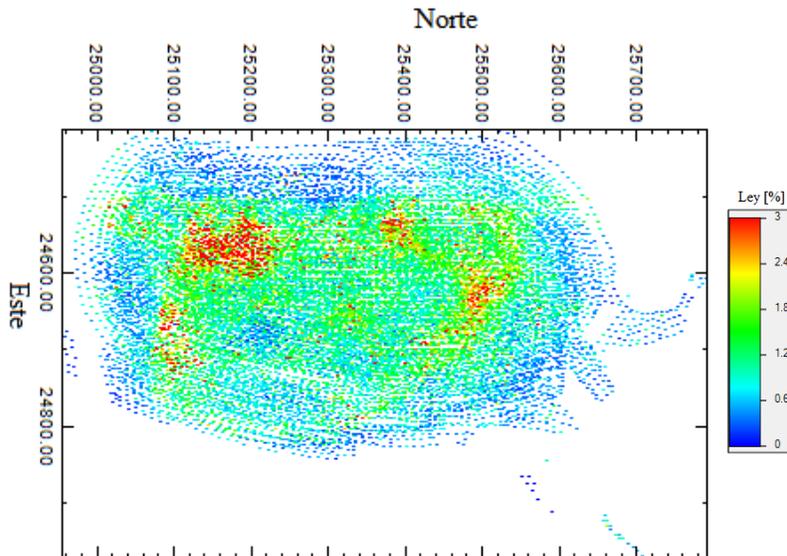
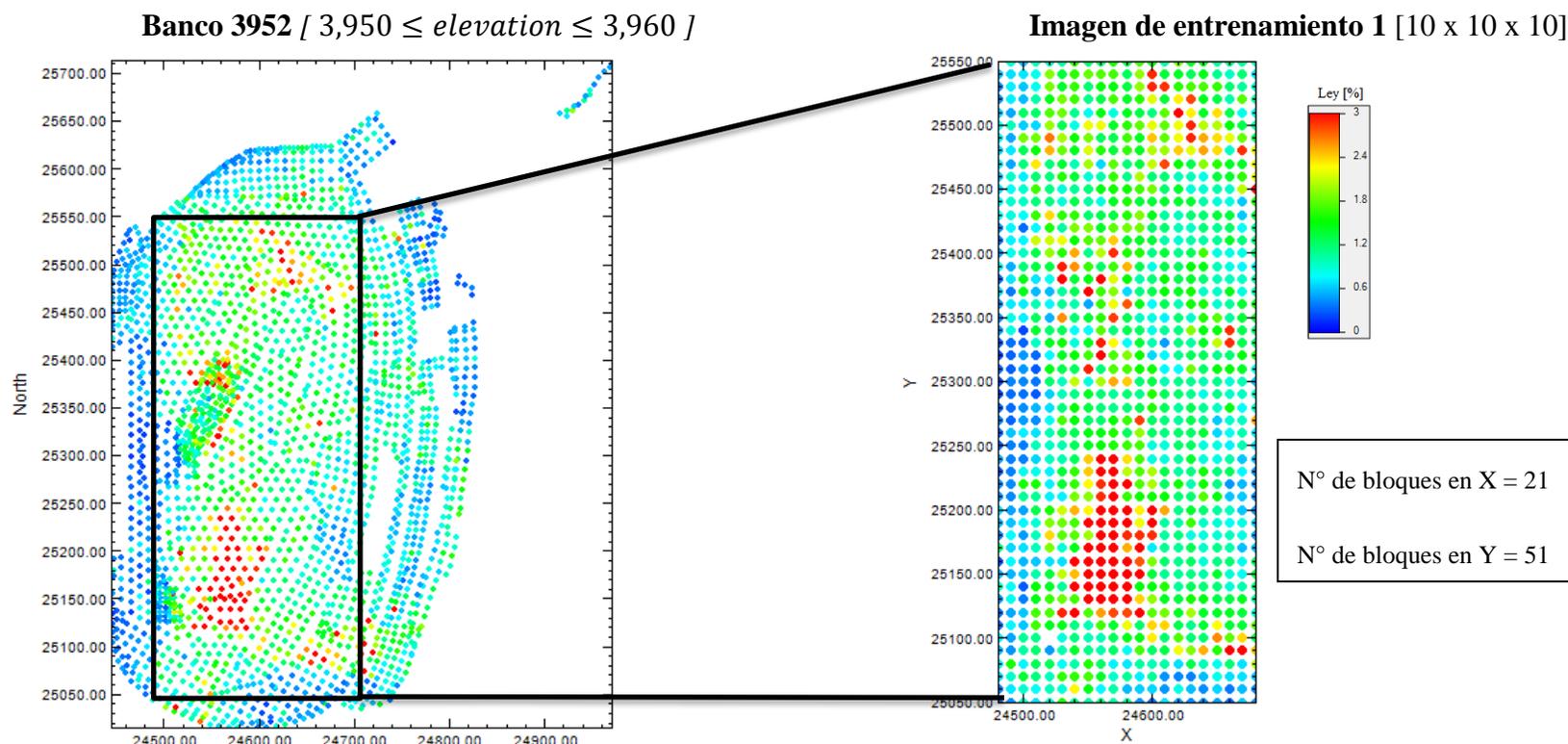


Figura 55 Vista XY del yacimiento completo

Por otro lado, la figura 55 presenta una vista en planta del yacimiento por completo. Se observa que en los sectores noreste y suroeste se concentran altas leyes (de color rojo). Cabe destacar, que todos los datos están sobre un mismo plano, es decir, no es un corte horizontal del yacimiento.

5.3 Selección imágenes de entrenamiento (T.I.)

Los bancos más superficiales son seleccionados para extraer las T.I.s. Los primeros 4 bancos presentan una gran cantidad de puntos no muestreados (ver anexo D) lo que imposibilita su utilización. Las figuras 56, 57, 58 muestran los bancos seleccionados.



Las imágenes de entrenamiento seleccionadas (3) provienen de los bancos 3952, 3940 y 3928. El resto de las posibles imágenes de entrenamiento se presentan en el Anexo D. La generación de las imágenes de entrenamiento a partir de los bancos se realizó mediante una

migración de los datos del banco sobre una grilla regular de 10 x 10 x 10. La técnica utilizada fue realizar una estimación del valor en el nodo dado el vecino más cercano con un radio de búsqueda de 5 x 5 x 5. Los criterios de selección fueron: obtener una grilla regular completa 100% informada (sin nodos vacíos) y con una distribución espacial de leyes que presenten un comportamiento característico (criterio sujeto a subjetividad).

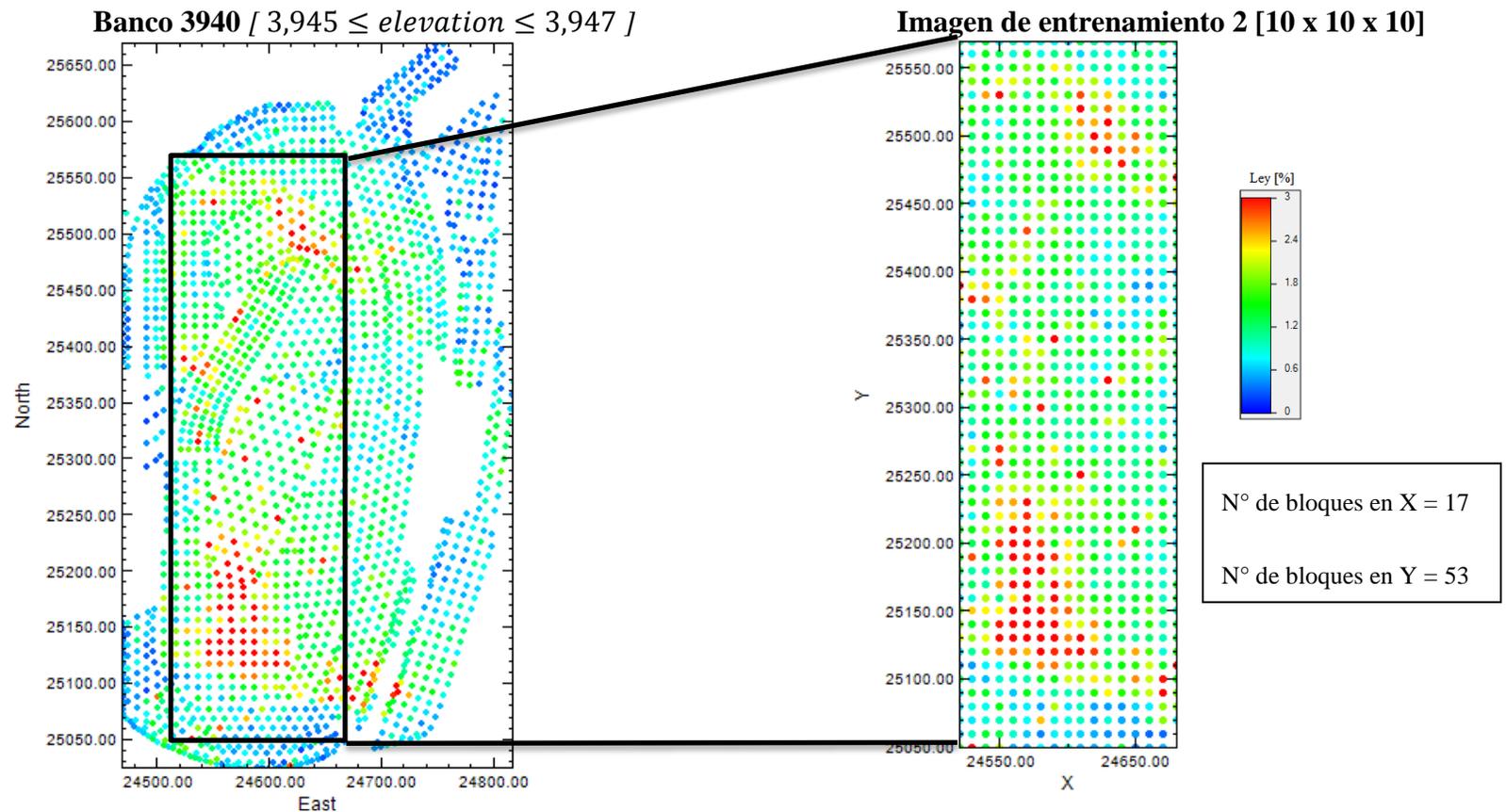


Figura 57 Selección segunda imagen de entrenamiento – Banco 3940

Banco 3928 [$3,925 \leq \text{elevation} \leq 3,935$]

Imagen de entrenamiento 3 [10 x 10 x 10]

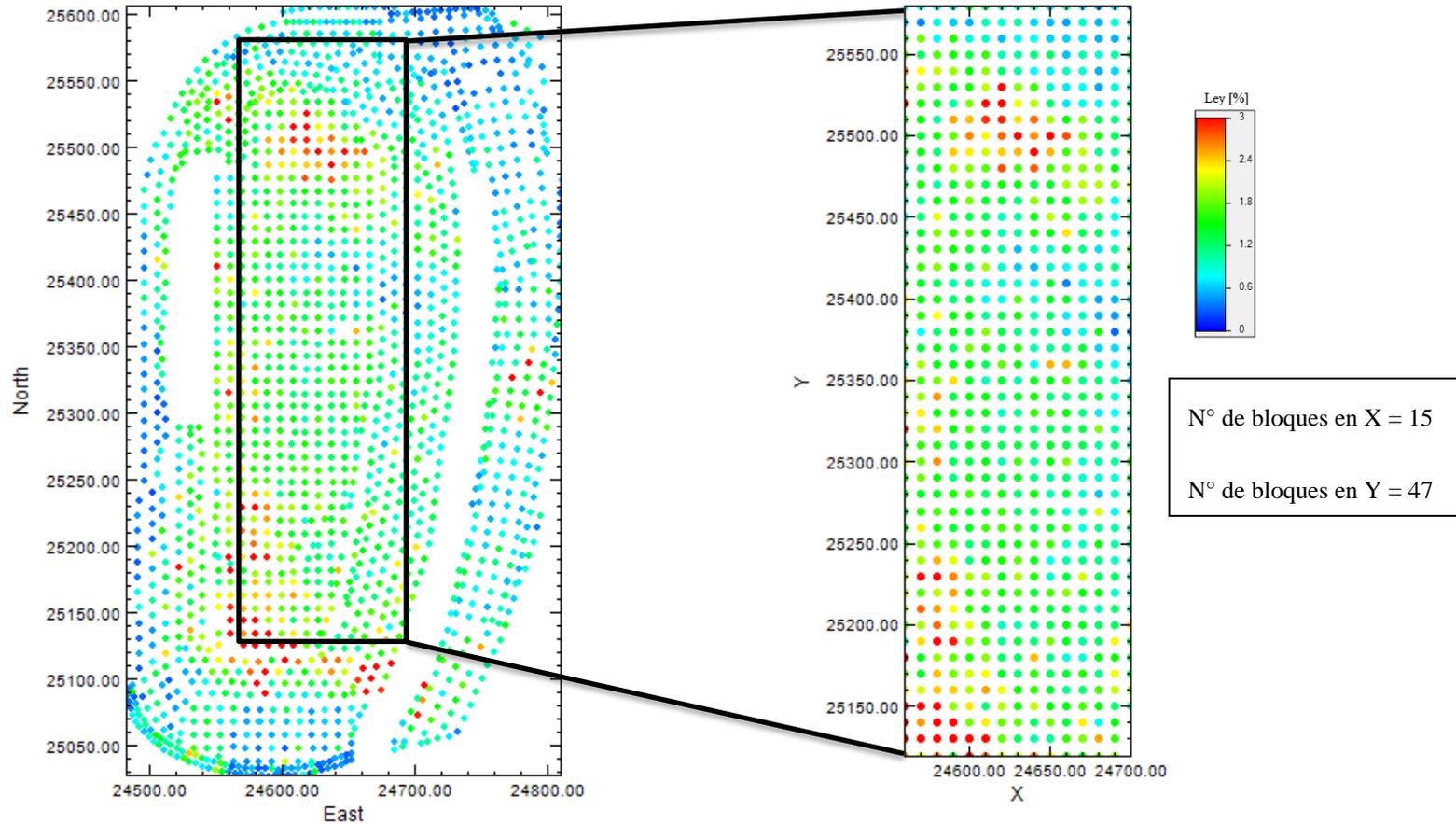


Figura 58 Selección tercera imagen de entrenamiento – Banco 3928

Finalmente, la primera T.I. tiene una mayor representatividad del comportamiento de las leyes bajas (azules) y la mayor cantidad de datos (1071 nodos), la segunda T.I. tiene presentada mayoritariamente valores medios-altos (verdes-rojos) con una cantidad de 901 nodos informados, mientras que la tercera T.I. tiene 705 nodos informados y una distribución similar a la segunda T.I. pero difieren espacialmente.

5.4 Selección de datos condicionantes: sondajes.

Tal como se presentó la descripción del yacimiento basado en los pozos de tronadura, esta sección presenta una visualización y descripción de los sondajes. La finalidad es determinar que bancos serán seleccionados para someterse a un proceso de simulación de leyes utilizando el algoritmo FILTERSIM. Cada banco a simular estará ligado a una sola T.I., dado que se asume que la representatividad de la mineralización no es tan solo a nivel horizontal sino también en profundidad. Por lo tanto, el banco a simular más superficial será simulado utilizando la T.I. más superficial, y del mismo modo con los otros dos bancos. La figura 59.a es una visualización de la totalidad de los sondajes en tres dimensiones mientras que la figura 59.b es una vista en planta de todos los sondajes sobre un plano.

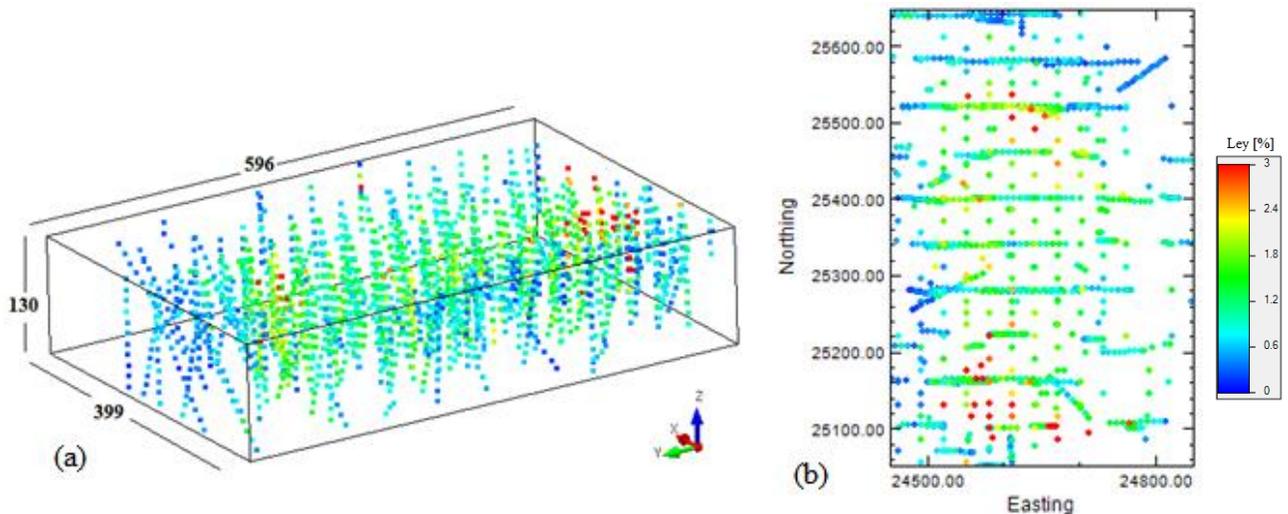


Figura 59 (a) Vista en planta de los sondajes, (b) vista en 3D de los sondajes.

Se observa tanto en las figuras 60 como 61 que la elevación de las muestras de sondajes es entre 3,820 y 3,950. El caso de estudio consiste en utilizar las imágenes de entrenamiento proveniente de los bancos superiores para luego simular las leyes de mineral en los bancos inferiores. Por lo tanto, la selección del banco a simular provendrá de las muestras de sondajes entre 3,820 y 3,890 asumiendo estos como los bancos inferiores.

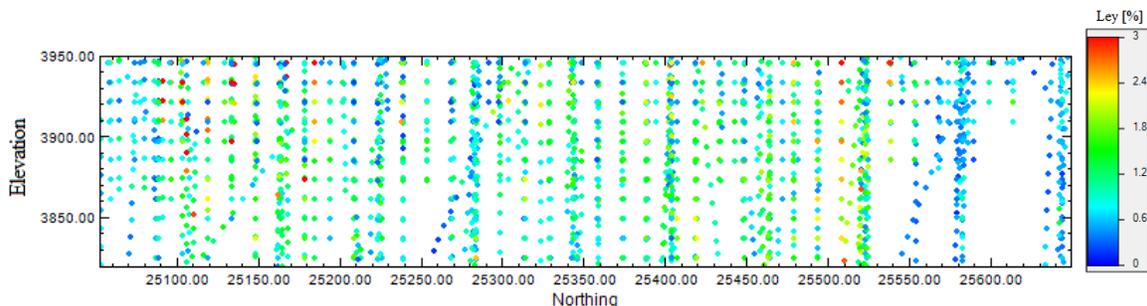


Figura 60 Vista lateral en dirección YZ del conjunto de sondajes

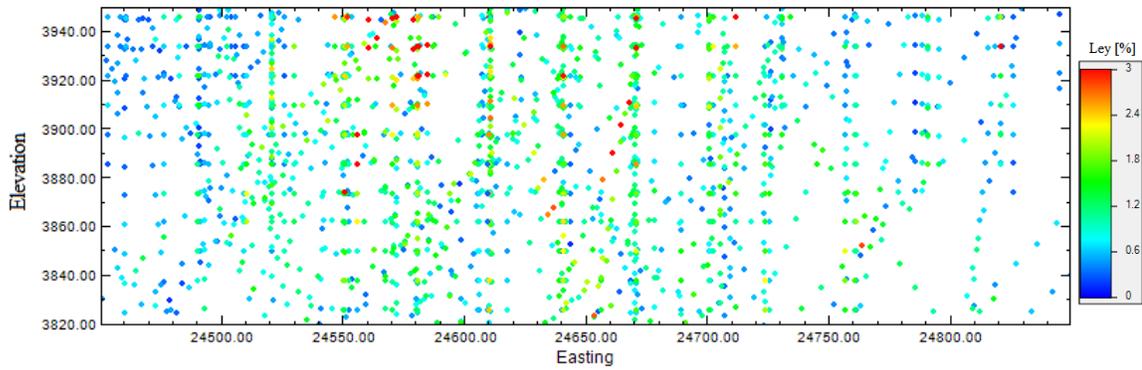


Figura 61 Vista lateral en dirección XZ del conjunto de sondajes

La figuras 62 y 63 presentan una vista en planta de cortes cada 10 metros entre 3,820 – 3,860 y 3,860 - 3,890 respectivamente. El propósito es presentar de manera visual la distribución espacial y la cantidad de muestras de sondajes asociadas a cada banco (asumiendo bancos de 10 metros).

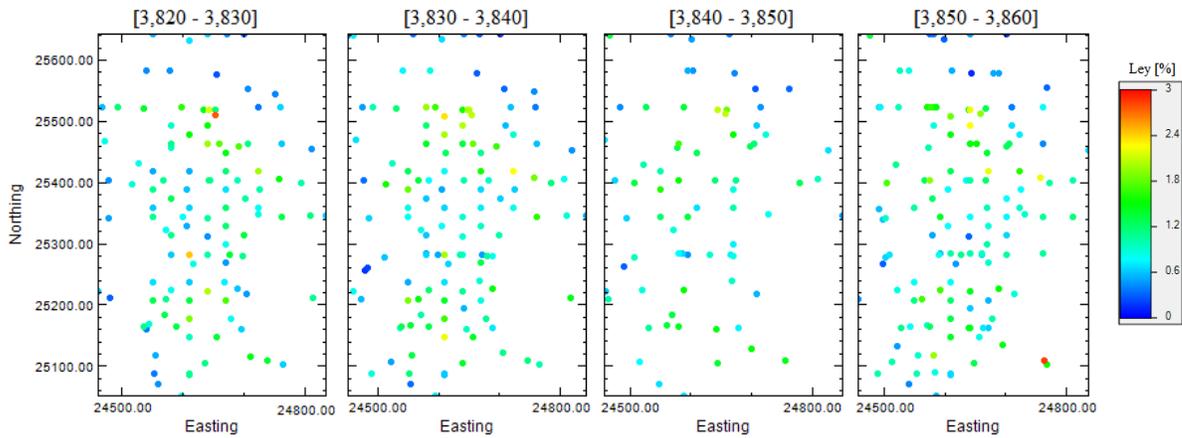


Figura 62 Vista en planta de sondajes en secciones de a 10 metros entre 3,820 y 3,860

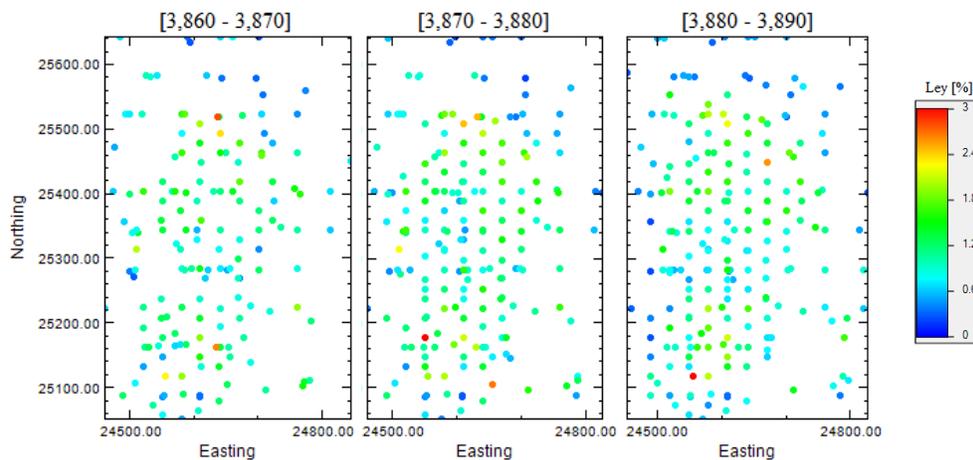


Figura 63 Vista en planta de sondajes en secciones de a 10 metros entre 3,860 y 3,890

Se observa en la figura 63 que los bancos entre 3,860 y 3,890 tienen mayor densidad de muestras de sondajes, mientras que la figura 62 presenta un sector, entre 3,840 – 3,850 con una baja cantidad de datos. El criterio para la elección del banco a simular tiende a ser mayormente subjetivo. No obstante, una cantidad adecuada de datos y una distribución espacial que cubran toda el área se asumen como condiciones necesarias. Varios bancos cumplen los requisitos anteriores ([3,820-3,830], [3,830-3,840], [3,850-3,860], [3,860-3,870]) de ellos se escogen los bancos [3,830-3,840], [3,850-3,860], y [3,860-3,870].

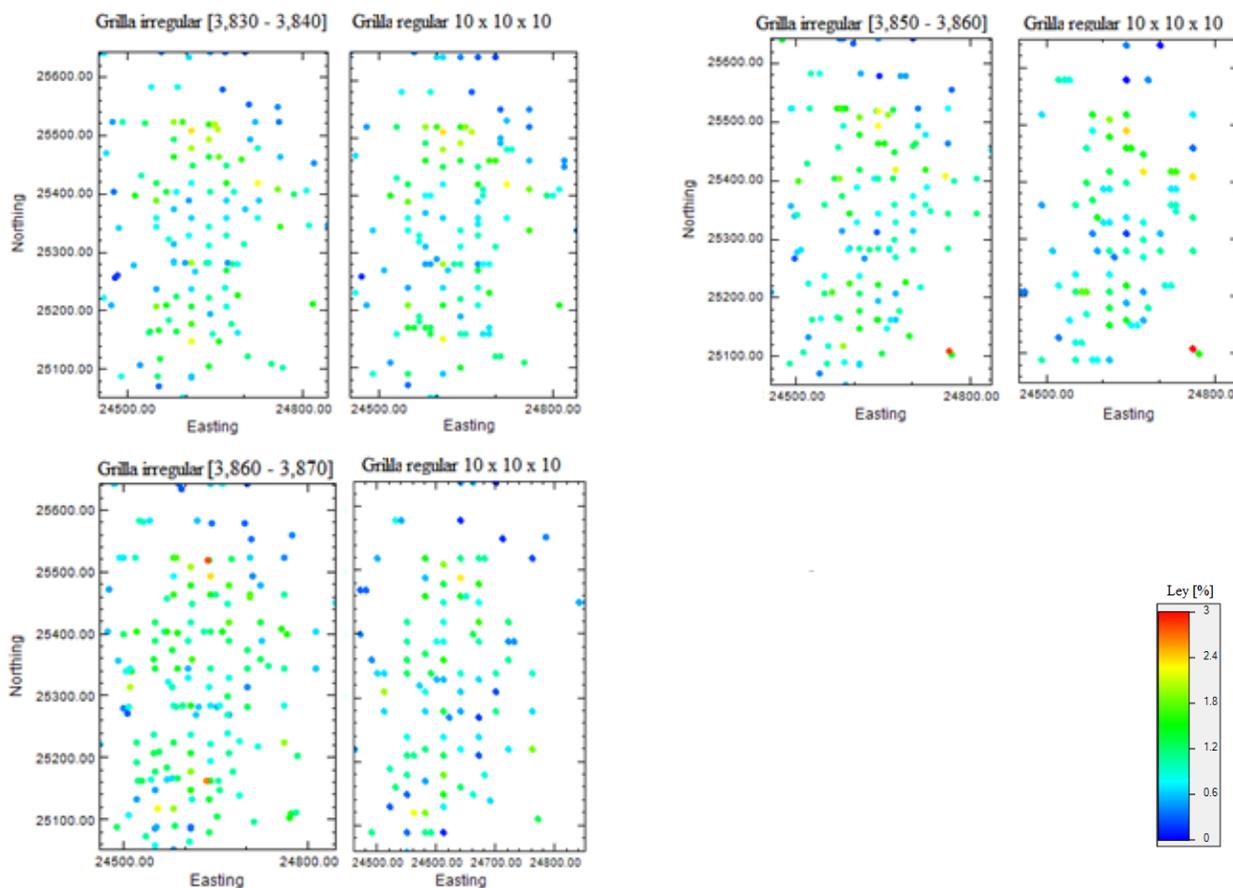


Figura 64 Bancos escogidos a simular. A la izquierda los datos en grilla irregular y a la derecha en grilla regular

Las muestras de sondajes vienen espaciadas de manera irregular. Para obtener resultados consistentes en la simulación los datos condicionantes deben migrarse a una grilla regular. Se utiliza una grilla de 10 x 10 x 10. El resultado final se observa en la figura 64, cada grilla regular se encuentra a la derecha de la grilla irregular.

Finalmente, se obtienen las siguientes asociaciones para el proceso de simulación:

- Imagen de entrenamiento 1 (banco 3952) asociada al banco [3,860-3,870].
- Imagen de entrenamiento 2 (banco 3940) asociada al banco [3,850-3,860].
- Imagen de entrenamiento 3 (banco 3928) asociada al banco [3,830-3,840].

El primer banco a simular consta de 40 bloques en el eje X [24,460-24,860], 59 bloques en el eje Y [25,060-25,650], y un solo bloque en el eje Z [3,860 – 3,870]. La totalidad de muestras de sondeos es de 101 y la distribución de leyes (histograma) es presentado en la figura 65.a.

El segundo banco a simular consta de 33 bloques en el eje X [24,450-24,780], 56 bloques en el eje Y [25,090-25,650], y un solo bloque en el eje Z [3,850 – 3,860]. La totalidad de muestras de sondeos es de 88 y la distribución de leyes (histograma) es presentado en la figura 65.b.

El tercer banco a simular consta de 40 bloques en el eje X [24,450-24,840], 61 bloques en el eje Y [25,050-25,650], y un solo bloque en el eje Z [3,830 – 3,840]. La totalidad de muestras de sondeos es de 137 y la distribución de leyes (histograma) es presentado en la figura 65.c.

El promedio de los datos condicionantes 1 es 1.03 %, el de los datos condicionantes 2 es 0.998 % y el de los datos condicionantes 3 es 1.02 %.

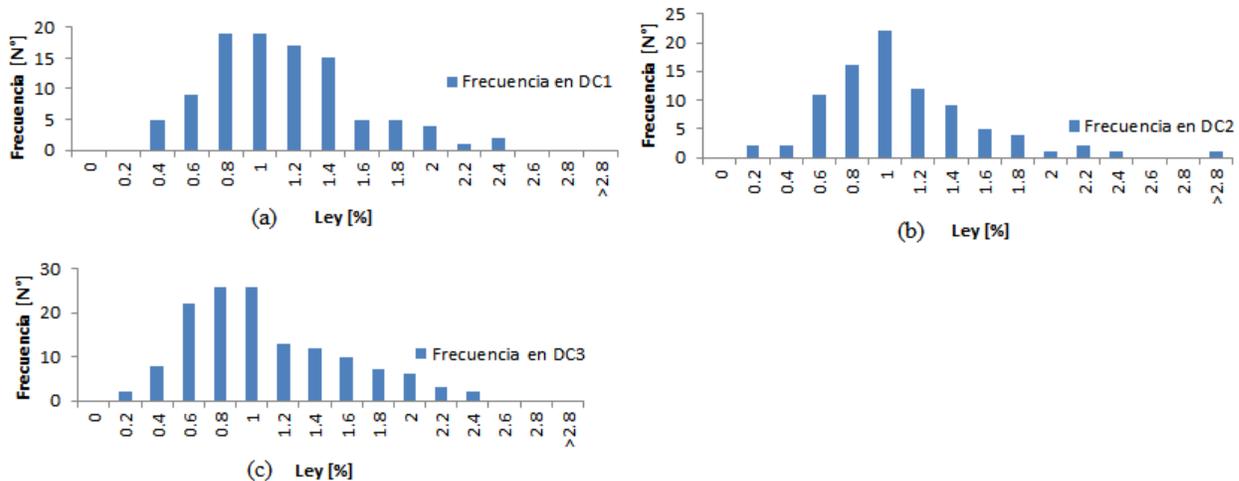


Figura 65 Histogramas de leyes en los datos condicionantes. (a) DC1, (b) DC2, (C) DC3

5.5 Simulación

Ya se cuenta con 3 T.I.s y un conjunto de datos condicionantes, elementos necesarios y suficientes para la realización de una simulación sobre una grilla regular definida utilizando el algoritmo FILTERSIM. Por cada T.I. se obtienen 1,000 realizaciones, es decir, 1,000 posibles escenarios de distribución de leyes. Luego, se someten las 1,000 realizaciones a un post-proceso de simulación. Este post-proceso consiste en estudiar la distribución de leyes medias, la varianza en la ley simulada en cada punto, y la generación de mapas de probabilidad de sobrepasar leyes umbrales: 1.0 %, 1.5% y 2.0%.

5.5.1 Parámetros de entrada FILTERSIM

Tal como se explicó en la sección 2.2.2, hay una serie de parámetros necesarios para el correcto funcionamiento de FILTERSIM. El valor de estos se fija basados en los resultados del caso sintético, sección 4.6, y corresponden a:

- Grilla a simular: elegir uno de los 3 bancos
- Imagen de entrenamiento: T.I. 1, ó T.I. 2, ó T.I. 3.
- Número de realizaciones: 1,000.
- Semilla: 211,175, ya que debe ser un número impar y suficientemente grande.
- *Search template dimension* (grilla de búsqueda): 9x9x1.
- *Inner patch dimension* (grilla interna): 7x7x1.
- Ajustar histogramas de los resultados respecto al histograma de la T.I.: no.
- Datos condicionantes: sondeos expuestos en la figura 59.b.
- Utilización de datos suaves o regiones: no.
- Número de multigrillas: 3, para que la grilla de mayor tamaño logre estar dentro de la grilla a simular.
- Mínimo número de réplicas por cada multigrilla: 10 para cada una de las multigrillas.
- Peso de los datos duros, simulados y otros: 0.5, 0.3, y 0.2. Al solo utilizar datos duros (condicionantes) y luego valores simulados, la ponderación real es 0.625 (0.5/0.8) para los datos duros y 0.375 (0.3/0.8) para los datos suaves.
- Tipo de partición: K-mean partition.

- Máximo número de grupos en el proceso K-mean: inicialización en 200 y 2 particiones secundarias máximo.
- Método para el cálculo de la distancia entre el evento condicionante y el patrón más cercano: *template pixel*. Se prefiere *template pixel* por sobre *filter score* dado que los resultados simulados en el caso de estudio anterior muestran mayor continuidad espacial de las estructuras a pesar de que el tiempo de cómputo es hasta 4 veces mayor.
- Tipo de filtros: filtros por defecto.

Con todo lo anterior, se procede a realizar los 3 procesos de simulación.

5.5.2 Resultados utilizando la T.I. 1 sobre el banco 3860

Al realizar las simulaciones, utilizando la imagen de entrenamiento 1, se busca identificar cuáles son los escenarios posibles de mineralización en el banco 3,860-3,870. Se determina la ley media de cada uno de los 1,000 escenarios y se obtiene un histograma de ellas (figura 66). La figura 67 entrega una visualización de 6 posibles distribuciones espaciales. Se observa que en variadas ocasiones se genera una acumulación de leyes altas en ciertos sectores, o bien una distribución más uniforme y aleatoria de estas. Otra buena herramienta de visualización son los mapas de probabilidad. Estos establecen un valor umbral en la ley y cuantifican la cantidad de veces que cada punto sobrepasa un valor y calcula su frecuencia de ocurrencia respecto al total de realizaciones. De esta forma se obtiene para cada punto (bloque) en la grilla simulada (banco) la probabilidad de pasar el valor umbral. La figura 68 presenta los mapas de probabilidad bajo los umbrales 1.0 %, 1.5 %, y 2.0 %, de izquierda a derecha



Figura 66 Histograma leyes media - Banco 3860

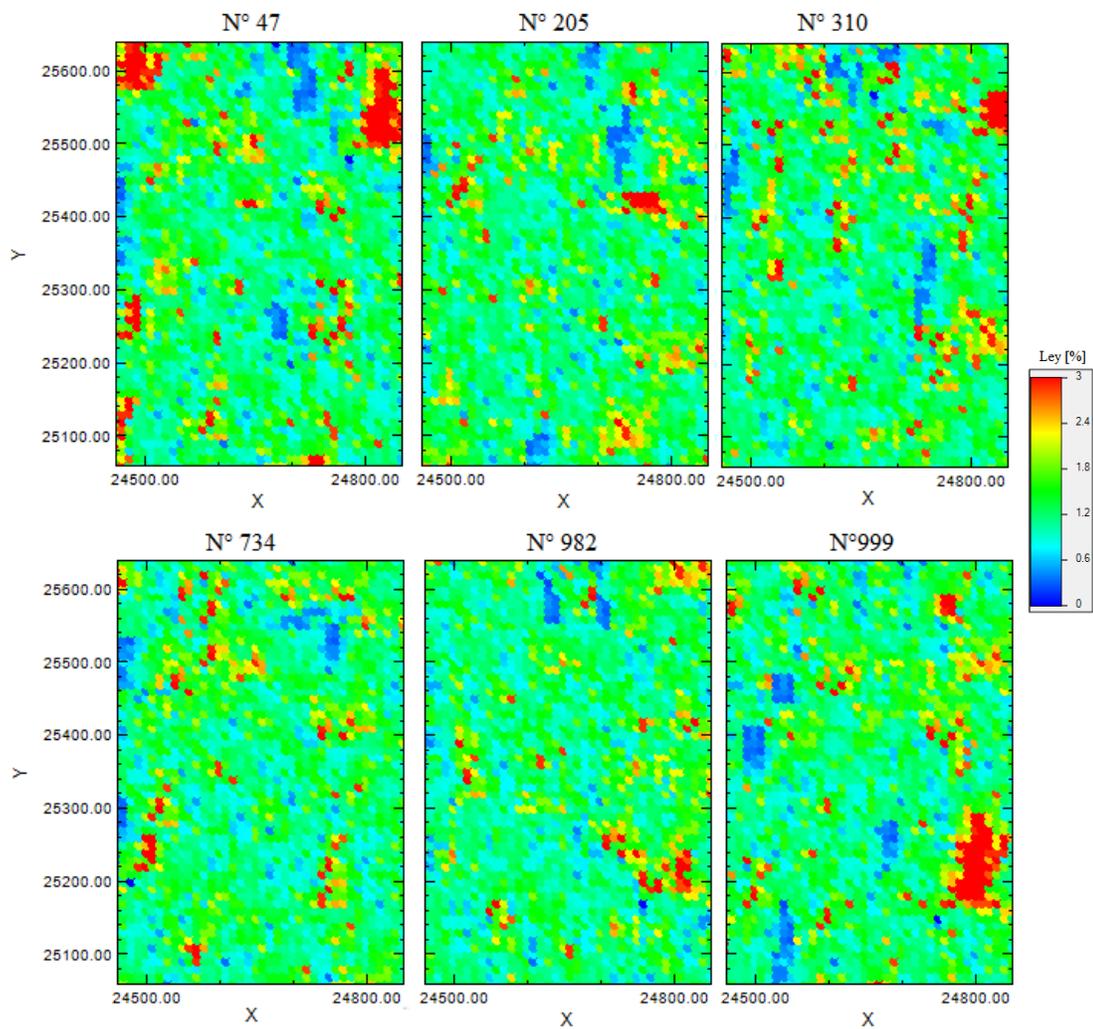


Figura 67 Seis realizaciones sobre el banco 3860

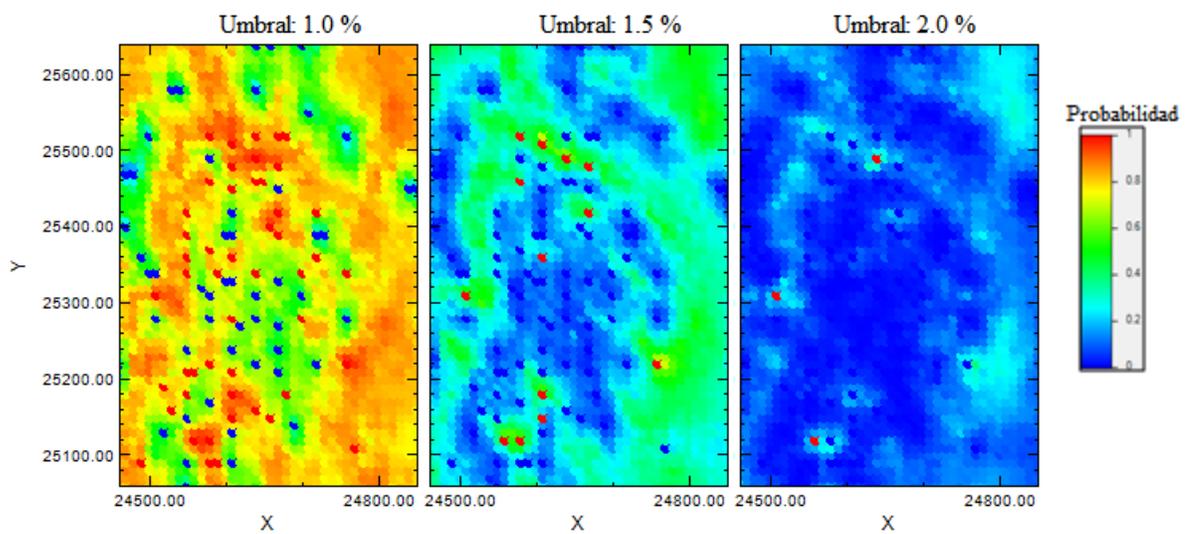


Figura 68 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3860

5.5.3 Resultados utilizando la T.I. 2 sobre el banco 3850

Tal como se realizó en el banco 3860, se procede ahora en el banco 3850. Sobre cada una de las 1,000 realizaciones se obtiene el valor promedio de la ley en el banco. Con la idea de obtener cual es la media del banco, bajo cierto intervalo de confianza, es que se genera el histograma de las leyes medias ilustrado en la figura 69.

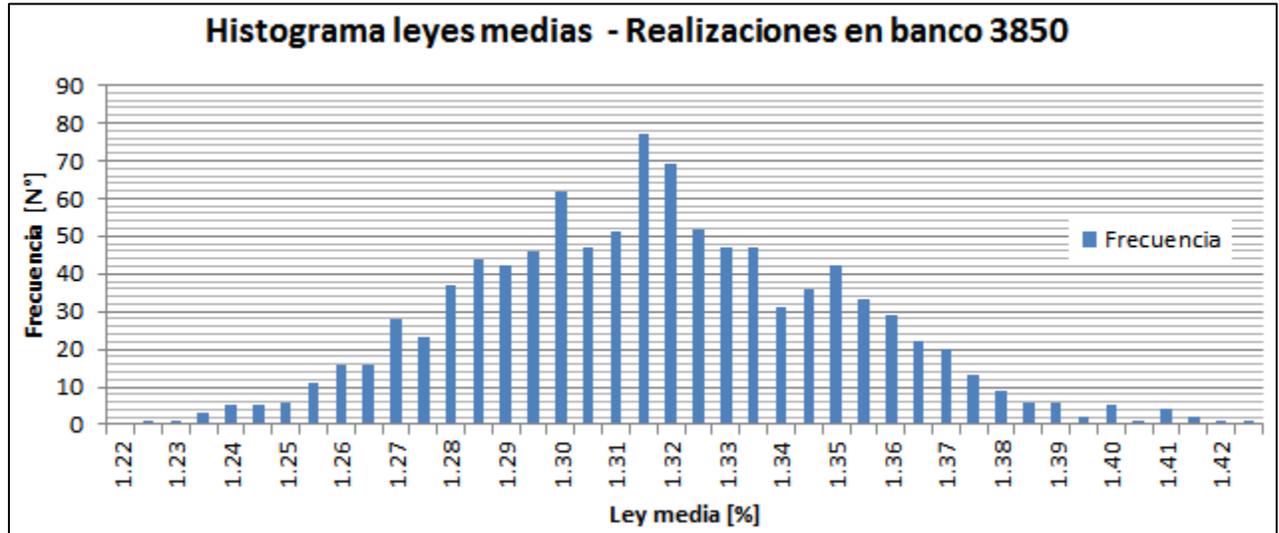


Figura 69 Histograma leyes media – Banco 3850

Resulta interesante notar que las realizaciones por separado, tal como se ilustra en la figura 70, presentan sectores preponderantes a leyes medias (celeste-verde) y pequeños sectores de altas leyes (rojos) de manera aleatoria. Dado que no es posible asegurar directamente la distribución espacial de las leyes en el banco 3850 es que se decide utilizar los mapas de probabilidad para obtener los sectores con mayor tendencia a sobrepasar un umbral determinado. De hecho, se observa en la figura 71 que hacia el sector Noreste (X [24,620 – 24,720], Y[25,400 - 25,540]) y hacia el sector Sureste (X[24,700 – 24,780], Y[25,100 – 25,160]) hay una alta tendencia a leyes sobre 1.5 %. Esto no es directamente inferido de una realización, por lo que los mapas de probabilidad juegan un rol importante al momento de decidir qué sectores explotar antes y cuales después.

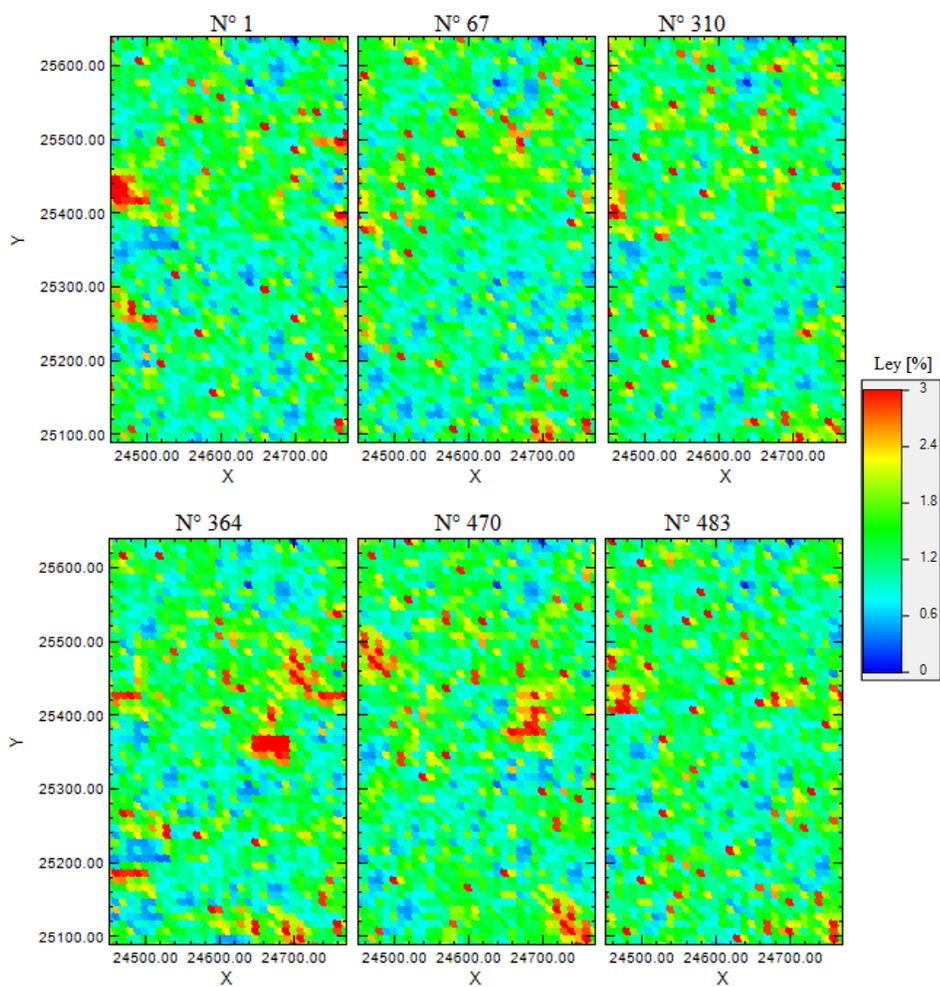


Figura 70 Seis realizaciones sobre el banco 3850

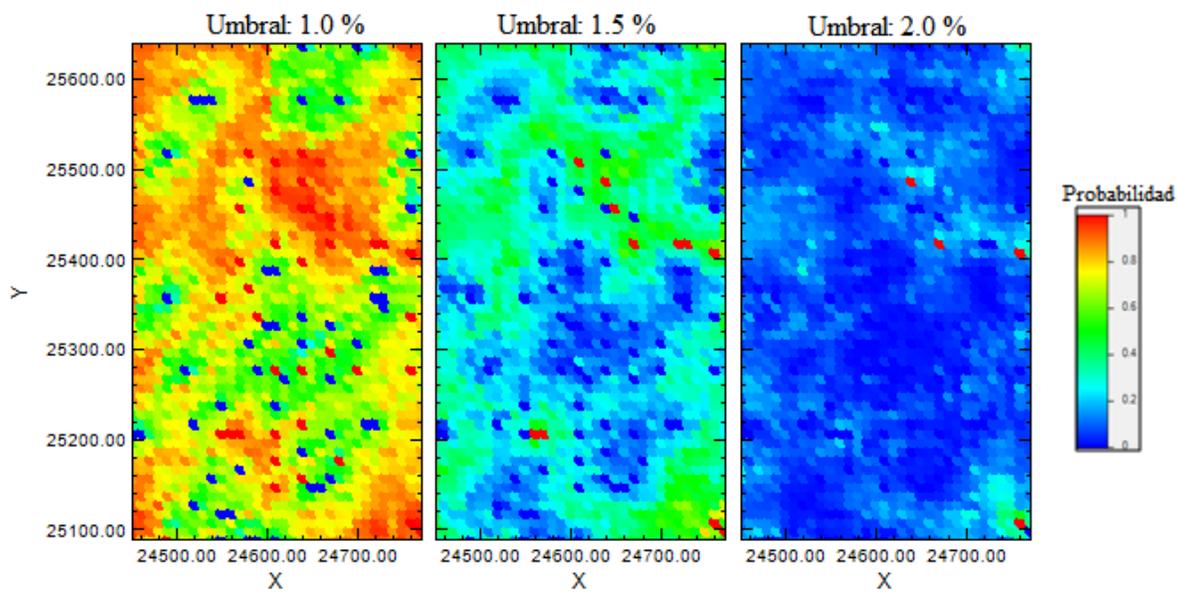


Figura 71 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3850

5.5.4 Resultados utilizando la T.I. 3 sobre el banco 3830

Del mismo modo que se analizaron los resultados para los dos bancos anteriores, se procede ahora con el banco 3830. Después de obtenida la ley media en el banco para cada realización se genera el histograma de la figura 72 de la cual se infiere que el banco tiene una ley media entre [1.195 – 1.135] % con mayor probabilidad a una ley media de 1.27 %.

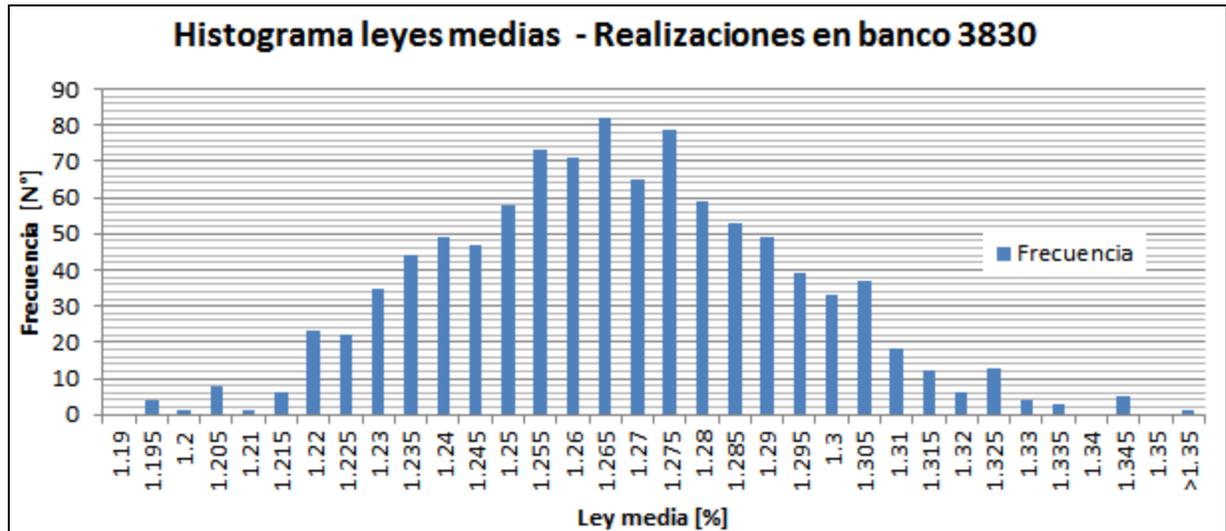


Figura 72 Histograma leyes media – Banco 3830

Respecto a las realizaciones, se observa en la figura 73 que cada una por separado entrega cúmulos de altas leyes dispersos de manera aleatoria. Además se observa una leve anisotropía en dirección N°45 O la cual se marca aún más en los mapas de probabilidad al sobrepasar el umbral 1.5 %.

Analizando los mapas de probabilidad presentes en la figura 74 se aprecian sectores preferentes a leyes medias-altas. En efecto, el sector elipsoidal comprendido entre X[24,550 – 24,800] y Y[25,400 – 25,550] garantiza con una probabilidad del 60% obtener leyes sobre 1.5 %. Situación similar ocurre en ciertos sectores pero de manera menos marcada o más pequeña.

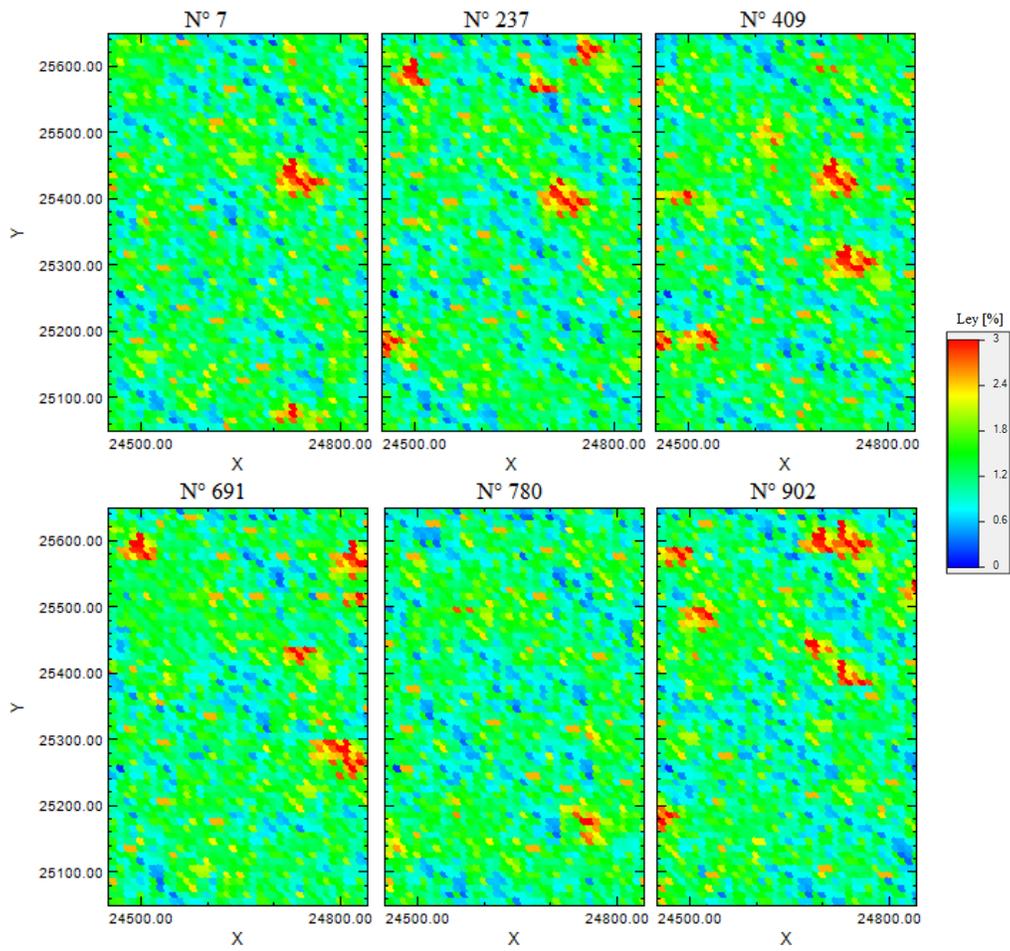


Figura 73 Seis realizaciones sobre el banco 3830

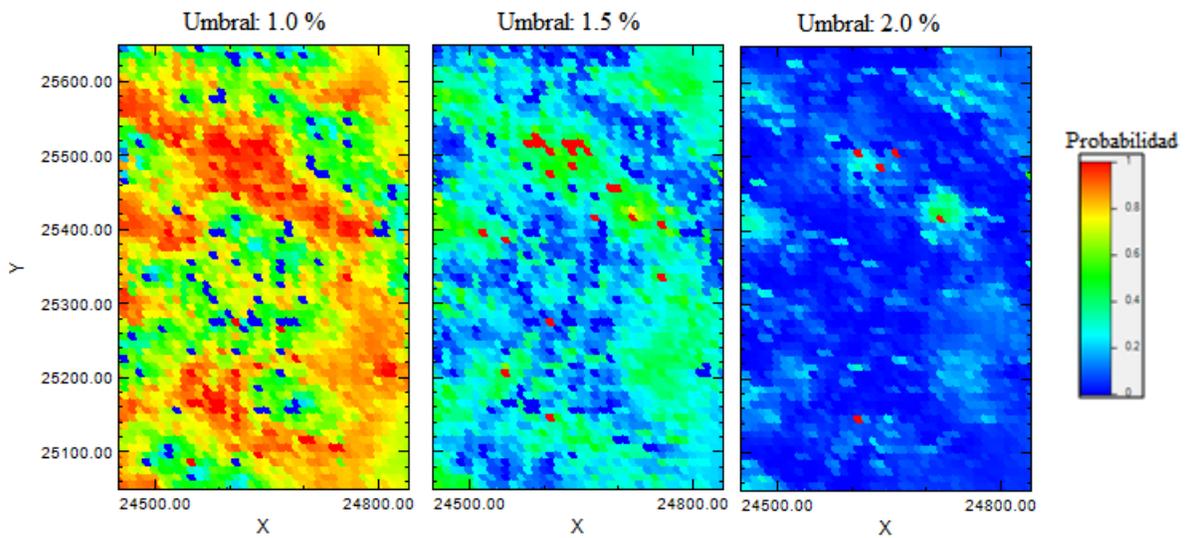


Figura 74 Mapas de probabilidad – Umbrales 1.0%, 1.5% y 2.0% - Banco 3830

5.6 Comparación con estimaciones mediante kriging ordinario

El presente capítulo presenta una comparación entre los resultados obtenidos a través de las simulaciones utilizando FILTERSIM y la estimación mediante kriging ordinario. Se presenta brevemente los variogramas experimentales y el variograma modelado, así como la vecindad de búsqueda utilizados en la estimación. Se estiman los 3 bancos basados en los datos de sondajes asumiendo que no existe anisotropía, es decir, se utiliza un variograma omni-direccional. Se realiza la estimación en el software SGeMS y se visualizan los resultados en el software U-Mine. Se comparan “Banco estimado v/s Promedio realizaciones simuladas” y “Varianza de kriging v/s Varianza condicionada”.

5.6.1 Resultados utilizando la T.I. 1 sobre el banco 3860

El variograma experimental sobre los datos condicionantes (sondajes) en el banco 3860 presenta el comportamiento presentado en la figura 75. Se modela con solo una estructura esférica tal como se presenta en la ecuación 12. Se estima utilizando una vecindad esférica de radio 150 con un mínimo de 2 datos y un máximo de 12.

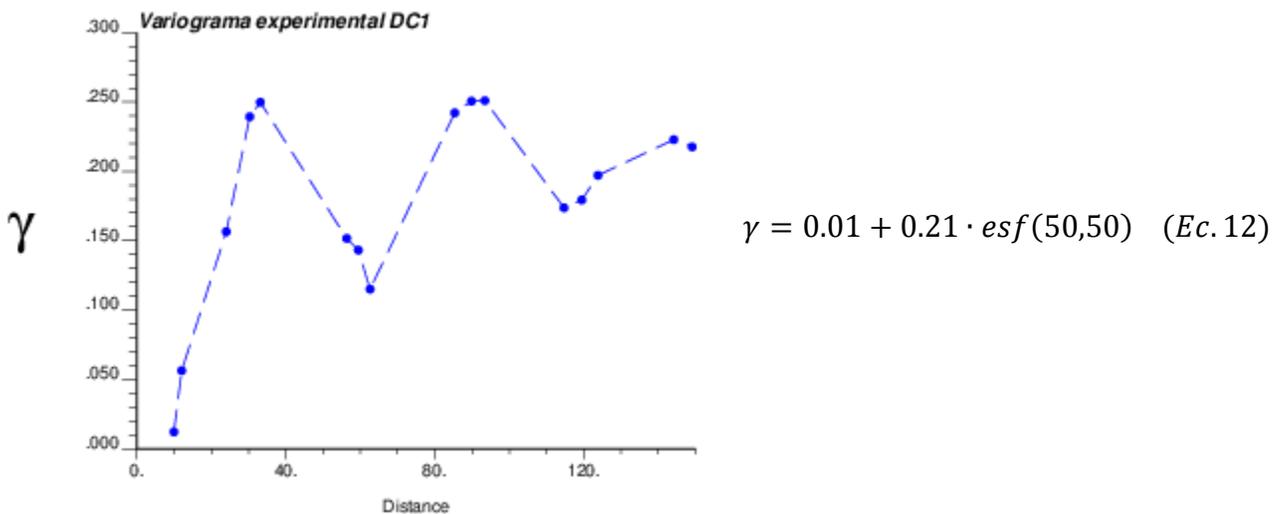


Figura 75 Variograma experimental omni-direccional sobre datos condicionantes 1

La figura 76 expone el banco estimado mediante kriging ordinario (K.O.) y el promedio de las 1,000 realizaciones mediante FILTERSIM. Se observa claramente que ambos resultados no son altamente similares, como abría de esperarse (ya que el promedio de las realizaciones debe tender

a la estimación mediante kriging). Para ver con más detalle la distribución de las leyes en el banco 3860 la figura 77 presenta la comparación entre ambos histogramas.

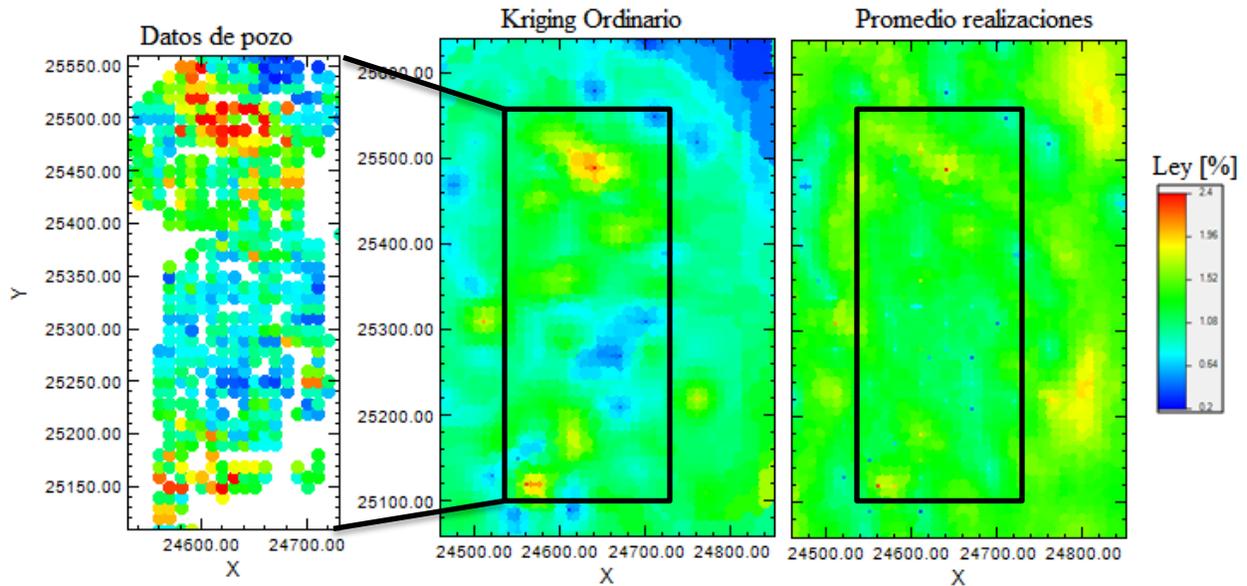


Figura 76 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3860

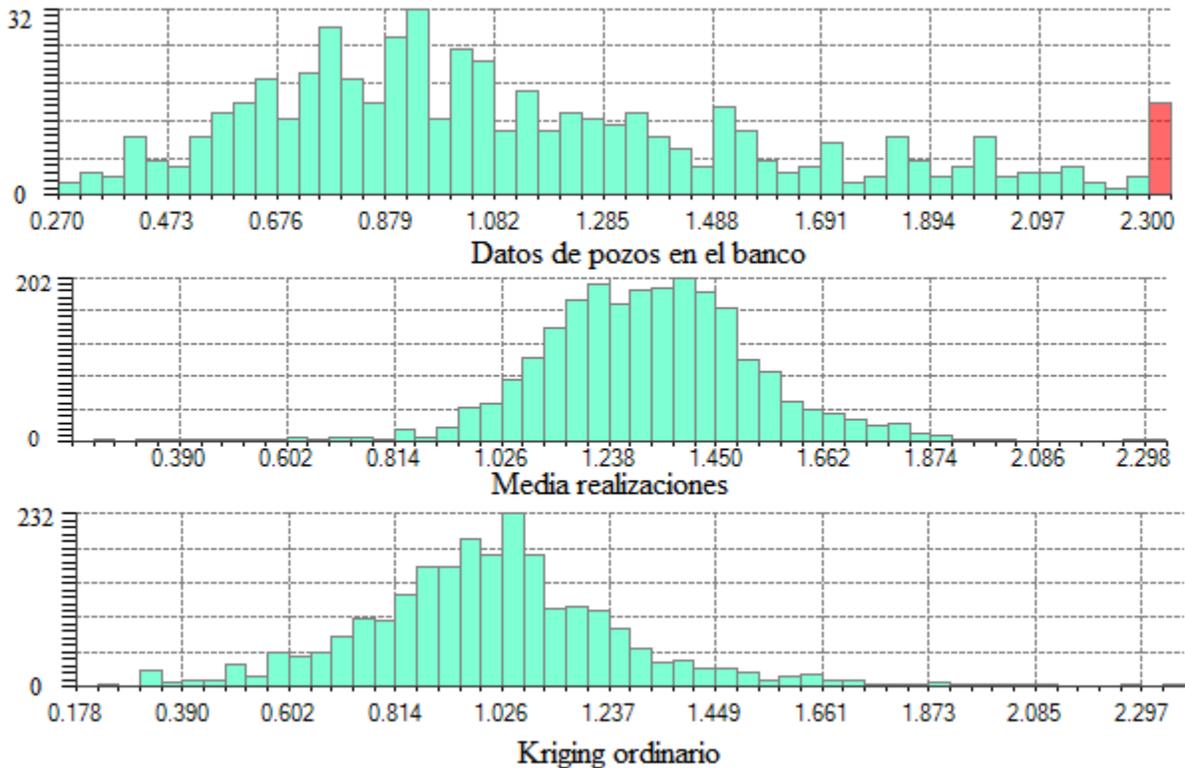


Figura 77 Histograma de valores en el banco 3860 mediante kriging, simulacion y datos de pozo

Otro punto a analizar es la comparación entre la varianza de kriging y la varianza condicional obtenida de la simulación. La figura 78 presenta los mapas de varianza entre ambos métodos. Cabe destacar que en la varianza de kriging los valores están en el rango $[0 - 0.33]$ mientras que en la varianza condicionada están en el rango $[0 - 1.985]$. No obstante, para una mejor visualización la escala de colores ha sido ajustada al intervalo $[0 - 1]$.

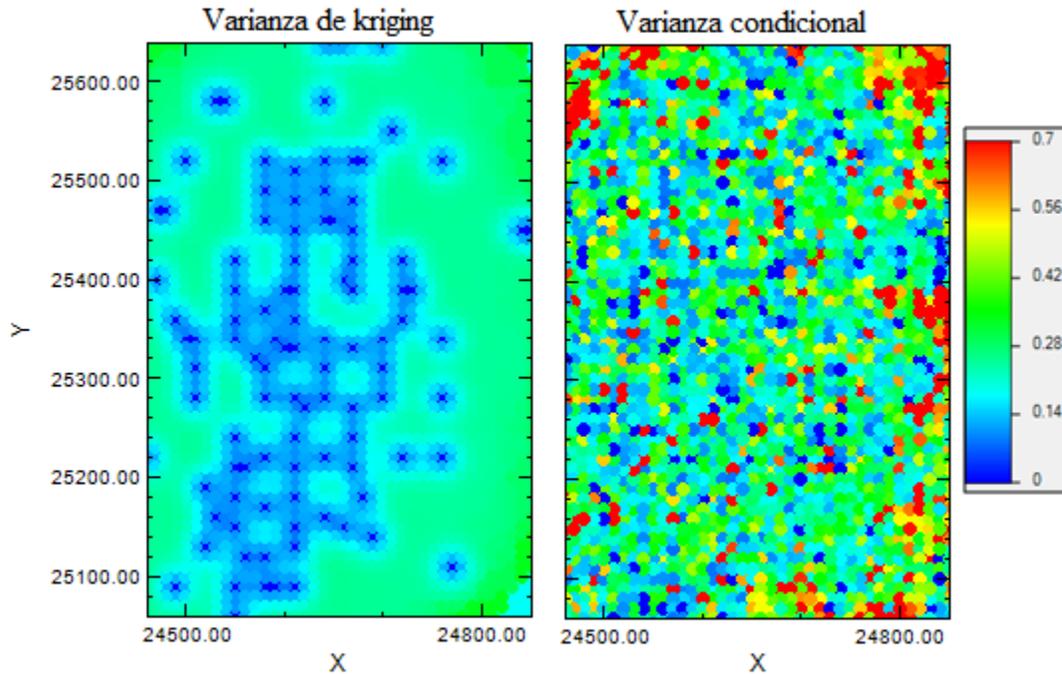


Figura 78 Varianza condicional y varianza de kriging en banco 3860

5.6.2 Resultados utilizando la T.I. 2 sobre el banco 3850

El mismo proceso anterior se lleva a cabo en el banco 3850. La figura 79 presenta el variograma experimental obtenido al analizar los datos condicionantes (sondajes) en el banco 3850. Es posible modelar este variograma utilizando 2 estructuras: una exponencial de alcance 40 y una esférica de alcance 120 tal como se observa en la ecuación 13.

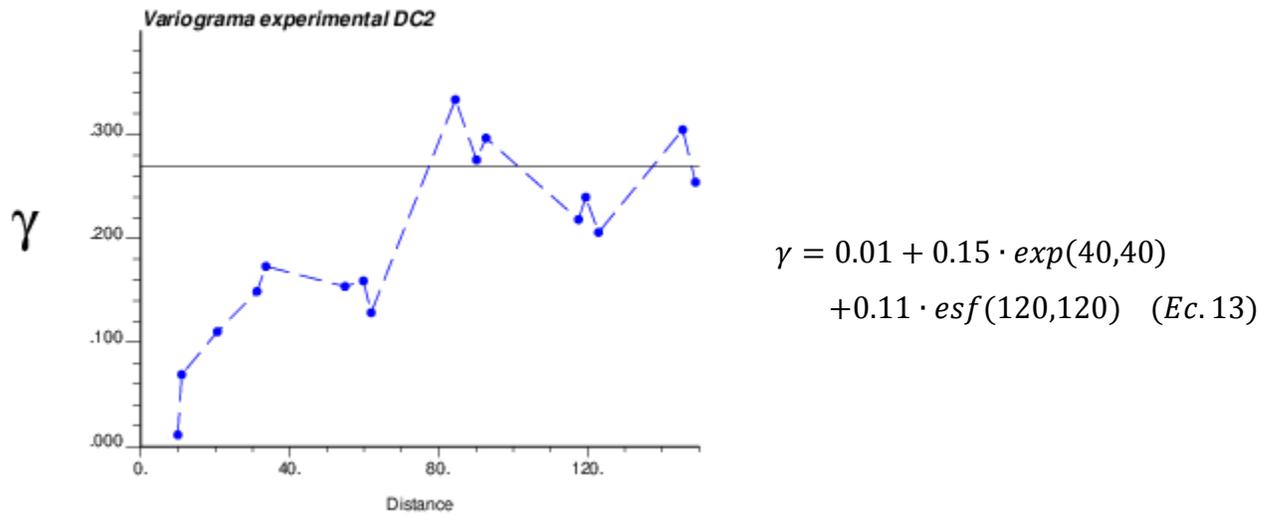


Figura 79 Variograma experimental omni-direccional sobre datos condicionantes 2

La figura 80 presenta la comparación entre los datos provenientes de los pozos de tronadura (migrados a una grilla regular), el promedio de las 1,000 realizaciones mediante FILTERSIM y la estimación mediante kriging.

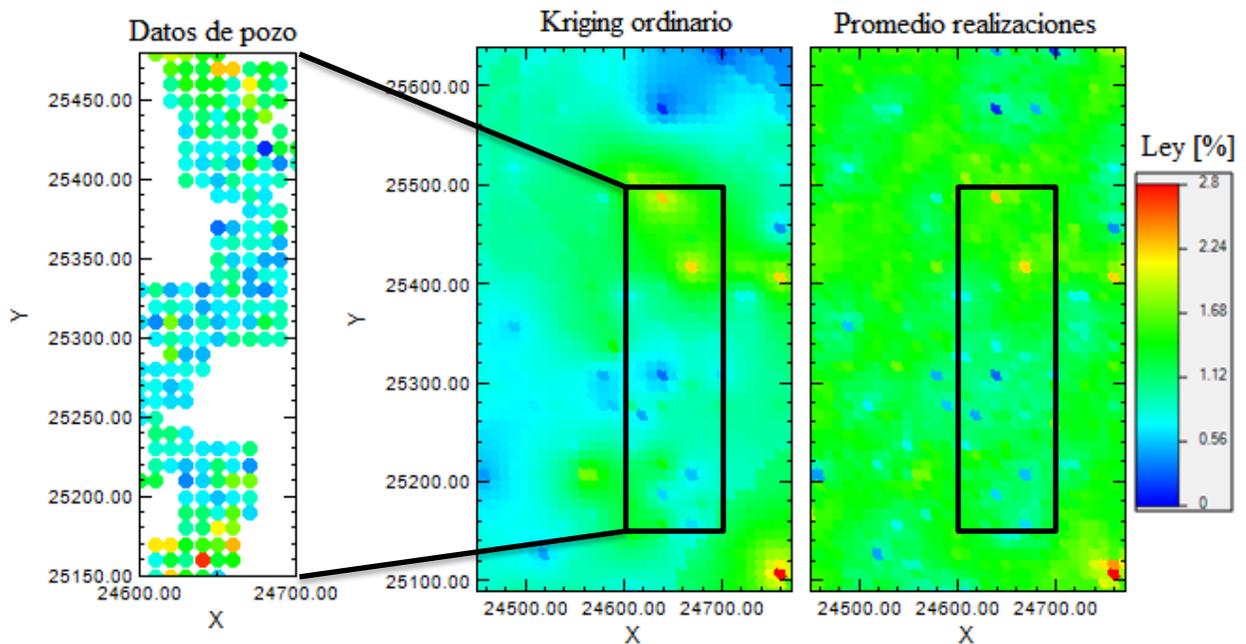


Figura 80 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3850

Se observa una mayor similitud entre la estimación mediante kriging y el promedio de las realizaciones mediante FILTERSIM, en comparación a lo que sucede en el banco 3860. Para tener una idea más clara de los datos obtenidos la figura 81 presenta los histogramas de los valores resultantes y del banco dado los pozos de tronadura.

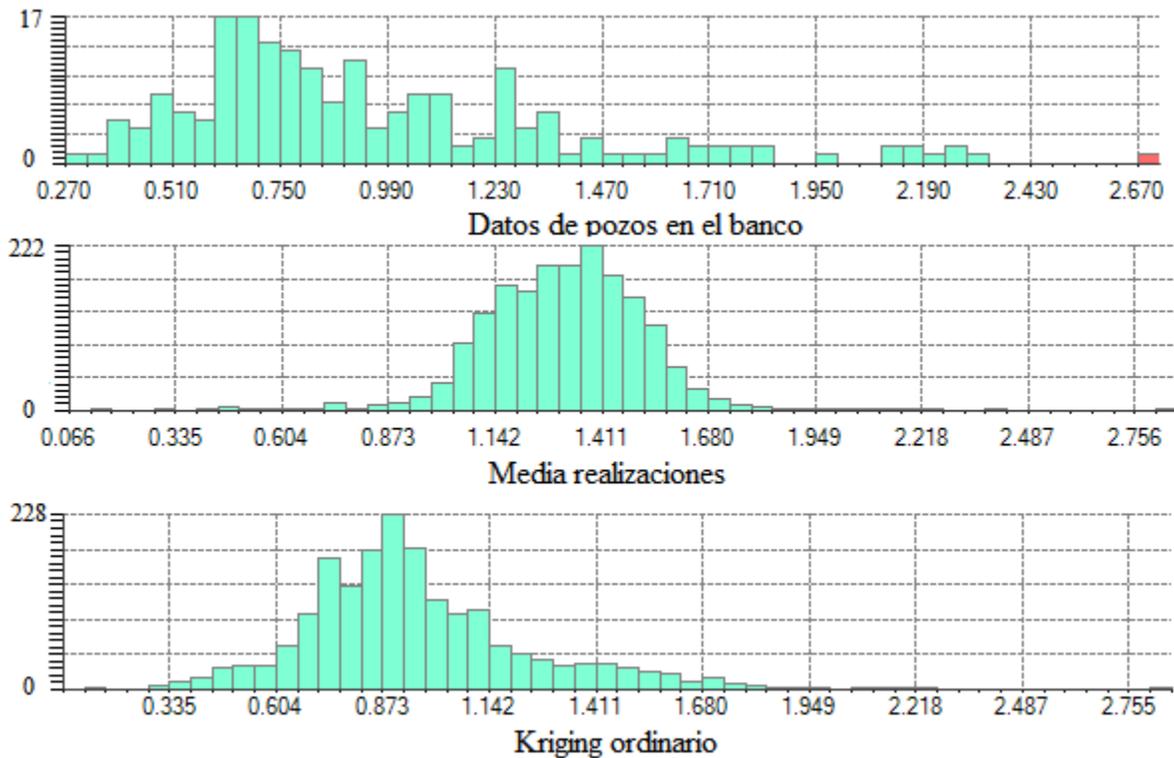


Figura 81 Histograma de valores en el banco 3850 mediante kriging, simulacion y datos de pozo

Nuevamente se compara la varianza condicional obtenida de las simulaciones respecto a la varianza de kriging. La primera presenta valores en el rango $[0 - 2.08]$ mientras que la segunda en el rango $[0 - 0.387]$. Para obtener una comparación visual más nítida se presentan ambos mapas de varianza en el rango $[0 - 0.7]$ tal como se ilustra en la figura 82.

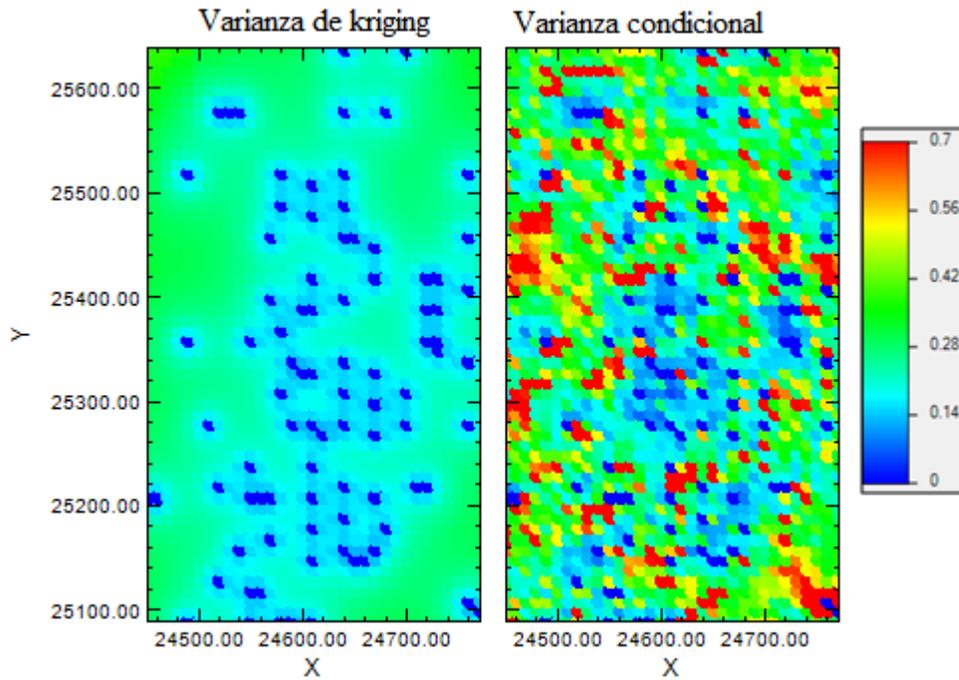


Figura 82 Varianza condicional y varianza de kriging en banco 3850

5.6.3 Resultados utilizando la T.I. 3 sobre el banco 3830

El mismo proceso llevado a cabo en los bancos 3860 y 3850 se realiza ahora en el banco 3830. El variograma experimental omni-direccional se ilustra en la figura 83 y su respectiva modelación en la ecuación 14.

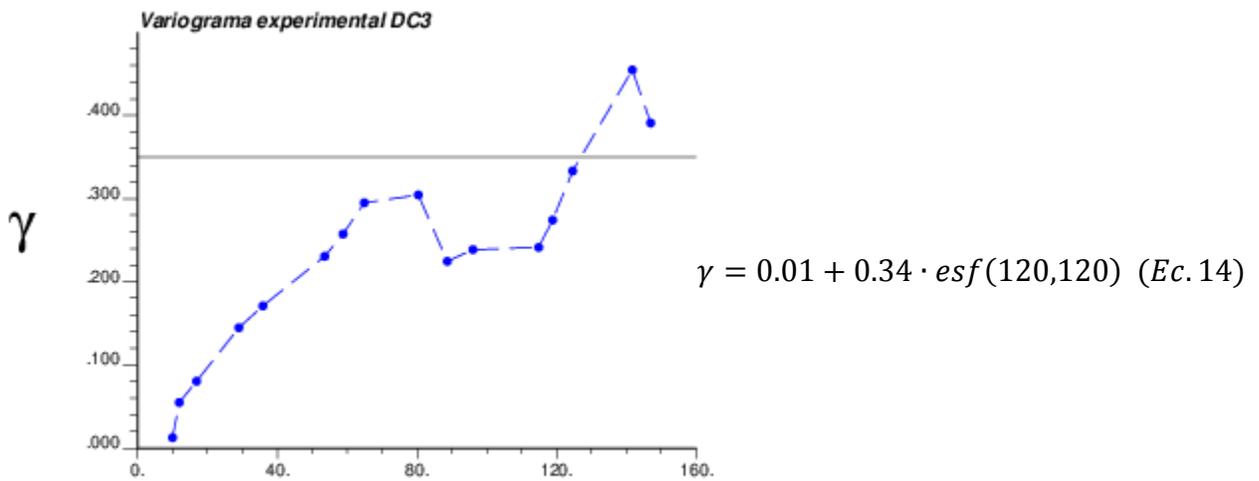


Figura 83 Variograma experimental omni-direccional sobre datos condicionantes 3

La figura 84 ilustra la estimación mediante kriging ordinario del banco 3830, el promedio de las 1,000 realizaciones mediante FILTERSIM y el sector del banco muestreado por pozos de tronadura.

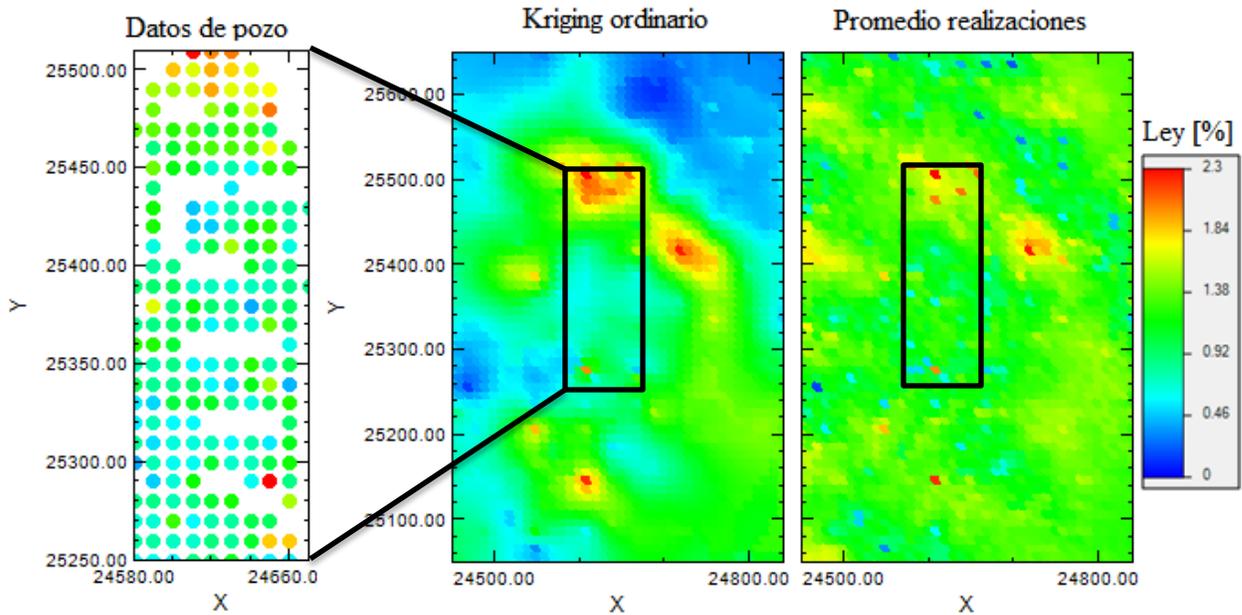


Figura 84 Comparación estimación mediante K.O. v/s promedio realizaciones - Banco 3830

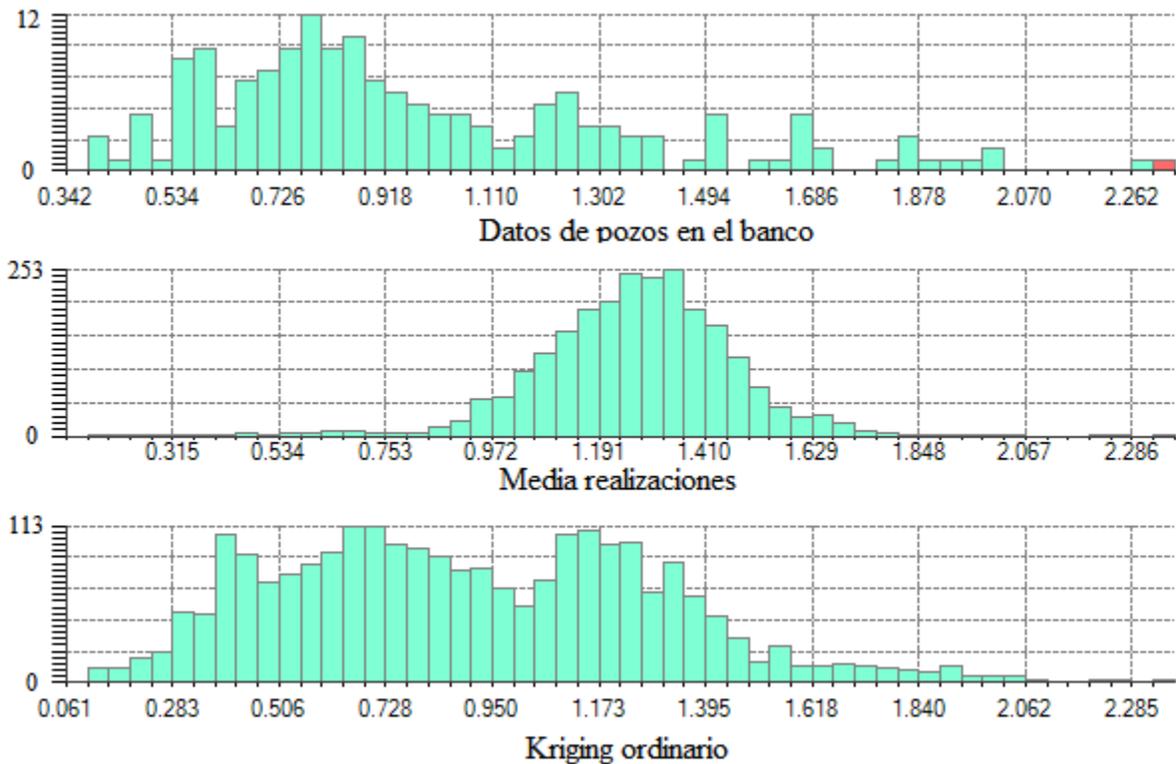


Figura 85 Histograma de valores en el banco 3830 mediante kriging, simulacion y datos de pozo

Al analizar el banco estimado con el simulado se observa claramente que la distribución de los valores estimados es considerablemente más apilada que el promedio de las realizaciones. Con más consistencia se observa entre los histogramas de la figura 85, en donde claramente se observa la gran dispersión de los datos en la estimación mediante kriging ordinario.

Finalmente, se analiza la varianza condicional obtenida en la simulación y la varianza de kriging. La varianza de kriging está en el rango $[0 - 0.608]$ mientras que la varianza condicional está en el rango $[0 - 0.831]$. Para tener una mejor visualización se opta por utilizar como leyenda el rango

$[0 - 0.7]$ tal como se ilustra en la figura 86.

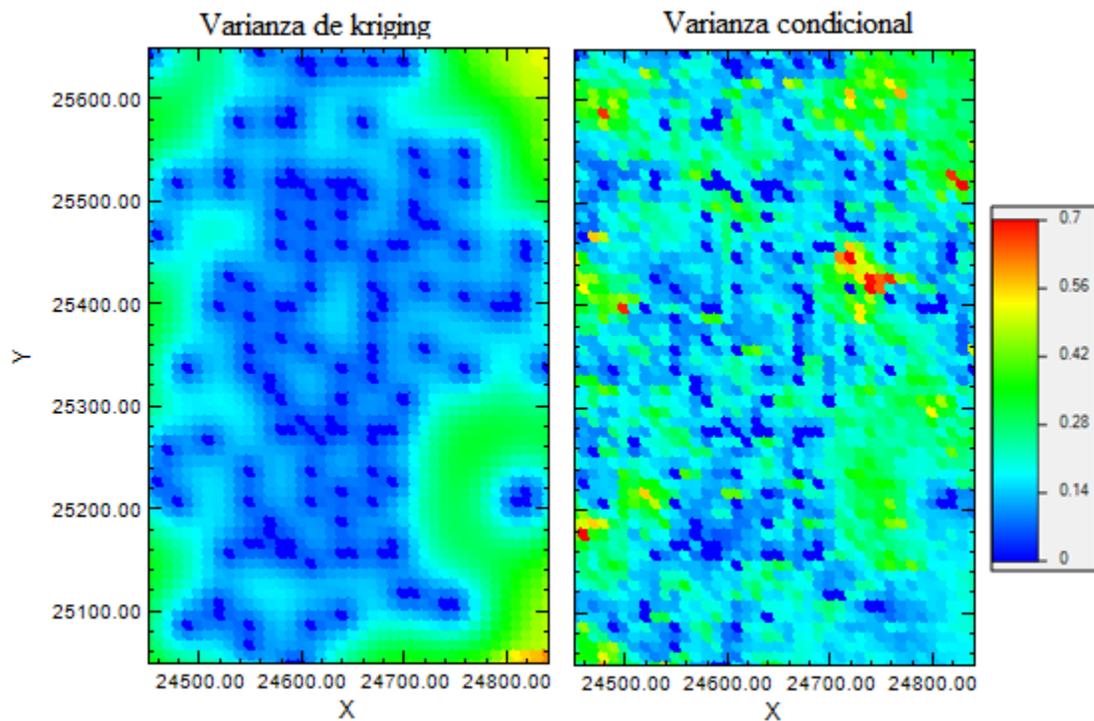


Figura 86 Varianza condicional y varianza de kriging en banco 3830

5.7 Curvas tonelaje-ley

El presente capítulo presenta las curvas tonelaje ley en los bancos simulados. Cabe destacar que las grillas simuladas en cada banco son de distinto tamaño y que las áreas abarcadas en las muestras de pozos de tronadura son de distinto tamaño también. Por lo anterior, se decide realizar una comparación normalizada en las curvas tonelaje-ley entre las realizaciones (3 por cada simulación), el promedio de las 1,000 realizaciones por simulación y su respectiva estimación mediante kriging ordinario expuesta en el capítulo anterior. Se entiende por normalización a la generación de un intervalo [0 – 1.0] considerando como 1.0 el tonelaje total en el banco.

5.7.1 Banco 3860

Al contar con las realizaciones y la estimación de kriging es posible comparar cómo se comportan las curvas tonelaje-ley. Para ello, se debe establecer cuanto es el tonelaje simulado o estimado sobre el banco con la finalidad de normalizar el tonelaje y poder realizar la comparación (ver figura 87). Los tonelajes totales, asumiendo bloques de 10m x 10m x 10m y una densidad de roca de 2.7 t/m³, son:

- Estimación por kriging ordinario y simulación utilizando FILTERSIM: 6.372 Mton.
- Datos medidos de pozos de tronadura: 1.49 Mton.

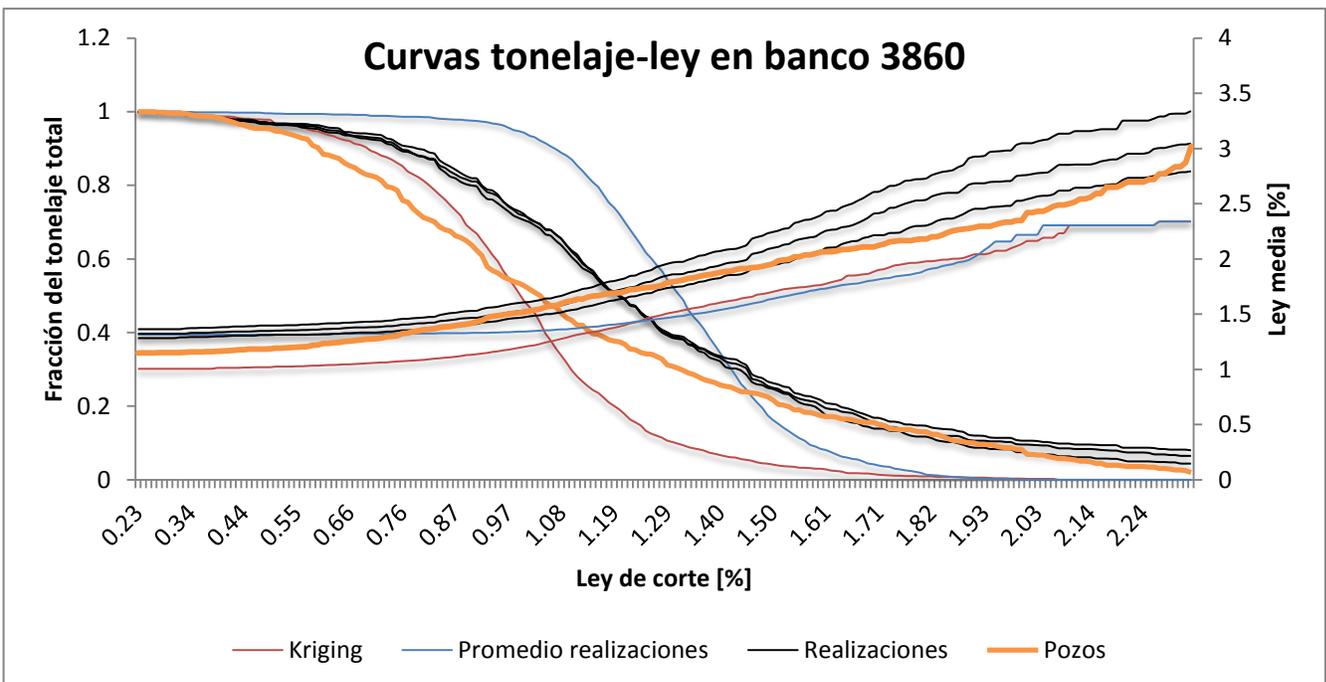


Figura 87 Curvas tonelaje-ley en banco 3860

Hay una gran diferencia entre las curvas representantes al promedio de las realizaciones y a la estimación mediante kriging ordinario. De hecho, la primera entrega un mayor tonelaje sobre leyes de corte equivalente debido posiblemente al no suavizamiento de las leyes bajas y altas. Por otro lado, las curvas de leyes promedio sobre una determinada ley de corte presentan un comportamiento similar entre el promedio de realizaciones y la estimación de kriging. Sin embargo, al analizar las realizaciones por separado, se alcanzan leyes promedio hasta por 0.7% por sobre las estimadas o el promedio de las realizaciones.

5.7.2 Banco 3850

Sobre el banco 3850 se realiza el mismo análisis anterior. Los tonelajes totales vienen dados por:

- Estimación mediante kriging ordinario y simulación utilizando FILTERSIM: 4.989 Mton.
- Datos medidos de pozos de tronadura: 0.54 Mton.

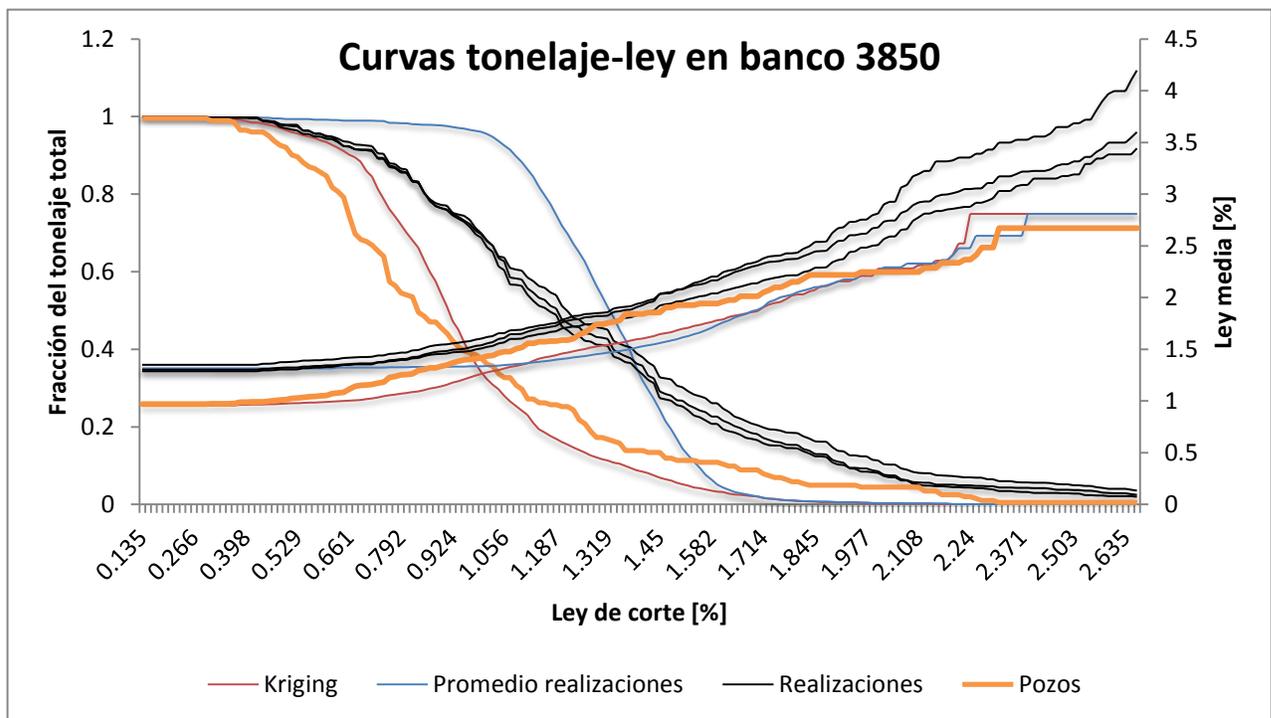


Figura 88 Curvas tonelaje -ley en banco 3850

La figura 88 presenta las curvas tonelaje ley para el banco 3850, las cuales se comportan muy similar a las obtenidas en la figura 87 para el banco 3860. Por lo tanto el análisis es similar, agregando en esta oportunidad que tanto cada realización, el promedio de estas o bien la

estimación sobre estiman el tonelaje real bajo una ley de corte igual o menor a 1.0 % (promedio del kriging). Por otro lado, las curvas de ley promedio presentan comportamientos similares.

5.7.3 Banco 3830

Por último, en el banco 3830 se vuelve a realizar el mismo procedimiento anterior considerando los siguientes tonelajes totales:

- Estimación mediante kriging ordinario y simulación utilizando FILTERSIM: 6.588 Mton.
- Datos medidos de pozos de tronadura: 0.521 Mton.

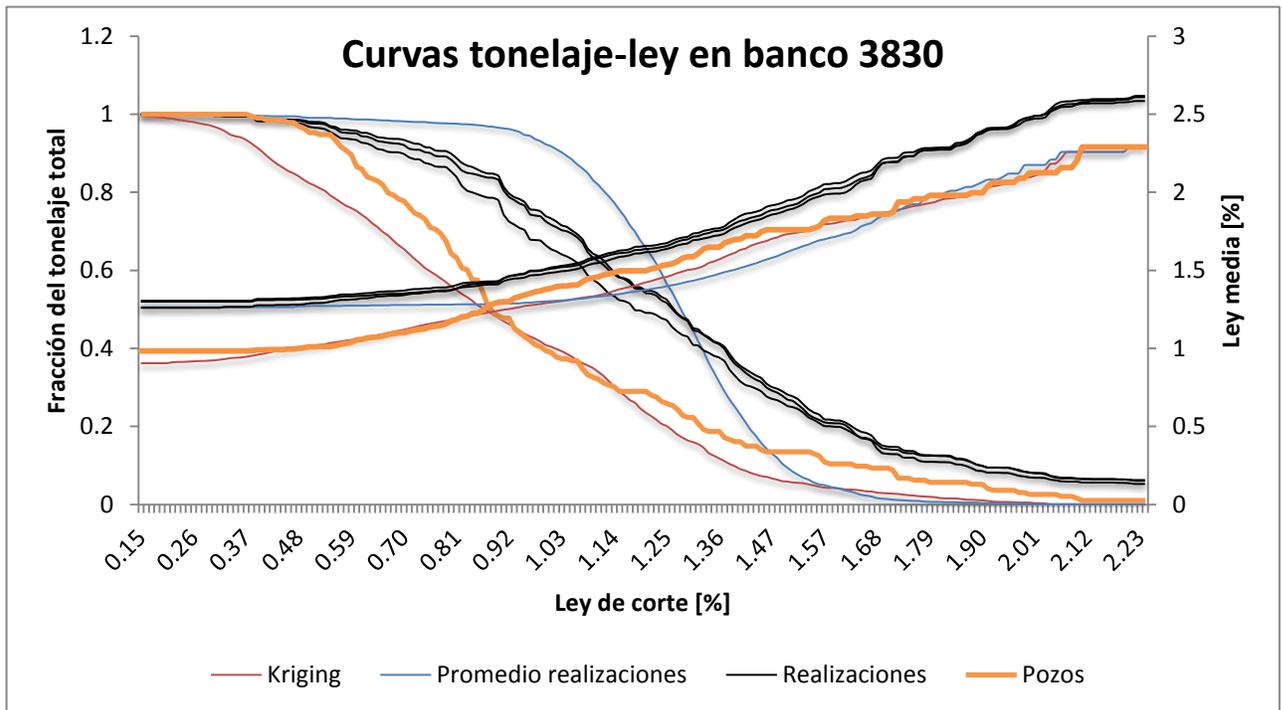


Figura 89 Curvas tonelaje-ley en banco 3830

Conclusiones similares a los casos anteriores se presentan en el banco 3830 representado en la figura 89. Lo peculiar en este banco es que las leyes obtenidas a través de pozos de tronadura están por sobre lo estimado mediante kriging. Por su parte, las realizaciones simuladas están por sobre las curvas tanto estimadas como de pozos tanto para la fracción del tonelaje total como de ley media.

5.8 Análisis y discusión

Al comparar la ley media en el banco simulado, tomando como ejemplo el banco 3850, se observa que la estimación mediante kriging entrega una media igual a 0.958 %, el 50% de las realizaciones al simular con FILTERSIM presentan una media en el intervalo [1.21 – 1.39] %, el promedio de las realizaciones de 1.31 %, los datos condicionantes una media de 0.998 % y el banco 3850 medido en los pozos de tronadura de 0.965 %. Analizando esto se observa que el promedio de las simulaciones no tiende a la estimación de kriging, algo que en teoría debería suceder. Por lo tanto, el problema puede deberse a una interpretación no adecuada de la correlación espacial de la variable (variograma experimental) a no considerar anisotropías, inclusive una no adecuada modelación. No obstante, se interpreta que las realizaciones tienden a una ley media en el banco al menos unos 0.2 % - 0.3 % más elevada que la entregada por estimación mediante kriging ordinario, hasta el mismo valor de las muestras de pozos en el banco. Lo mismo sucede para los dos bancos restantes 3860 y 3830.

Se interpreta de la varianza condicional que esta, a diferencia de la varianza de kriging, incorpora el efecto proporcional, es decir, que localmente la desviación estándar está en proporción cuadrática a la media. Visto de otra forma, se desprende que la variable regionalizada en la simulación presenta un comportamiento log-normal en cada una de las realizaciones.

El mapa de probabilidad bajo un umbral de 1.5 %, en los tres bancos simulados, coincide considerablemente con la estimación mediante kriging. Más aún, si esta última fuese normalizada por el máximo valor para obtener una estimación en el intervalo [0 – 1.0] ambos mapas tendrían alta similitud. Por lo tanto, una muy potente herramienta son los mapas probabilísticos para decidir qué sectores explotar en la vida real al tener leyes de corte variables (posiblemente en las etapas temprana de ingeniería para la planificación de la mina).

6 ALTERNATIVAS DE MODIFICACION AL CÓDIGO FILTERSIM

Habiendo definido previamente FILTERSIM como el algoritmo de simulación multi-punto a ser modificado para la incorporación de la media local en las simulaciones ahora se presentan 2 alternativas (una más acertada que otra) que serán detalladas presentando además un análisis de su real representatividad a la hora de simular.

Importante es la incorporación de la media local ya que es posible influir en los resultados de las simulaciones incorporando como dato secundario tendencias de la media. Concretamente es posible influenciar en los recursos totales (leyes medias en los bancos de un yacimiento, distribución espacial de leyes altas/bajas, etc.) lo cual es relevante para los procesos aguas abajo a partir de la evaluación de recursos.

La primera alternativa apunta a modificar la función que determina el patrón más cercana al evento condicionante, mientras que la segunda busca modifica el los patrones almacenados al incorporar un nuevo filtro previo al proceso de agrupamiento.

6.1 Modificación de la función distancia

Se propone mantener las metodologías de agrupamiento, considerando la aplicación de los 3 filtros (para el cálculo de los 3 primeros momentos) en las principales direcciones del espacio (2 en 2D y 3 en 3D), pero el cambio sustancial se realizaría en la determinación del prototipo más cercano a la condición de evento. Se propone agregar un ponderador, W_s , que presenta valores en el intervalo $[0,1]$ en donde: si vale 0 no existe incorporación de la media local, y si vale 1 sólo se considera la diferencia absoluta entre la media del evento condicionante y del prototipo en comparación. Cabe destacar que no todos los datos del prototipo son considerados en la media, sino sólo los que están en la misma ubicación espacial, respecto al nodo central, que los nodos informados en el evento condicionante.

Si se incorpora esta técnica a la metodología *template pixel* se tendría la ecuación 15

$$D_l(u_o) = \left(\sum_{d=1}^3 w(d) \text{Average} |dat_d(u_o + u) - prot^{(l)}(u)| \right) (1 - w_s) + w_s (|dat_m - prot_m^{(l)}|) \quad \text{con } w_s \in [0,1] \quad (\text{Ec. 15})$$

O bien, para la metodología *filter score* se tendría la ecuación 16.

$$D_l(u_o) = \left(\sum_{i=1}^F |S_{dat}^{(i)}(u_o) - S_{prot^{(i)}}^{(i)}(u)| \right) (1 - w_s) + w_s (|dat_m - prot_m^{(l)}|)$$

con $w_s \in [0,1]$ (Ec. 16)

Interesante sería aplicar ambas alternativas, la primera al caso de estudio real ya que utilizó *template pixel* y la segunda al caso de estudio sintético que utilizó *filter score*. De esta forma es posible contrastar los resultados y establecer el real impacto en los resultados.

6.2 Consideración de la media local en la determinación del patrón final, nuevo filtro.

Esta alternativa propone incorporar la media local en los patrones iniciales de FILTERSIM, es decir, agregar un nuevo filtro en el proceso de *clustering* de patrones inferidos de la imagen de entrenamiento. Este nuevo filtro es directamente la media de los datos dentro de la grilla de búsqueda.

Recordando que la grilla de búsqueda tiene un tamaño de $(2n + 1) \times (2n + 1)$ el nuevo filtro (media local) para los datos dentro de la grilla viene dado por

$$\text{Nuevo filtro: media local, } f_7: \frac{1}{(2n + 1) \times (2n + 1)}, \quad \forall u, v \in [-n, n]$$

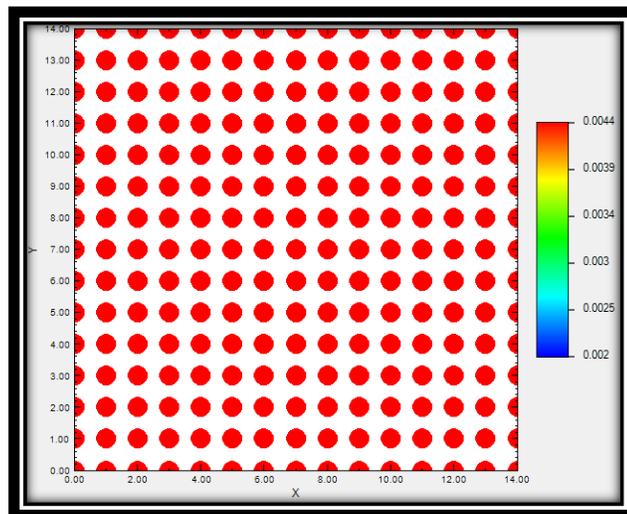


Figura 90 Séptimo filtro

El filtro debe ser aplicado tanto a los patrones inferidos de la imagen de entrenamiento como a los eventos condicionantes en la grilla de simulación. De esta manera, no es necesaria la modificación de las funciones para el cálculo de la distancia entre el evento condicionante y el patrón más cercano. Si se desea aumentar la ponderación de la media local en los resultados, se propone incluir más de una vez el mismo filtro recién expuesto.

Al realizar esto los parámetros a definir por parte del usuario no se verán modificados si no que la incorporación de la media local vendrá de datos secundarios y afectaran en el proceso inicial de agrupamiento. Interesante resultaría analizar la incorporación de este filtro una, dos o más veces y ver como influencia en los resultados. Para tener un contraste se recomienda tomar los datos del yacimiento del caso real y reformular el análisis y en la comparación entre kriging ordinario y FILTERSIM actual agregar este nuevo FILTERSIM modificado.

7 CONCLUSIONES

Al analizar el estado del arte en el campo de la simulación estocástica multipuntos se optó por estudiar, aplicar y generar propuestas de modificación al algoritmo *Filter-Based Simulation* (FILTERSIM) dado que es veloz al simular, su estructura posibilita la incorporación de la media local en la simulación, es posible simular sobre variables continuas sin necesidad de generar intervalos y está incorporado al software libre *Stanford Geostatistical Modeling Software* (SGeMS) cuyo código fuente es de carácter libre y abierto.

Se proponen dos alternativas para la incorporación de la media local: la primera consiste en la modificación de la función distancia al agregar un ponderador a la diferencia que existe entre la media local del patrón y la media local esperada, sin modificación en el proceso de agrupamiento de patrones. La segunda consiste en la incorporación de un séptimo filtro capaz de inferir la media local en cada patrón y almacenarlo como una nueva dimensión. De esta manera se ve afectado el proceso de agrupamiento de patrones. Para que coincidan las dimensiones de los patrones con los eventos condicionantes se debe incorporar la media local esperada a estos últimos. En esta segunda alternativa no se ven modificadas las funciones para el cálculo de la distancia.

Respecto al caso de estudio sintético y en el análisis de cómo repercuten las variaciones de los parámetros de entrada en la calidad de las simulaciones y su proceso global se tiene:

- Un aumento en el tamaño de la grilla de búsqueda, fijando el resto de los parámetros lleva a un aumento en el tiempo de cómputo y una disminución en la continuidad espacial de la estructura.
- Un aumento en el tamaño de la grilla interna, fijando el resto de los parámetros lleva a una disminución en el tiempo de cómputo y un leve aumento en la continuidad espacial de la estructura en desmedro del comportamiento a escala local.
- Aumentar 50 veces la cantidad de datos condicionantes lleva a resultados más similares a la imagen de entrenamiento (caso sintético) y una mayor continuidad espacial en la estructura, no obstante, las estadísticas básicas y la correlación espacial entre los valores simulados son similares.

- Utilizar *template pixels* lleva a simulaciones cuyas estructuras son más continuas, con bordes más delineados y cambios con el entorno más abruptos, mientras que al utilizar *filter scores* se llega a simulaciones cuyas estructuras son más dispersas, una menor continuidad espacial y cambios con el entorno más diseminados.
- Utilizar *template pixels* lleva a tiempos de cómputo entre 3 a 4 veces más elevados que utilizar *filter scores* bajo las mismas configuraciones.

Por todo lo anterior, se concluye que la configuración óptima es relativa y depende sustancialmente de lo que el usuario desee llevar a cabo. En otras palabras, la determinación del valor para cada uno de los parámetros de entrada y las metodologías a implementar para los distintos sub-procesos de la simulación deben balancear las aristas: velocidad en el tiempo de cómputo, calidad en la continuidad espacial de las estructuras, captura de los comportamientos a mayor o menor escala, disminución en la varianza de las realizaciones, entre otras.

Respecto al caso de estudio real, aplicado a un yacimiento cuprífero, se utilizan los parámetros óptimos y recomendados como resultado del caso de estudio sintético. Se observa que tanto la ley media de los bancos estimados en cada realización y en el promedio de estas supera en 0.2 % - 0.3 % a la ley media determinada por los pozos de tronadura, los sondajes como datos condicionantes y estimación mediante kriging ordinario. No obstante, al analizar los mapas de probabilidad se tiene que al utilizar un umbral de 1.5% el comportamiento es muy semejante al resultado del mapa al estimar mediante kriging ordinario. El valor umbral 1.5% puede ser reemplazado por la media de las leyes en las realizaciones y posiblemente se obtengan comportamientos más similares aun entre ambos mapas. Por otro lado, al analizar la varianza condicional se observa claramente la incorporación del efecto proporcional, algo no considerado por parte de la varianza de kriging.

8 RECOMENDACIONES A FUTURO

Considerando los alcances de la presente memoria se desprenden las siguientes recomendaciones a futuro:

Incorporar la media local al algoritmo FILTERSIM. Para ello se recomienda utilizar ambas alternativas de modificación. Al tener generada ambas técnicas realizar el caso de estudio real y contrastar los resultados obtenidos en esta memoria con los nuevos. Se recomienda su implementación en el software SGeMS a través del uso de un plug-in que tome como base estructural el código fuente de FILTERSIM para variables continuas. Cabe mencionar que dicho algoritmo está escrito en el lenguaje C++ lo que presenta un desafío más por el lado computacional que por el ámbito minero.

Respecto al caso de estudio sintético expuesto en la presente memoria se propone desarrollar nuevas simulaciones, bajo las mismas configuraciones, pero sin considerar datos condicionantes, y aplicar el algoritmo modificado (incluyendo la media local como parámetro de entrada) con la finalidad de comparar la calidad de los resultados bajo las mismas configuraciones.

Respeto al caso de estudio real (yacimiento cuprífero) tomar mayor consideración en el análisis variográfico, determinando direcciones de anisotropía y generando variogramas modelados más ajustados.

9 BIBLIOGRAFÍA

Austin, 2007, “*Stanford Geostatistical earth modeling software. SGeMS short course::FILTERSIM*”, Austin, Texas, pp. 9-10.

Caers, J., Boucher, A., 2008, “*Multiple-point Geostatistics: mastering training-image based algorithms*”, MPS Short course, Chile, pp. 23.

Deutsch, C. and Journel, A., 1998, “*GSLIB: Geostatistical Software Library and User’s Guide*”, Segunda edición, Oxford University Press, New York, pp. 20-24.

Dujardin, B., 2007, “*Sensitivity analysis of Filtersim and histogram reproduction*”, Tesis de doctorado, Universidad de Stanford, Stanford, CA, pp. 6-28.

Emery, X., 2012, “*Simulación estocástica y geoestadística no lineal*”. Facultad de ciencias físicas y matemáticas, Universidad de Chile, Santiago, Chile, pp. 1-7.

Emery, X., 2012, “*Lección 11: simulación no paramétrica*”, Cátedra simulación geoestadística, Universidad de Chile, Santiago, Chile. pp. 62.

Guardiano, F., Srivastava, M., 1993, “*Multivariate geostatistics: Beyond bivariate moments*”, in: Soares, A., ed., *Geostatistics-Troia*: Kluwer Academic, Dordrecht, pp. 133-144.

Hartigan, J. A., 1975, “*Clustering algorithms*”, John Wiley & Sons, Inc.

Izenman, A. J., 2008, “*Modern multivariate statistical techniques: regression, classification, and manifold learning*”, Springer, pp. 423-424.

Remy, N., Boucher, A. and Wu, J., 2009, “*Applied Geostatistics with SGeMS: A User’s Guide*”, Cambridge University Press, New York, USA. Pp. 191-214.

Strebelle, S., 2000, “*Sequential Simulation Drawing Structures from Training Images*”, Tesis de doctorado, Universidad de Stanford, Stanford, CA, pp. 34-53.

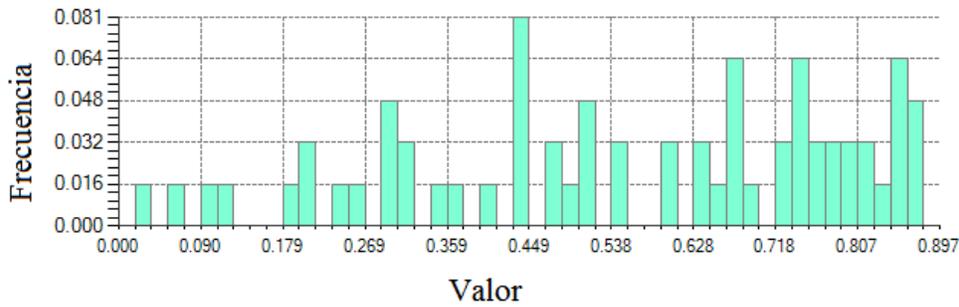
Strebelle, S., 2002, “*Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics*”, *Mathematical Geology*, 34(1), pp. 3-12.

Wu, J., Boucher, A., Zhang, T., 2008, “*A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM*”, *Computers & Geosciences*, 34(12), 1863–1876.

Zhang T., Switzer, P., and Journel, A., 2006, “*Filter-Based Classification of Training Image Patterns for Spatial Simulation*”, *Mathematical Geology*, 38(1), pp. 63-80.

10 ANEXO A

Datos condicionantes 1

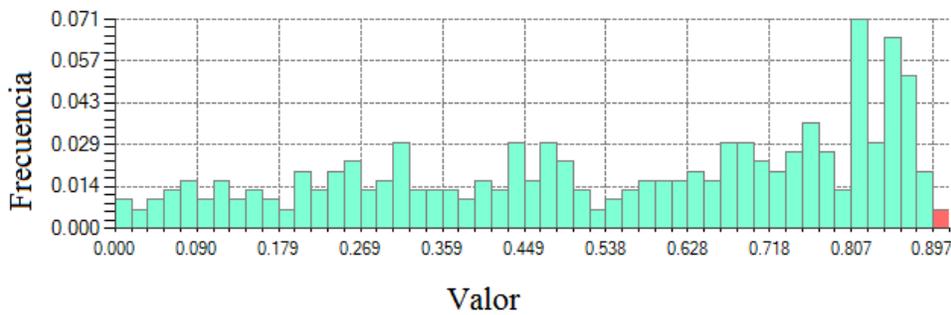


Estadísticas

Promedio	0.556
Varianza	0.056
Mínimo	0.036
Máximo	0.877

Anexos 1 Histograma y estadística datos condicionantes 1

Datos condicionantes 2

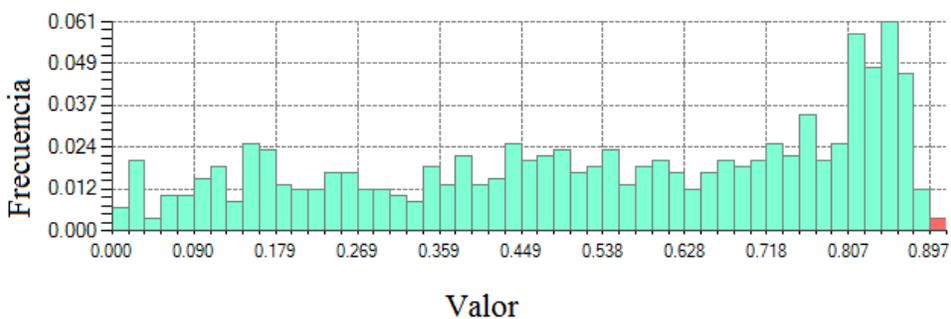


Estadísticas

Promedio	0.550
Varianza	0.068
Mínimo	0.000
Máximo	0.897

Anexos 2 Histograma y estadística datos condicionantes 2

Datos condicionantes 3

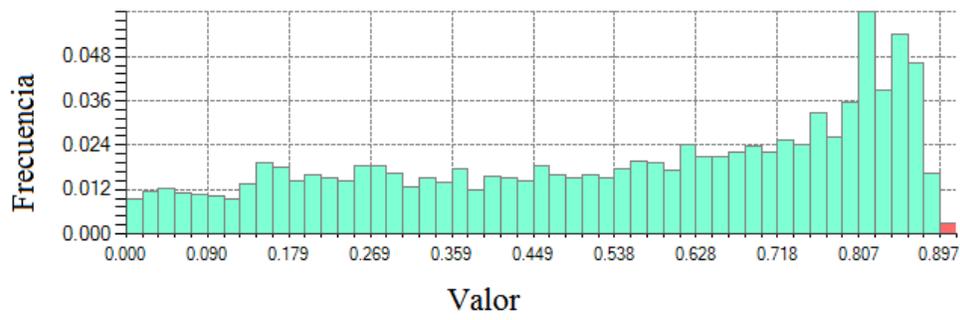


Estadísticas

Promedio	0.539
Varianza	0.069
Mínimo	0.000
Máximo	0.897

Anexos 3 Histograma y estadística datos condicionantes 3

Datos condicionantes 4



Estadísticas

Promedio	0.547
Varianza	0.068
Mínimo	0.000
Máximo	0.897

Anexos 4 Histograma y estadística DC4

11 ANEXO B

Análisis estadístico (promedio y varianza), sobre cada realización, para cada configuración de parámetros.

Anexos 5 Tabla de los promedios de cada realización para cada combinación de estado de las variables. Método patrón más cercano: template pixels.

Método patrón más cercano: Template Pixels									
Configuración		11x11x7x7				21x21x7x7			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.6015	0.5893	0.5823	0.5652	0.6205	0.5821	0.5705	0.5685
	Primer cuartil	0.6063	0.5984	0.5869	0.5686	0.6271	0.5882	0.5749	0.5741
	Promedio	0.6115	0.6011	0.5907	0.5706	0.6304	0.5909	0.5777	0.5758
	Tercer cuartil	0.6163	0.6050	0.5936	0.5730	0.6339	0.5936	0.5811	0.5780
	Máximo	0.6246	0.6121	0.5996	0.5778	0.6453	0.6001	0.5895	0.5813
Configuración		21x21x13x13				41x41x7x7			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.5934	0.5750	0.5660	0.5637	0.5948	0.5683	0.5657	0.5604
	Primer cuartil	0.6056	0.5800	0.5706	0.5690	0.5998	0.5758	0.5740	0.5646
	Promedio	0.6104	0.5837	0.5746	0.5733	0.6027	0.5779	0.5756	0.5664
	Tercer cuartil	0.6157	0.5881	0.5783	0.5767	0.6054	0.5797	0.5780	0.5681
	Máximo	0.6276	0.5957	0.5841	0.5821	0.6119	0.5873	0.5803	0.5720
Configuración		41x41x13x13				41x41x25x25			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.5893	0.5719	0.5703	0.5620	0.5782	0.5656	0.5646	0.5578
	Primer cuartil	0.5986	0.5769	0.5722	0.5643	0.5868	0.5734	0.5712	0.5622
	Promedio	0.6012	0.5790	0.5759	0.5665	0.5908	0.5773	0.5751	0.5648
	Tercer cuartil	0.6052	0.5810	0.5790	0.5685	0.5951	0.5810	0.5792	0.5678
	Máximo	0.6132	0.5916	0.5857	0.5740	0.6063	0.5888	0.5864	0.5714

Anexos 6 Tabla de las varianzas de cada realización para cada combinación de estado de las variables. Método patrón más cercano: template pixels.

Método patrón más cercano: Template Pixels									
		11x11x7x7				21x21x7x7			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0595	0.0608	0.0621	0.0668	0.0541	0.0621	0.0635	0.0655
	Primer cuartil	0.0605	0.0625	0.0642	0.0680	0.0571	0.0633	0.0646	0.0660
	Promedio	0.0616	0.0631	0.0646	0.0684	0.0576	0.0639	0.0652	0.0666
	Tercer cuartil	0.0624	0.0638	0.0653	0.0689	0.0585	0.0645	0.0658	0.0670
	Máximo	0.0642	0.0652	0.0666	0.0696	0.0596	0.0656	0.0667	0.0683
Configuración		21x21x13x13				41x41x7x7			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0576	0.0628	0.0627	0.0645	0.0610	0.0641	0.0640	0.0646
	Primer cuartil	0.0602	0.0640	0.0652	0.0658	0.0618	0.0647	0.0646	0.0654
	Promedio	0.0608	0.0647	0.0657	0.0666	0.0623	0.0651	0.0651	0.0658
	Tercer cuartil	0.0616	0.0655	0.0662	0.0673	0.0629	0.0655	0.0656	0.0662
	Máximo	0.0640	0.0664	0.0673	0.0695	0.0638	0.0667	0.0664	0.0675
Configuración		41x41x13x13				41x41x25x25			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0607	0.0637	0.0624	0.0640	0.0617	0.0627	0.0620	0.0629
	Primer cuartil	0.0616	0.0645	0.0639	0.0652	0.0622	0.0642	0.0643	0.0648
	Promedio	0.0623	0.0650	0.0648	0.0659	0.0631	0.0651	0.0651	0.0657
	Tercer cuartil	0.0630	0.0655	0.0657	0.0664	0.0639	0.0657	0.0659	0.0666
	Máximo	0.0639	0.0669	0.0666	0.0673	0.0658	0.0675	0.0676	0.0679

Anexos 7 Tabla de los promedios de cada realización para cada combinación de estado de las variables. Método patrón más cercano: filter scores.

Método patrón más cercano: Filter scores									
Configuración		11x11x7x7				21x21x7x7			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.5798	0.5616	0.5620	0.5678	0.5881	0.5713	0.5677	0.5633
	Primer cuartil	0.5860	0.5731	0.5695	0.5719	0.5922	0.5763	0.5755	0.5703
	Promedio	0.5887	0.5770	0.5723	0.5748	0.5960	0.5804	0.5792	0.5736
	Tercer cuartil	0.5917	0.5797	0.5761	0.5764	0.5996	0.5832	0.5833	0.5774
	Máximo	0.6022	0.5884	0.5812	0.5828	0.6077	0.5909	0.5879	0.5839
Configuración		21x21x13x13				41x41x7x7			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.5765	0.5663	0.5645	0.5646	0.5779	0.5791	0.5765	0.5691
	Primer cuartil	0.5846	0.5737	0.5738	0.5703	0.5892	0.5869	0.5851	0.5774
	Promedio	0.5897	0.5782	0.5781	0.5732	0.5949	0.5904	0.5881	0.5799
	Tercer cuartil	0.5939	0.5831	0.5817	0.5768	0.6009	0.5933	0.5916	0.5829
	Máximo	0.6098	0.5895	0.5908	0.5808	0.6114	0.6001	0.6001	0.5905
Configuración		41x41x13x13				41x41x25x25			
Promedio	Datos Condicionantes	DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
	Mínimo	0.5849	0.5765	0.5768	0.5728	0.5757	0.5759	0.5718	0.5733
	Primer cuartil	0.5884	0.5832	0.5835	0.5776	0.5832	0.5865	0.5824	0.5776
	Promedio	0.5922	0.5872	0.5880	0.5799	0.5888	0.5901	0.5860	0.5814
	Tercer cuartil	0.5950	0.5895	0.5914	0.5825	0.5933	0.5936	0.5909	0.5845
	Máximo	0.6031	0.6070	0.6000	0.5887	0.6032	0.5978	0.6019	0.5893

Anexos 8 Tabla de las varianzas de cada realización para cada combinación de estado de las variables. Método patrón más cercano: filter scores.

Método patrón más cercano: Filter scores									
		11x11x7x7				21x21x7x7			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0568	0.0595	0.0607	0.0629	0.0577	0.0612	0.0613	0.0620
	Primer cuartil	0.0588	0.0607	0.0616	0.0641	0.0594	0.0624	0.0623	0.0635
	Promedio	0.0595	0.0614	0.0621	0.0646	0.0600	0.0627	0.0631	0.0640
	Tercer cuartil	0.0601	0.0622	0.0626	0.0651	0.0606	0.0633	0.0641	0.0645
	Máximo	0.0616	0.0640	0.0637	0.0664	0.0620	0.0639	0.0662	0.0655
Configuración		21x21x13x13				41x41x7x7			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0593	0.0618	0.0622	0.0636	0.0604	0.0609	0.0608	0.0616
	Primer cuartil	0.0613	0.0636	0.0636	0.0647	0.0615	0.0623	0.0618	0.0628
	Promedio	0.0620	0.0643	0.0644	0.0653	0.0624	0.0628	0.0624	0.0632
	Tercer cuartil	0.0629	0.0650	0.0652	0.0660	0.0632	0.0635	0.0629	0.0635
	Máximo	0.0644	0.0665	0.0669	0.0676	0.0656	0.0642	0.0641	0.0644
Configuración		41x41x13x13				41x41x25x25			
Datos Condicionantes		DC1	DC2	DC3	DC4	DC1	DC2	DC3	DC4
Varianza	Mínimo	0.0599	0.0600	0.0602	0.0608	0.0598	0.0594	0.0591	0.0603
	Primer cuartil	0.0616	0.0623	0.0615	0.0621	0.0612	0.0611	0.0608	0.0612
	Promedio	0.0625	0.0629	0.0622	0.0627	0.0621	0.0617	0.0615	0.0617
	Tercer cuartil	0.0632	0.0636	0.0628	0.0631	0.0627	0.0624	0.0620	0.0620
	Máximo	0.0646	0.0648	0.0638	0.0641	0.0649	0.0638	0.0633	0.0629

Anexos 9 Tabla con los promedio sobre cada simulación (promedio de las realizaciones)

Promedio	Método	Template pixels						Filter score							
	Grilla de búsqueda	11x11	21x21			41x41			11x11	21x21			41x41		
	Grilla interna	7x7	7x7	13x13	7x7	13x13	25x25	7x7	7x7	13x13	7x7	13x13	25x25		
	Datos condicionantes 1	0.612	0.630	0.610	0.603	0.601	0.591	0.589	0.596	0.590	0.595	0.592	0.589		
	Datos condicionantes 2	0.601	0.591	0.584	0.578	0.579	0.577	0.577	0.580	0.578	0.590	0.587	0.590		
	Datos condicionantes 3	0.591	0.578	0.575	0.576	0.579	0.575	0.572	0.579	0.578	0.588	0.588	0.586		
	Datos condicionantes 4	0.571	0.576	0.573	0.566	0.567	0.565	0.575	0.574	0.573	0.580	0.580	0.581		

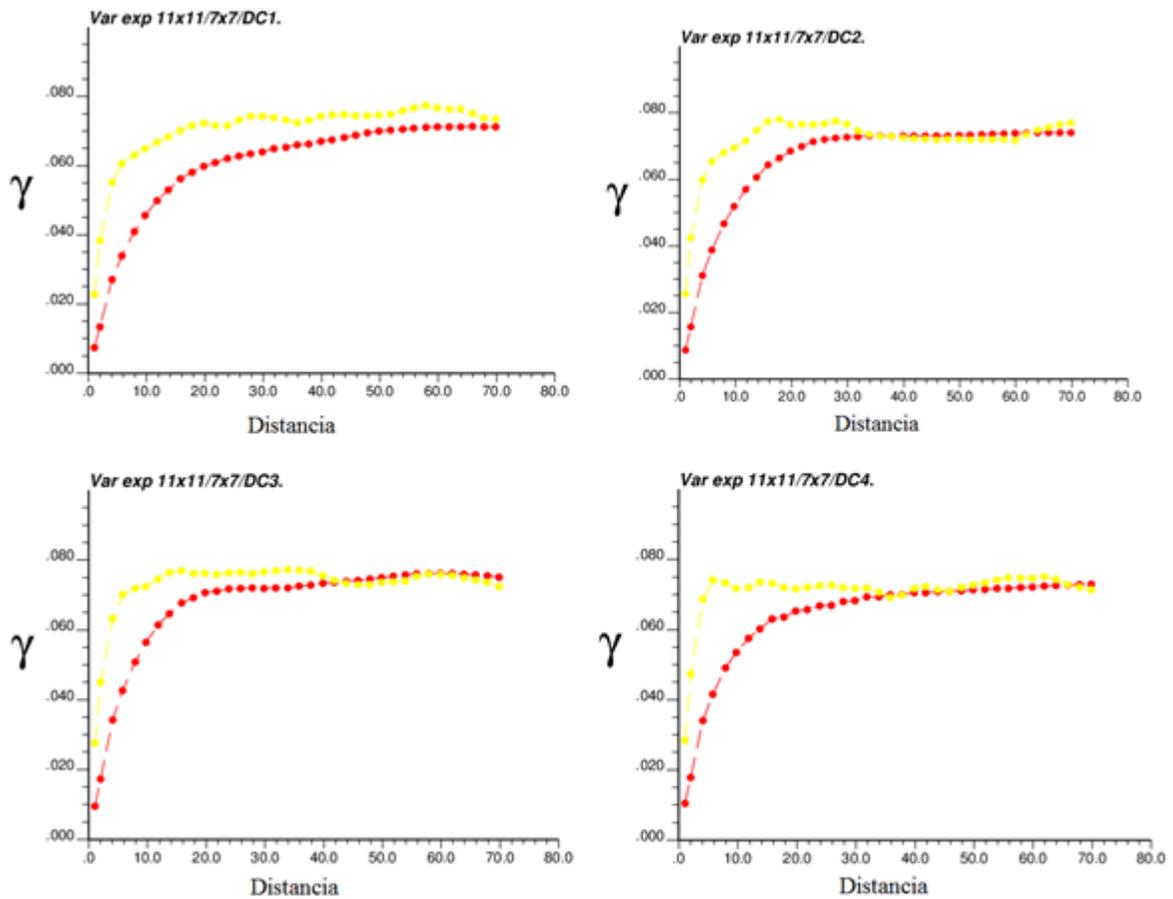
Anexos 10 Tabla con las varianzas sobre cada simulación (promedio de las realizaciones)

Varianza	Método	Template pixels						Filter score							
	Grilla de búsqueda	11x11	21x21			41x41			11x11	21x21			41x41		
	Grilla interna	7x7	7x7	13x13	7x7	13x13	25x25	7x7	7x7	13x13	7x7	13x13	25x25		
	Datos condicionantes 1	0.0034	0.0045	0.0039	0.0058	0.0050	0.0044	0.0034	0.0032	0.0033	0.0034	0.0032	0.0030		
	Datos condicionantes 2	0.0093	0.0114	0.0098	0.0120	0.0100	0.0090	0.0078	0.0056	0.0063	0.0048	0.0046	0.0005		
	Datos condicionantes 3	0.0130	0.0142	0.0120	0.0140	0.0126	0.0110	0.0092	0.0064	0.0068	0.0049	0.0050	0.0063		
	Datos condicionantes 4	0.0282	0.0225	0.0200	0.0235	0.0210	0.0200	0.0164	0.0110	0.0120	0.0083	0.0086	0.0104		

12 ANEXO C

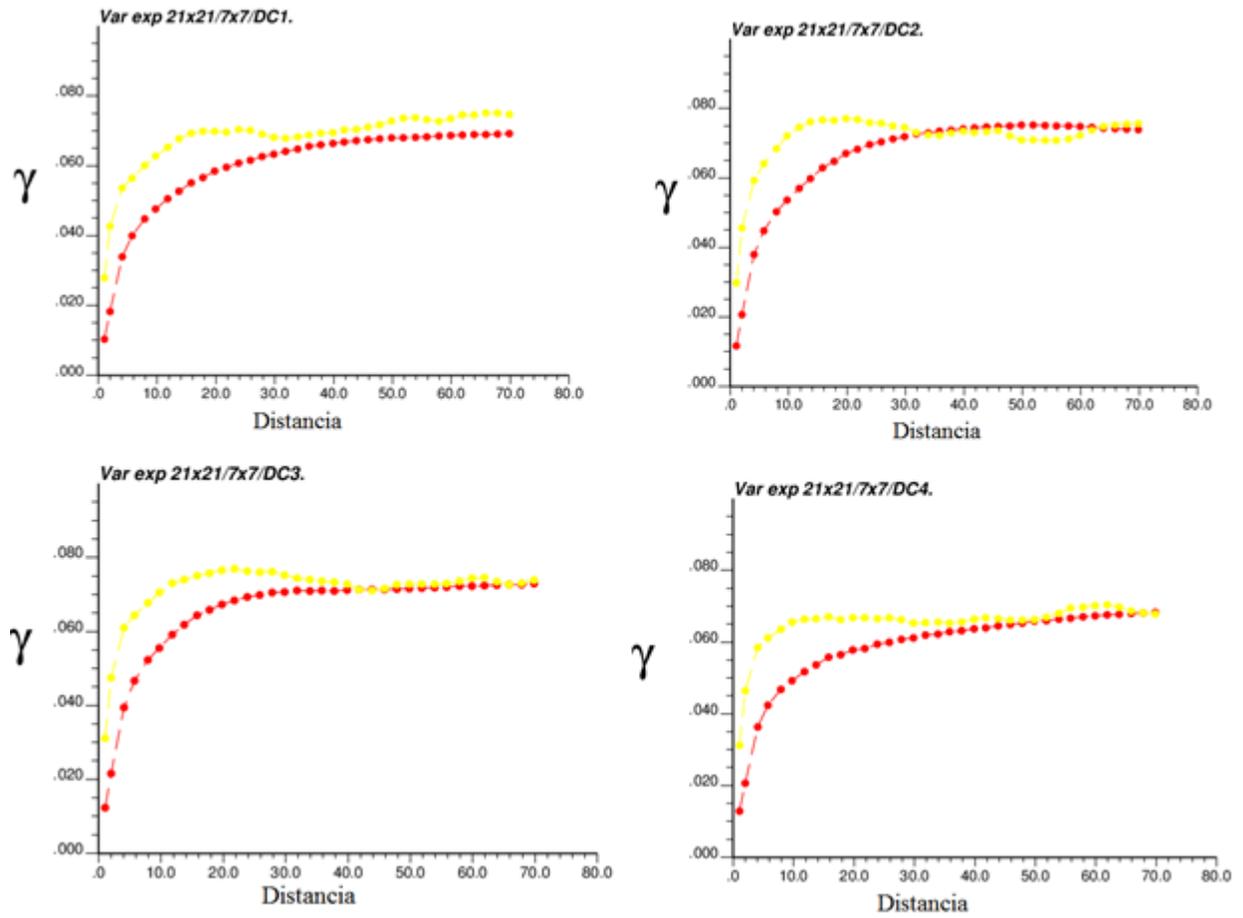
Variogramas experimentales asociados a una de las 35 realizaciones (elegida aleatoriamente). Largo del paso: 2, número de pasos: 35, tolerancia en el paso: 1, y se analizan 2 direcciones: N85°E (color rojo) y N5°O (color amarillo) ya que son las direcciones de anisotropía.

Método: *template pixels*. Grilla de búsqueda: 11x11. Grilla interna: 7x7



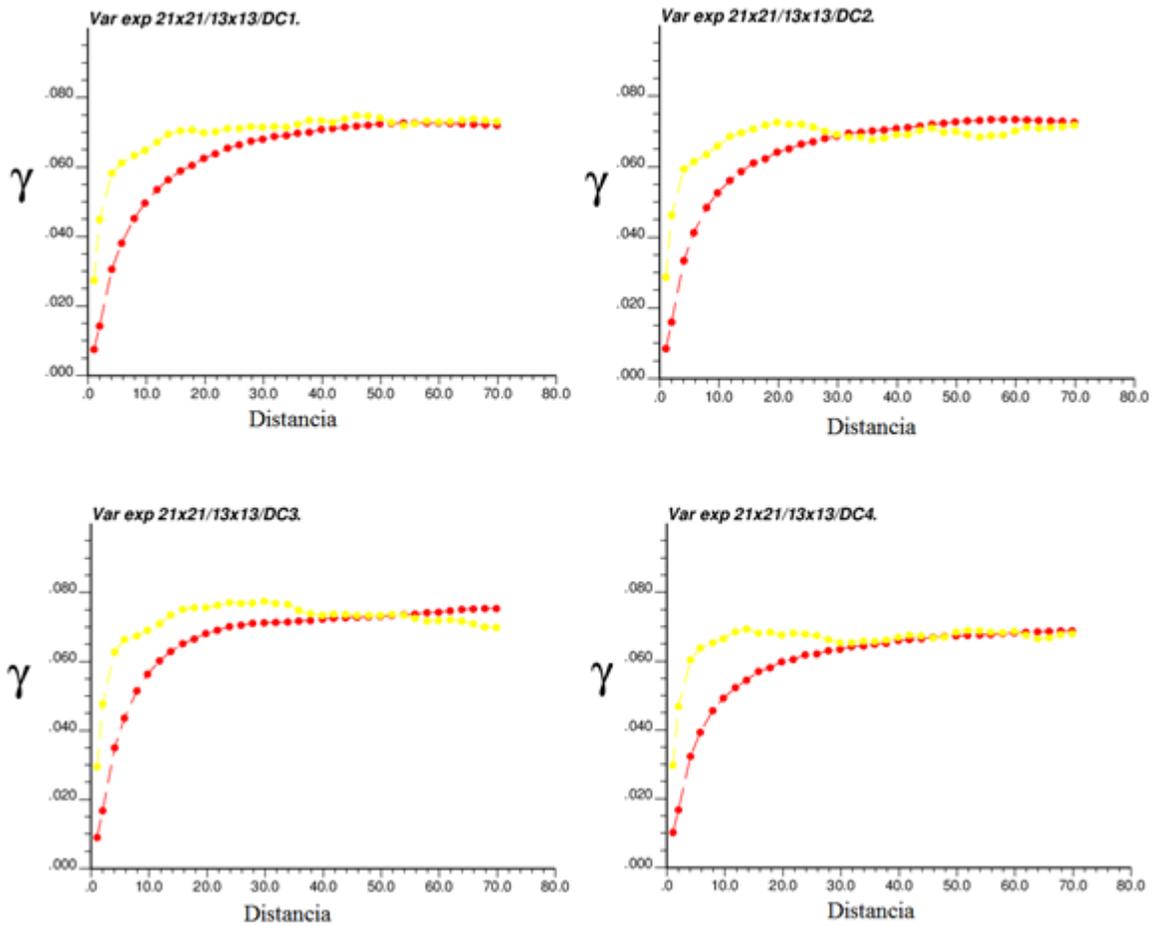
Anexos 11 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 11x11. Grilla interna: 7x7

Método: *template pixels*. Grilla de búsqueda: 21x21. Grilla interna: 7x7



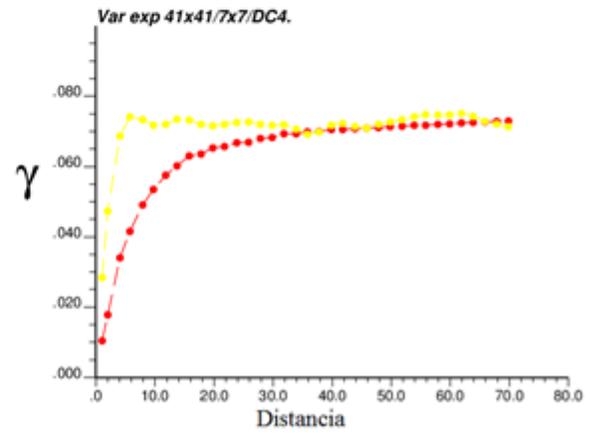
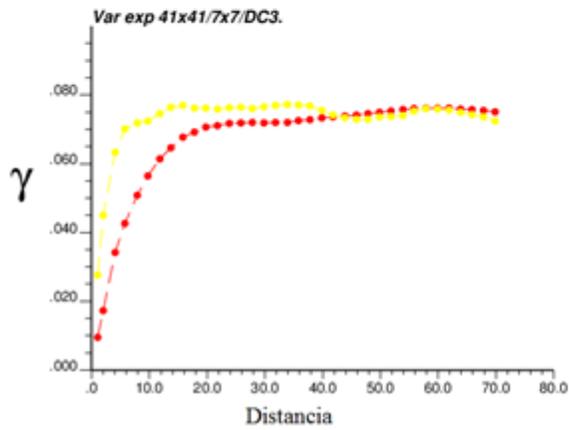
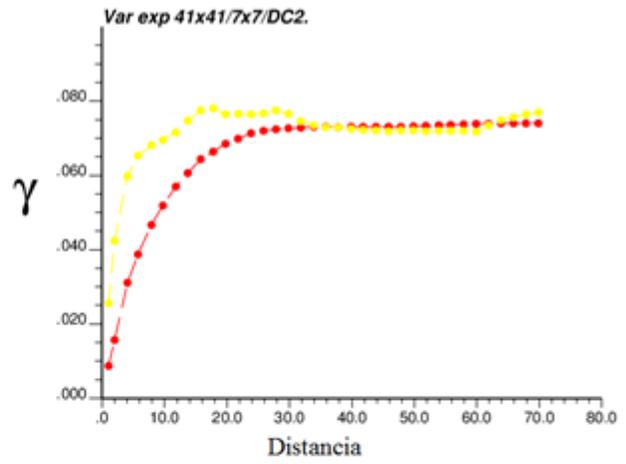
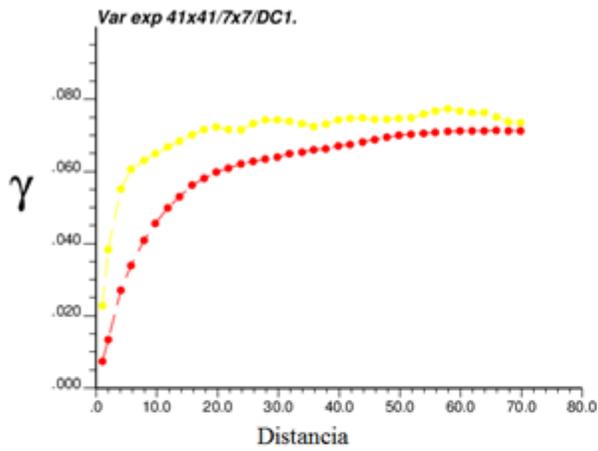
Anexos 12 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 21x21. Grilla interna: 7x7

Método: *template pixels*. Grilla de búsqueda: 21x21. Grilla interna: 13x13



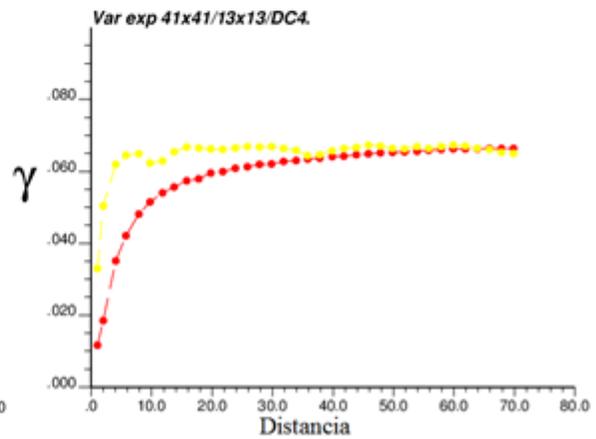
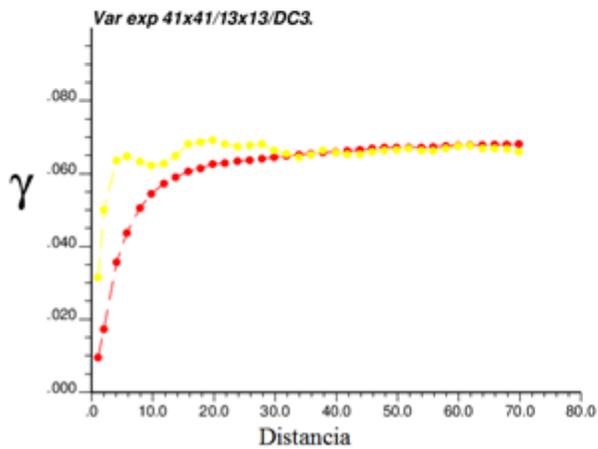
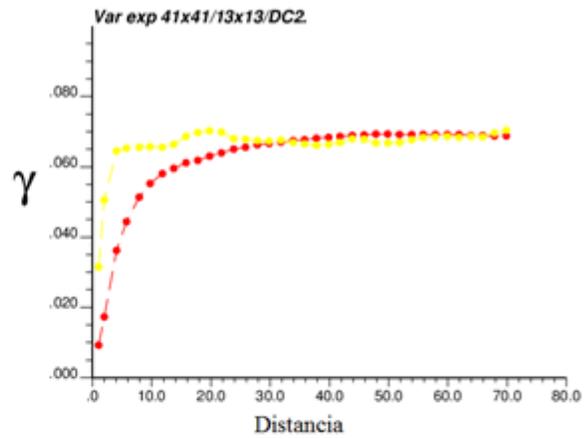
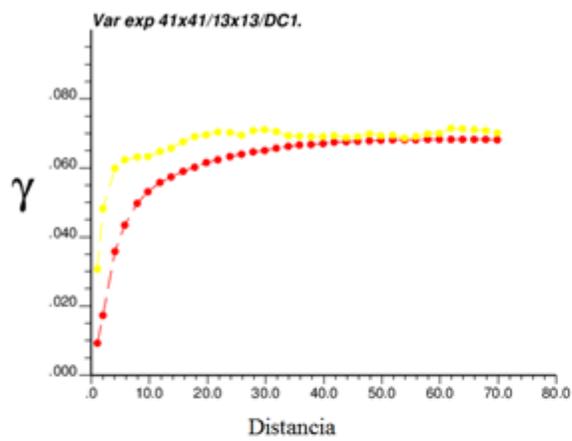
Anexos 13 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 21x21. Grilla interna: 13x13

Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 7x7



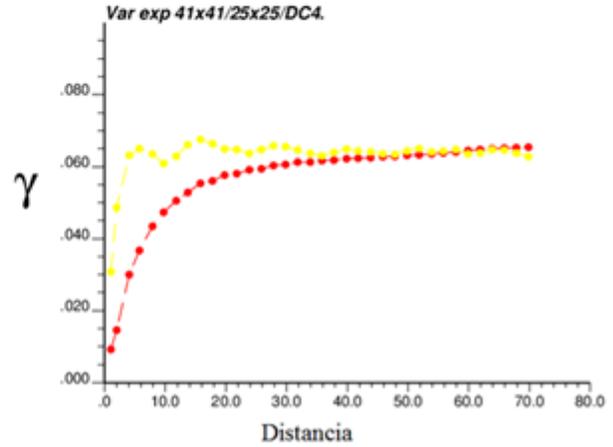
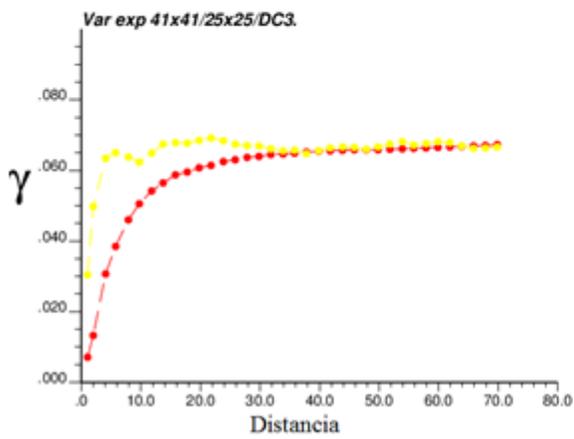
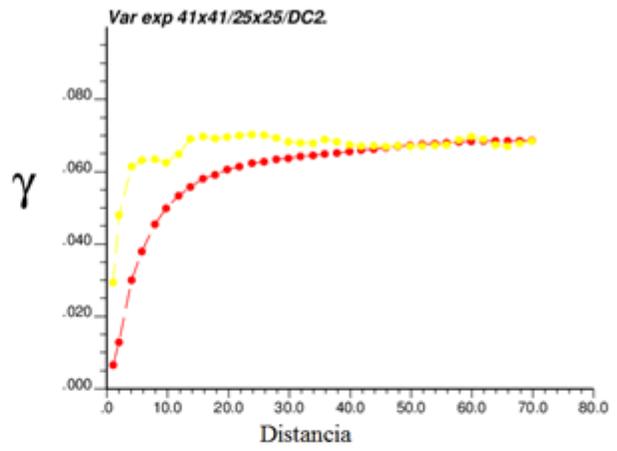
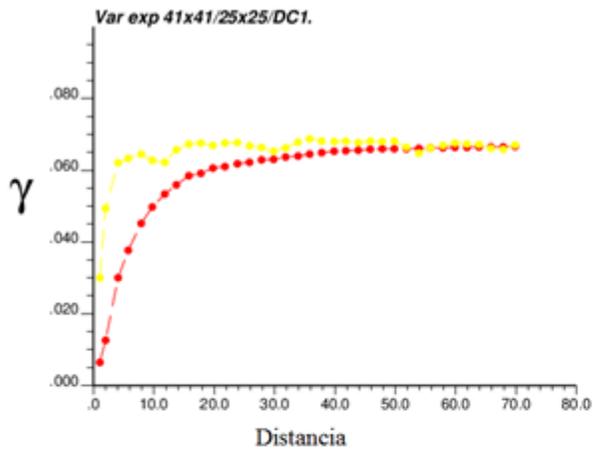
Anexos 14 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 7x7

Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 13x13



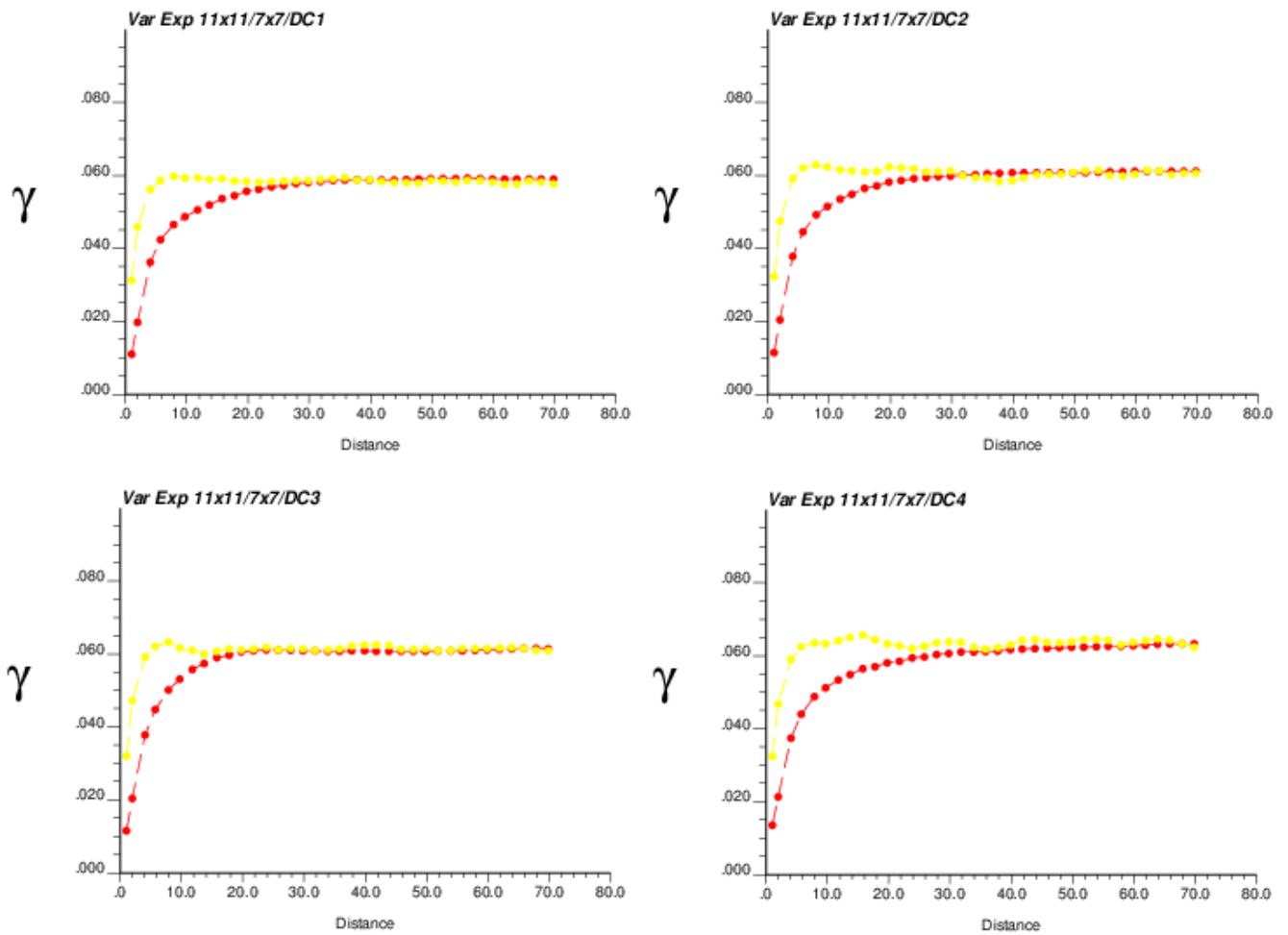
Anexos 15 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 13x13

Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 25x25



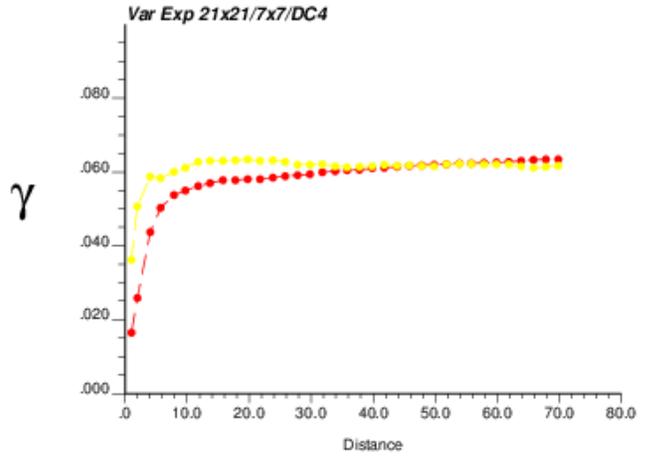
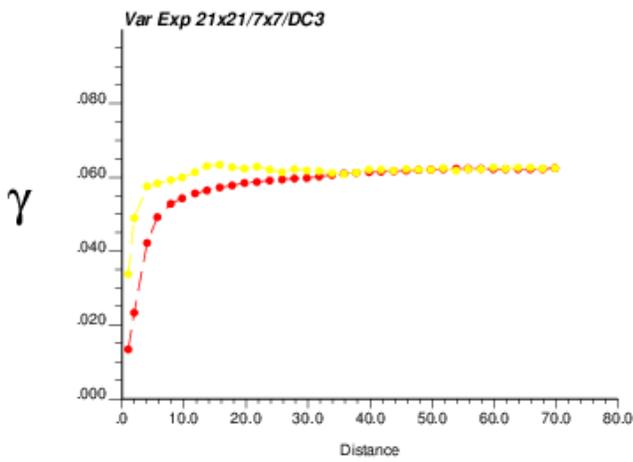
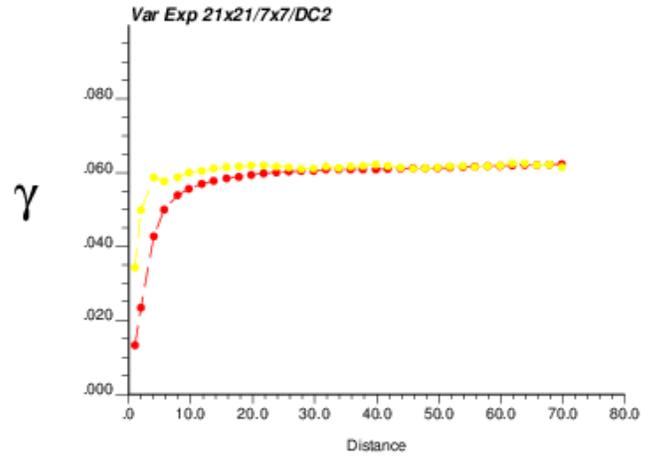
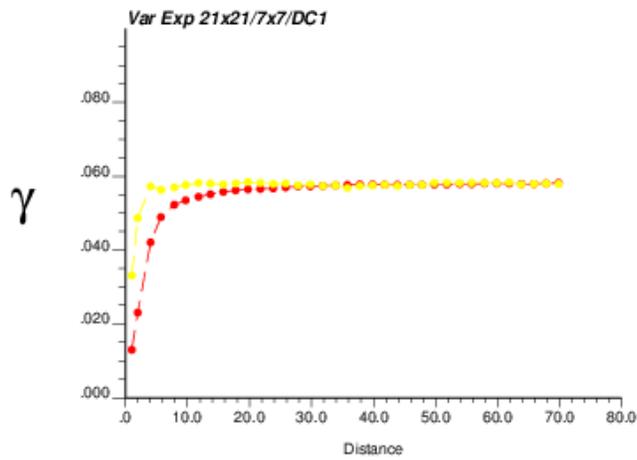
Anexos 16 Variogramas experimentales. Método: *template pixels*. Grilla de búsqueda: 41x41. Grilla interna: 25x25

Método: *filter scores*. Grilla de búsqueda: 11x11. Grilla interna: 7x7



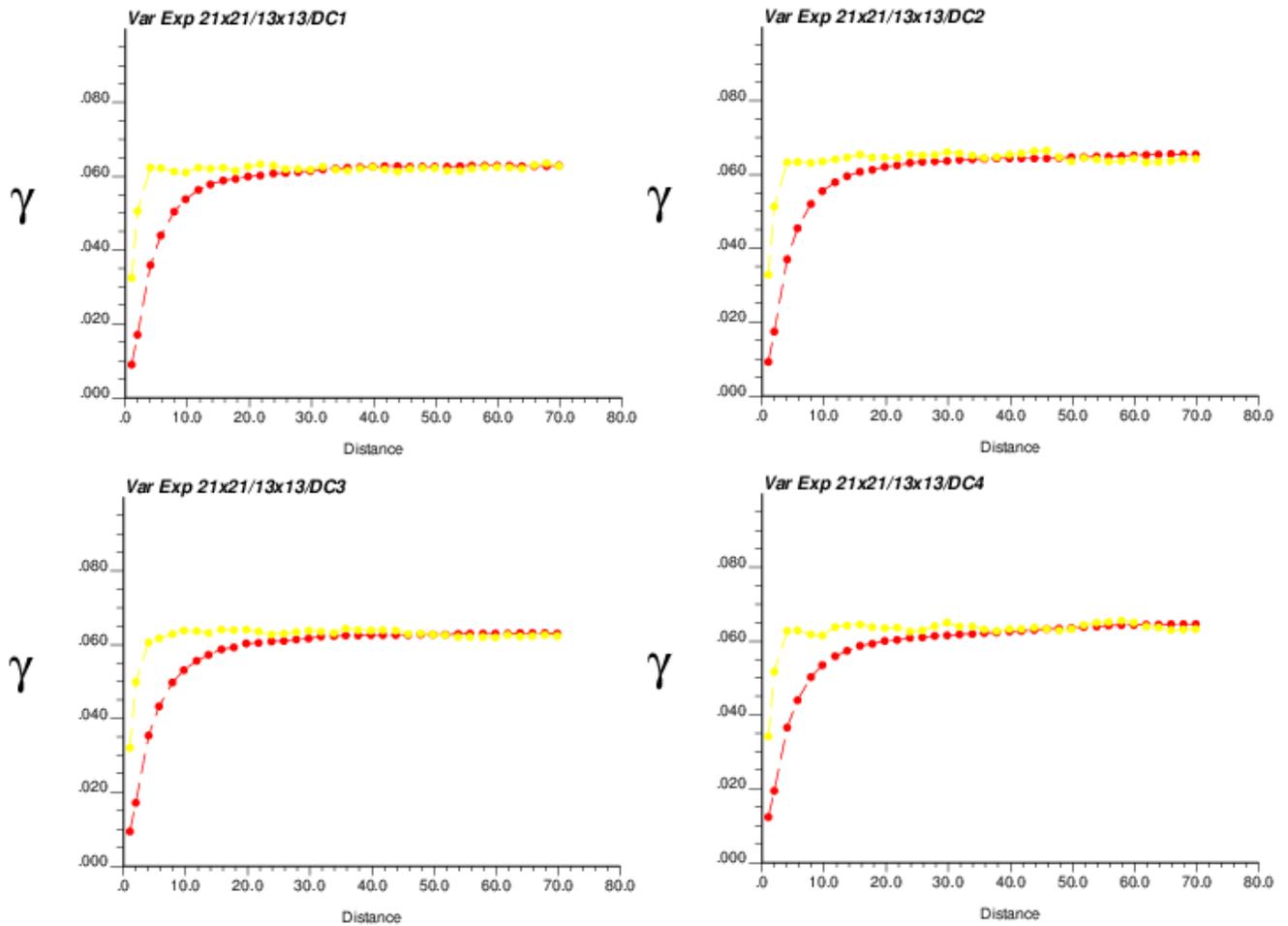
Anexos 17 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 11x11. Grilla interna: 7x7

Método: *filter scores*. Grilla de búsqueda: 21x21. Grilla interna: 7x7



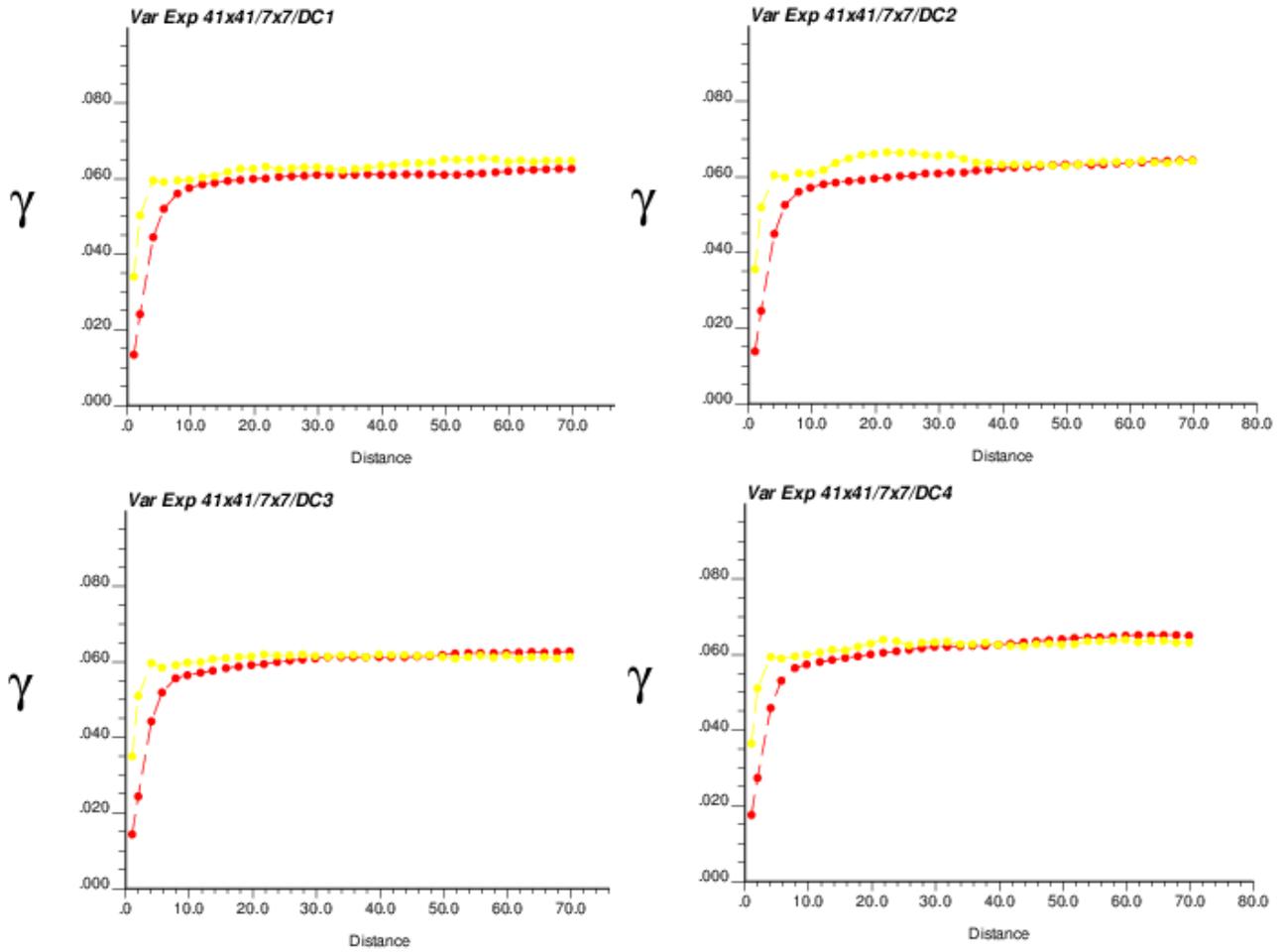
Anexos 18 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 21x21. Grilla interna: 7x7

Método: *filter scores*. Grilla de búsqueda: 21x21. Grilla interna: 13x13



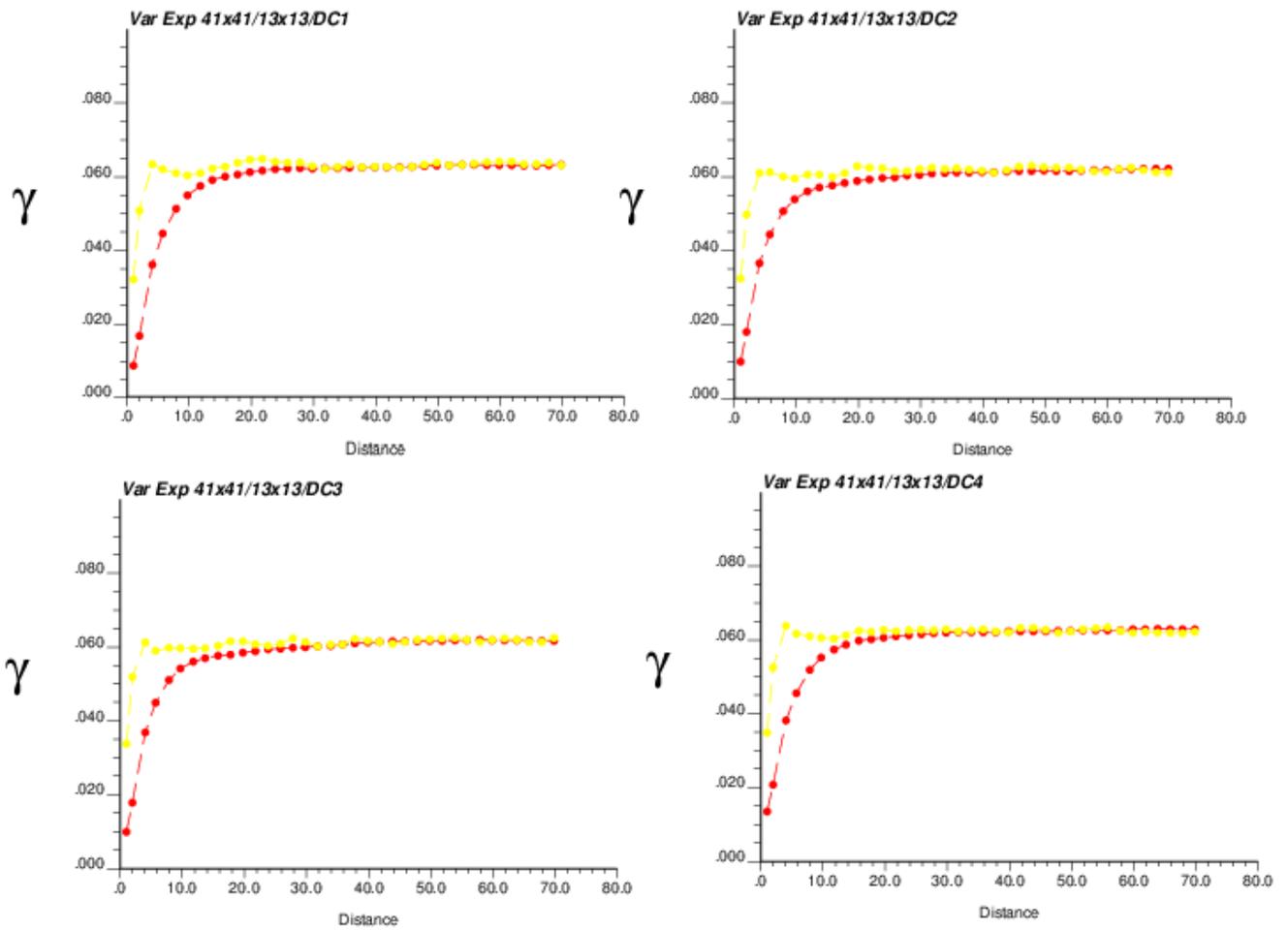
Anexos 19 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 21x21. Grilla interna: 13x13

Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 7x7



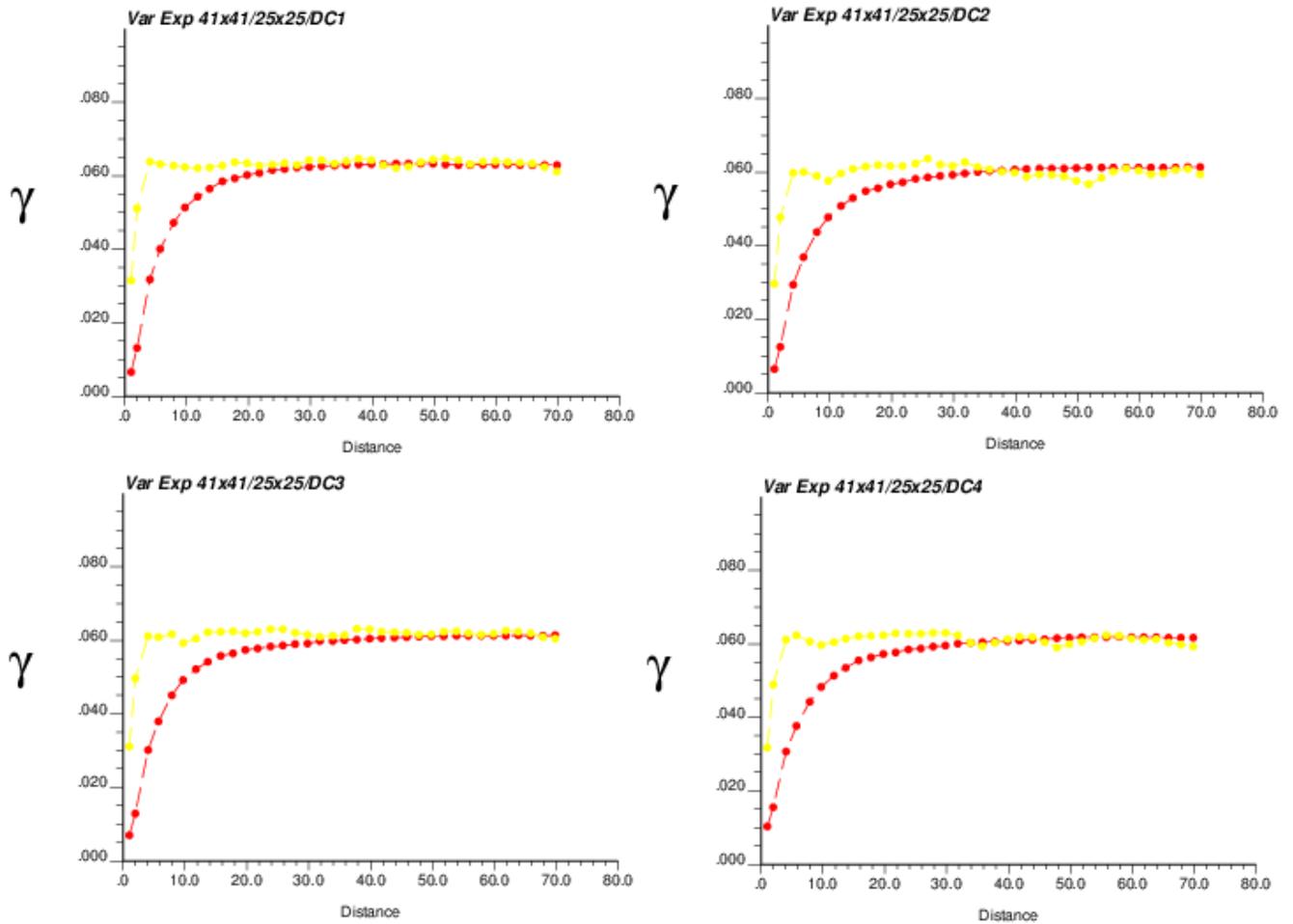
Anexos 20 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 7x7

Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 13x13



Anexos 21 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 13x13

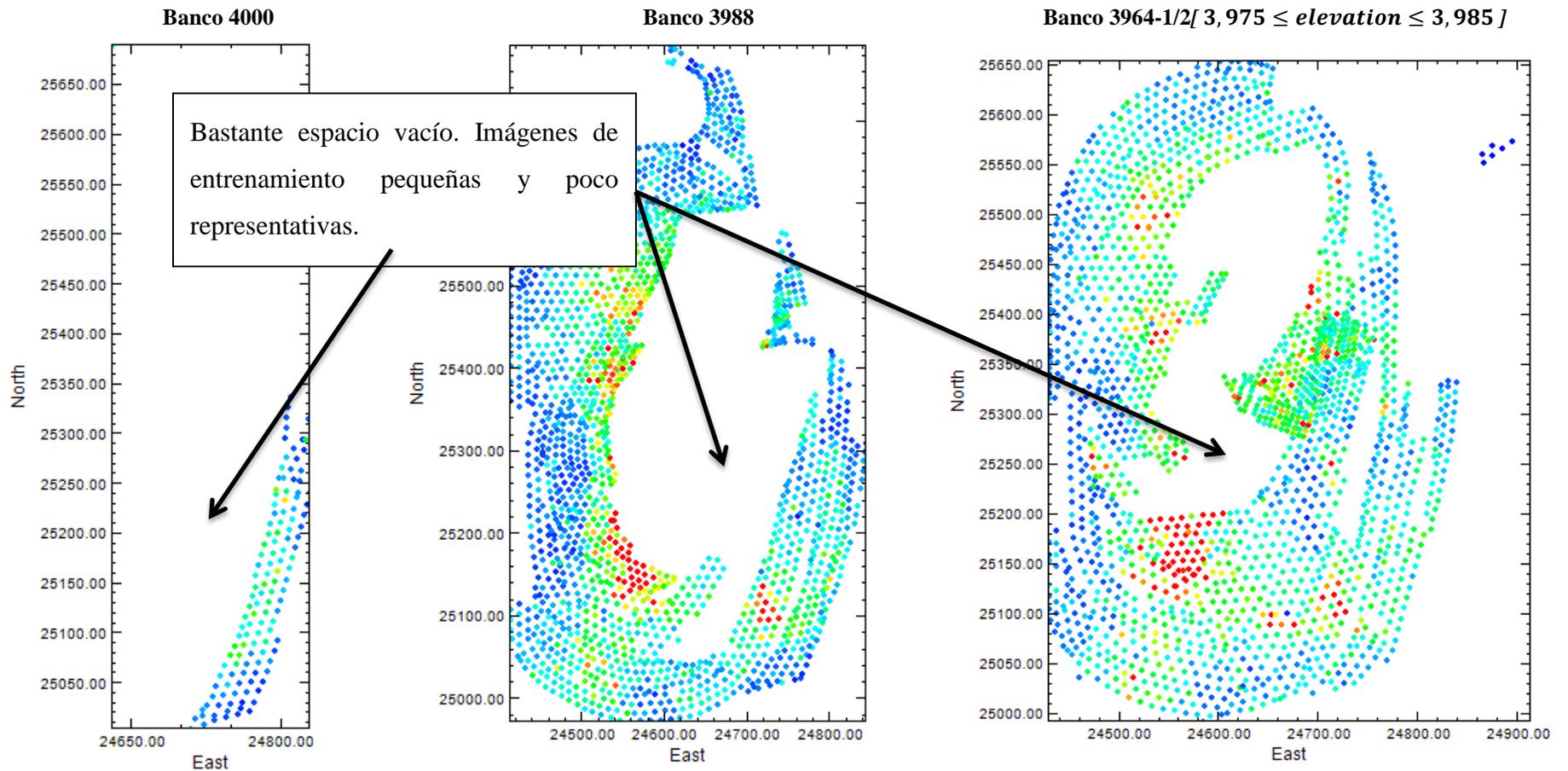
Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 25x25



Anexos 22 Variogramas experimentales. Método: *filter scores*. Grilla de búsqueda: 41x41. Grilla interna: 25x25

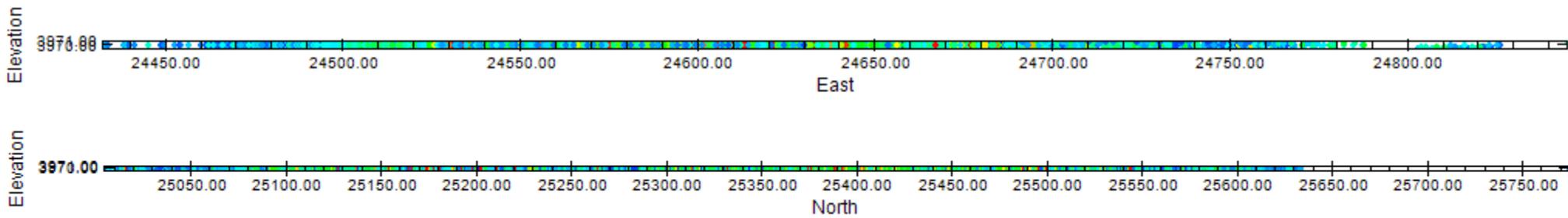
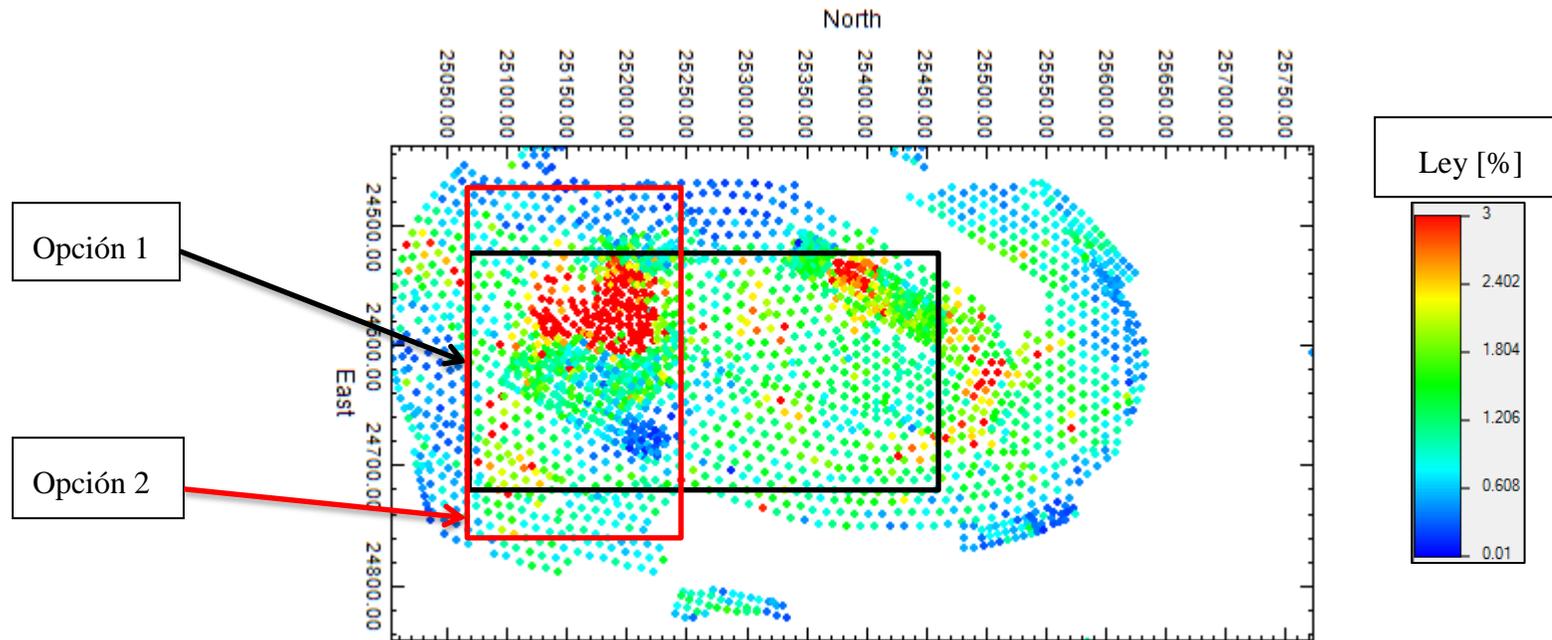
13 ANEXO D

El anexo presenta las posibles imágenes de entrenamiento. *(Las vistas XZ e YZ están para presentar solo el espesor)*



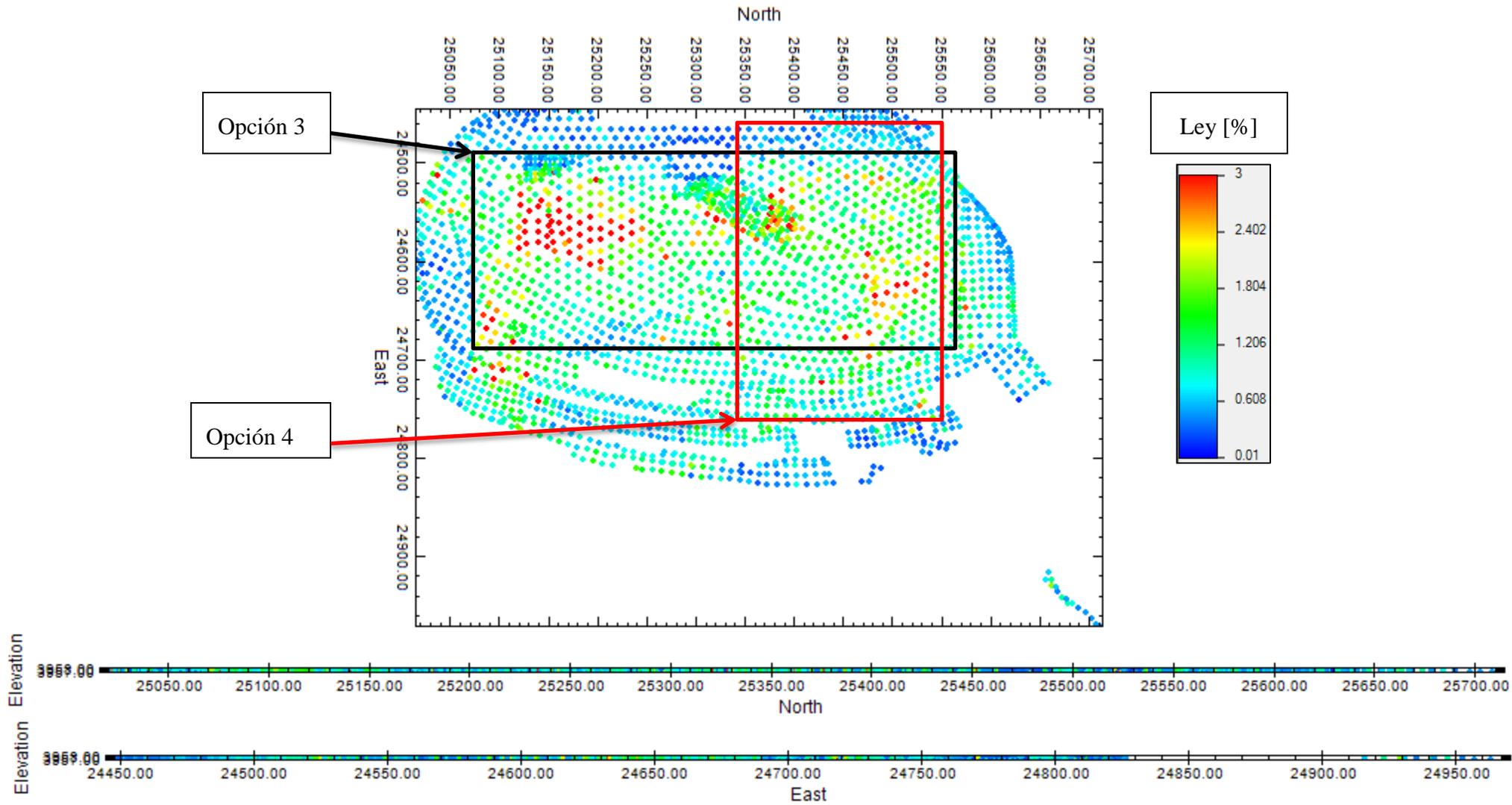
Anexos 23 Primeros 3 bancos - pozos de tronadura

Banco 3964-2/2 [3,965 ≤ elevation ≤ 3,974.9]



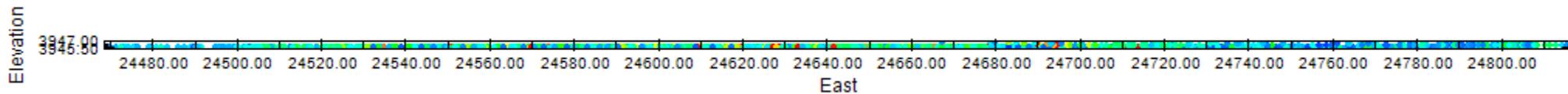
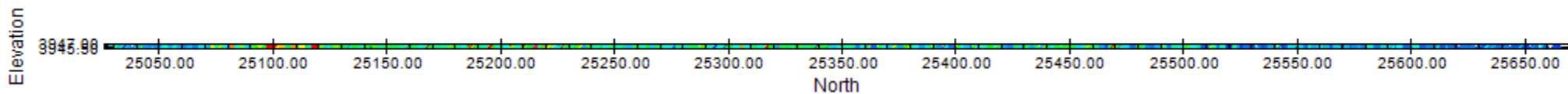
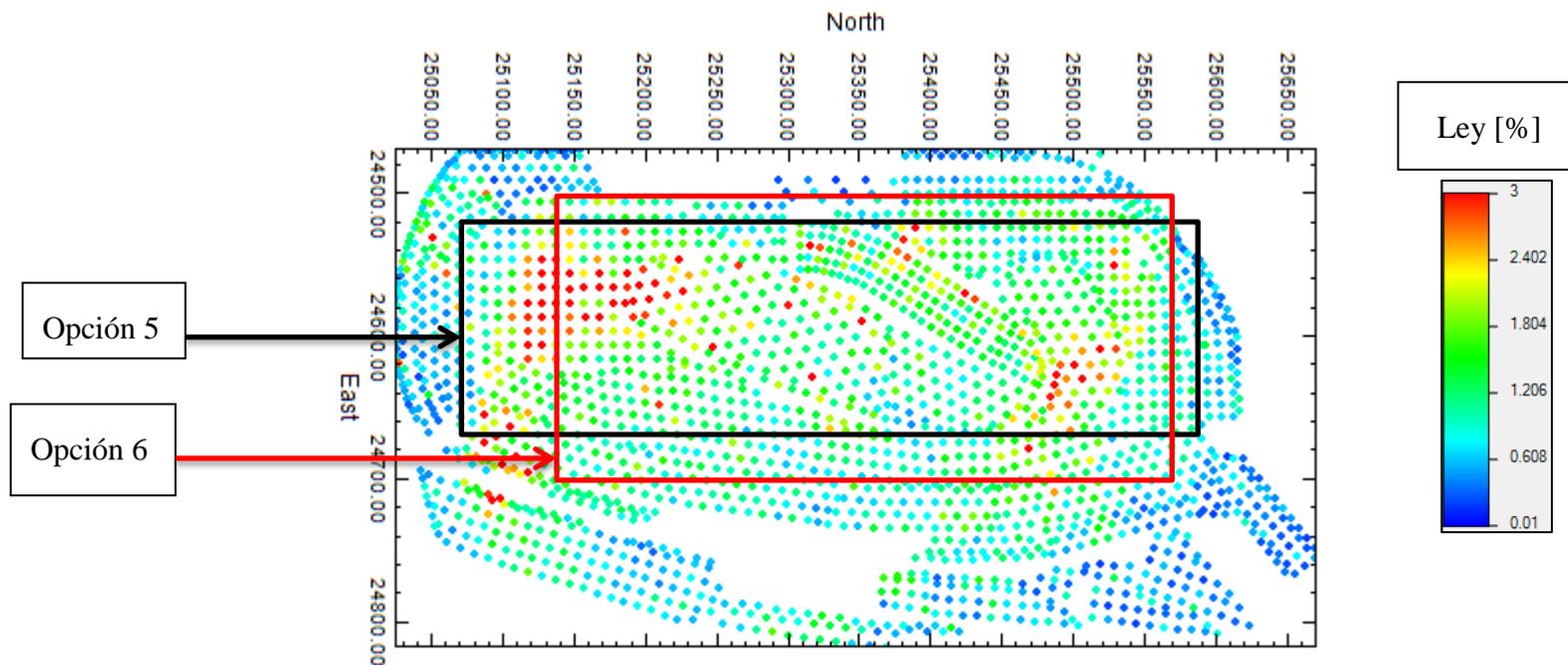
Anexos 24 Posibles T.Is en Banco 3964-2/2

Banco 3952 [3,950 ≤ elevation ≤ 3,960]



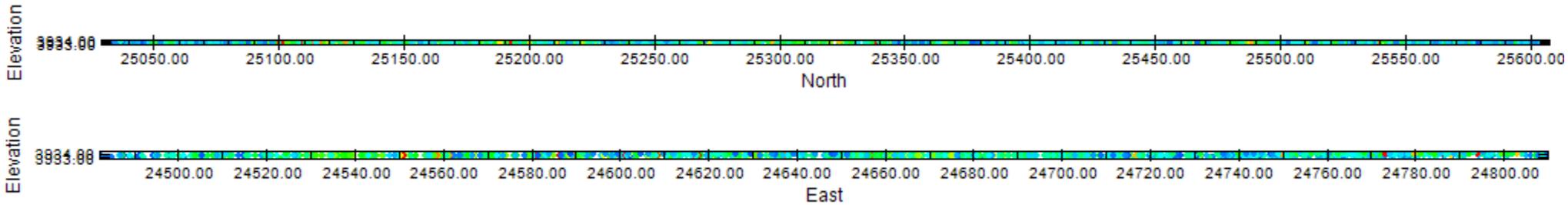
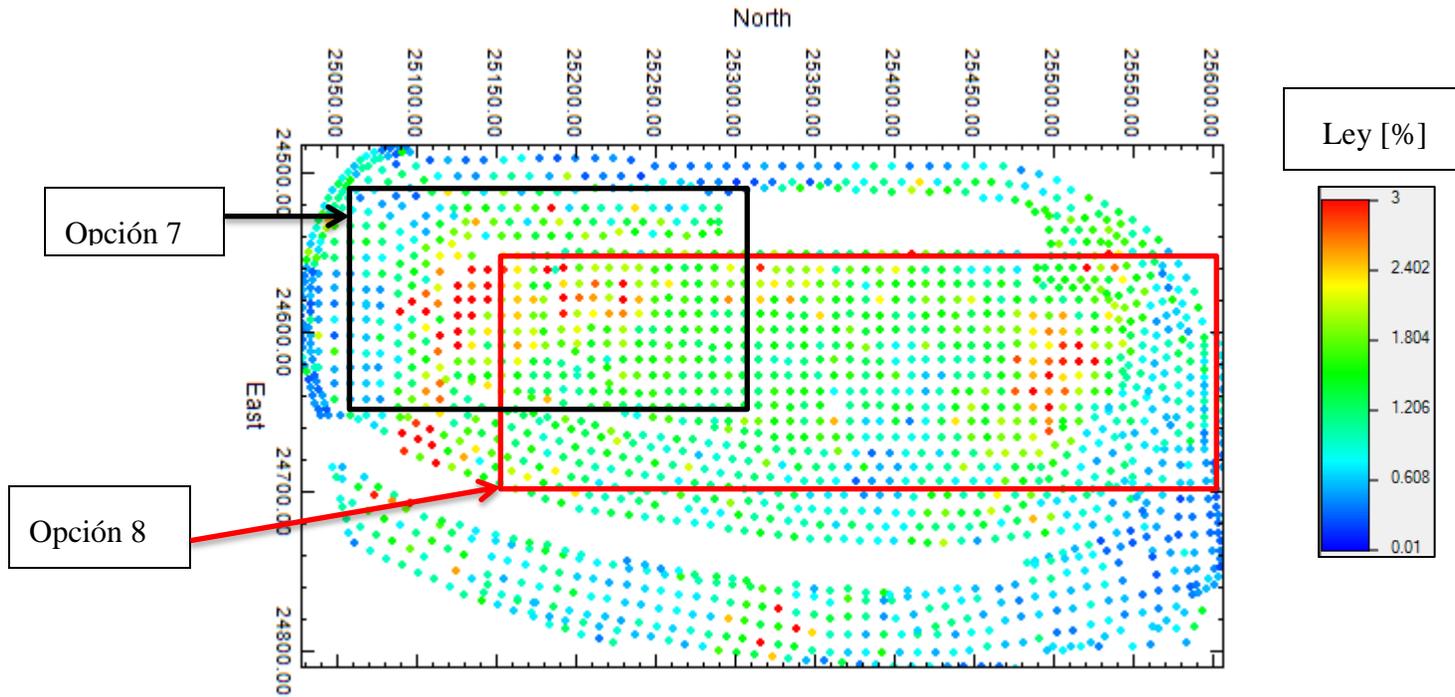
Anexos 25 Posibles T.I.s en banco 3952

Banco 3940 [3,945 ≤ elevation ≤ 3,947]



Anexos 26 Posibles T.I.s en banco 3940

Banco 3928 [3,925 ≤ elevation ≤ 3,935]



Anexos 27 Posibles T.I.s en banco 3928