



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

SIMULACIÓN MULTIPUNTOS UTILIZANDO EL ALGORITMO FILTERSIM:
INCORPORACIÓN DE DERIVAS EN LA MEDIA LOCAL

TESIS PARA OPTAR AL GRADO DE MAGISTER EN MINERÍA

SEBASTIÁN ALEJANDRO ÁVALOS SOTOMAYOR

PROFESOR GUÍA:
JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:
XAVIER EMERY
BRIAN TOWNLEY CALLEJAS
PAULA LARRONDO

SANTIAGO DE CHILE
2014

Resumen

Durante los últimos años, el desarrollo y aplicación de algoritmos basados en estadísticas multipuntos ha ido en aumento. La presente tesis considera la incorporación de derivas externas, asociadas a la media local, al algoritmo Filter-Based Simulation (Filtersim). Para lograrlo se programó en Python el algoritmo original. Luego, una modificación interna al algoritmo permite incorporar derivas en la media. Para ello un nuevo ponderador, w_0 , mide el nivel de influencia que tendrá en la elección del patrón en el proceso de simulación.

Dos casos de estudio son llevados a cabo. El primero busca reproducir el comportamiento sinusoidal que presentan los valores altos de una variable continua en la imagen del corte pulido de una roca sedimentaria. De ella se infiere el comportamiento estructural mientras que de otra zona conocida, con características similares, son extraídos los datos condicionantes y las derivas externas. Se simula a distintos grados de información por parte de la deriva y distinto número de datos condicionantes. Se deriva de lo anterior que un óptimo porcentaje de influencia de la deriva en la media oscila entre un 40 % y un 60 %, logrando caracterizar adecuadamente el comportamiento sinusoidal y la continuidad espacial de estas. De la variación en la cantidad de datos condicionantes se concluye que considerar entre un 0.5 % y un 1.0 % del dominio informado ayuda a mejorar los resultados, y las derivas obtenidas de ellos tienen un alto grado de información. Además, se reafirma la capacidad de los algoritmos multipuntos de reproducir estructuras complejas en contraste con la incapacidad de los métodos de estimación/simulación convencionales basados en el variograma. El segundo caso de estudio se aplica en el yacimiento de Compañía Minera Cerro Colorado, específicamente en el sector Sur-Oeste entre los bancos 2395 y 2355. Se busca reproducir la ley de cobre total de los bancos 2385 al 2355 extrayendo del banco 2395 la imagen de entrenamiento para la obtención de patrones mediante Filtersim. Las simulaciones se condicionan con los sondeos de exploración y se contrastan con los pozos de tronadura asumiendo su representación de la realidad. Se utiliza un 40 % de influencia en la media local. Los resultados muestran una buena aproximación a la realidad al utilizar Filtersim convencional, y al incorporar derivas se disminuye la dispersión entre los resultados obtenidos, mejorando con esto la reproducción de las leyes. No se observa un beneficio claro y considerable a la hora de incorporar derivas en la media. Esto debido, posiblemente, a que el proceso geológico de formación asociado al yacimiento es más bien masivo careciendo de estructuras geológicas con forma y continuidad bien marcadas. En específico se obtiene una mejor cuantificación de los recursos, a través de Curvas Tonelaje-Ley, al incorporar deriva local que sin ella. Así como también, la capacidad de reproducir el comportamiento de la deriva en los distintos bancos.

Abstract

During the last years, algorithms based on multi-points statistics have been increased their development and application. The present thesis considers the external drift incorporation, associated to local mean, to the algorithm Filter-Based Simulation (Filtersim). In order to do it, the original Filtersim was programmed/reconstructed in Python 2.7.5. Then, an internal modification allows the internal drift incorporation. Specifically, a new weighting, w_0 , measures the influence level that will have during the pattern selection across the simulation process.

Two study cases are conducted. The first seeks to reproduce the sinusoidal behavior shown by high values of a continuous variable in a sedimentary rock image. From it its structural behavior is inferred, while the hard data and external drift is extracted from another known place with similar features called Ground Truth. Then, the simulation uses different drift information degrees and different number of hard data. As results, an optimal influence fluctuates between 40 % and 60 %, managing to adequately characterize the sinusoidal behavior and spatial continuity. For the other hand, results varying the among of hard data shown that considering between an 0.5 % and 1.0 % informed domains bring accuracy local mean drift and enough conditioning information, using Filtersim. In addition, the ability to reproduce complex structure carried out by multi-points simulation is reaffirmed, in contrast to the inability of assessment by traditional variogram-based methods. The second study case is applied in Compañía Minera Cerro Colorado ore deposit, specifically in the South-West sector between the bench 2395 and 2355. The main goal is reproduce the total cooper from the bench 2385 to 2355 inferring the mineralization behavior (patterns) from the bench 2395 using Filtersim. The simulations are conditioned with exploration drilling data and contrasted with blasting holes data assuming their reality representation. A 40 % of influence on the local mean is used. The results show a good reality approximation when conventional Filtersim is used, and incorporation drift decreases the variability between obtained realizations, improving the grade reproduction. A clear and significant benefit, when incorporating drift in local mean, is not well defined. It's due, possibly, to the geological process of genesis associated to the reservoir is rather massive than local, implying a lack of geological structures with well-defined shape and continuity. Specifically, the incorporation of drift get better resources estimation, through grade-tonnage curves. As well as, the ability to reproduce the behavior of the drift in different bench.

*Love the life you live,
Live the life you love*

Agradecimientos

Agradezco al profesor Julián Ortiz por haber propuesto el tema de investigación y el apoyo académico, financiero y personal durante el desarrollo de la tesis. A los profesores y profesionales de comisión Xavier Emery, Brian Townley y Paula Larrondo por el tiempo dedicado en leer la tesis para corregir y enriquecer su contenido.

Agradezco a mis padres Muriel S. y Rodrigo Á. por su entrega incondicional a nuestra familia, por inculcarnos que la educación es uno de los pilares de la vida y mostrarnos con ejemplos que el esfuerzo es el motor de la creación. A mis hermanos Nicolás y Esteban por el apoyo entregado en las etapas finales de pregrado y postgrado, deseándoles el mejor de los éxitos en sus respectivas carreras.

Agradezco a mi polola Carolina A. por el apoyo, paciencia y comprensión durante toda esta etapa. A los mineros de siempre Yavar, Martínez, Peps, Sebita, Juanito, Kiwi y Toño por creer en este proceso y todos los buenos momentos, deseándoles el mejor de los éxitos. A mis amigas incondicionales Carito, Ceci, Kena y Nicole por los momentos de relax y todas esas conversaciones existencialistas para cambiar el mundo, que al fin de cuentas enriquecen el alma. A mis amigos Emilio, Vale, Huaso y Camila por haber crecido juntos en la Universidad.

Agradezco a la Universidad de Chile, como institución, por apoyar el desarrollo de la ciencia e ingeniería. Es inspirador ser parte de una Universidad que busca el crecimiento social por sobre el personal. En especial, al departamento de Ingeniería de Minas.

Agradezco a la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) por otorgarme la Beca de Magister Nacional 2013.

Tabla de contenido

Resumen	I
Abstract	II
1. INTRODUCCIÓN	1
1.1. Motivación	1
1.2. Objetivos	3
1.2.1. Objetivo general	3
1.2.2. Objetivos específicos	3
1.3. Alcances	3
2. ESTADO DEL ARTE	4
2.1. Modelamiento de variables en el espacio	4
2.1.1. Derivas	6
2.2. Modelamiento geológico	7
2.3. Estimación mediante Kriging	8
2.3.1. Linealidad	8
2.3.2. Insesgo	8
2.3.3. Optimalidad	9
2.3.4. Vecindad del Kriging	9
2.3.5. Kriging simple	9
2.3.6. Kriging ordinario	10
2.4. Simulación	10
2.4.1. Simulación de variables categóricas	11
2.4.2. Simulación de variables continuas	11
2.4.3. Simulaciones no paramétricas	12
2.5. Simulación estocástica multipuntos	13
2.5.1. Algoritmo Snesim	13
2.5.2. Algoritmo Filtersim	18
2.5.3. Algoritmo Direct Sampling	30
2.5.4. Algoritmo de síntesis textural	32
2.5.5. Algoritmo DisPat	34
2.5.6. Algoritmo Impala	35
2.5.7. Algoritmo Simpat	37
2.5.8. Comentarios sobre el estado del arte y elección del algoritmo a utilizar - Filtersim	38
2.5.9. Aplicaciones de los algoritmos MPS	40
3. METODOLOGÍA	42
4. FILTERSIM INCORPORANDO DERIVAS	44
4.1. Formas de integrar deriva	44
4.1.1. Modificación función distancia evento condicionante \leftrightarrow prototipo	44
4.1.2. Nuevo filtro y modificación del proceso de clustering	45
4.1.3. Nuevo filtro e imposición de una deriva externa en la media local	45

5. APLICACIONES Y ANALISIS	46
5.1. Primer Caso de estudio – Corte roca sedimentaria	46
5.1.1. Presentación del caso de estudio	46
5.1.2. Metodología	49
5.1.3. Parámetros de simulación	52
5.1.4. Resultados - Visualización, estadísticas básicas y variografía	53
5.1.5. Validación y análisis de resultados	58
5.2. Segundo Caso de estudio - Yacimiento Cerro Colorado	79
5.2.1. Introducción y objetivos	79
5.2.2. Descripción Yacimiento Cerro Colorado	79
5.2.3. Metodología	81
5.2.4. Estudio exploratorio de datos	84
5.2.5. Simulación	90
5.2.6. Resultados y validación	101
5.2.7. Análisis de resultados	109
6. DISCUSIÓN Y CONCLUSIONES	117
6.1. Estado del arte e incorporación de derivas externas	117
6.2. Resultados	117
6.3. Conclusiones generales	120
6.4. Recomendaciones y trabajo futuro	121
BIBLIOGRAFÍA	122
7. ANEXOS	124
7.1. Anexo A	124
7.2. Anexo B	129
7.3. Anexo C	146
7.4. Anexo D	152
7.5. Anexo E	155
7.6. Anexo F	171
7.7. Anexo G	173
7.8. Anexo H	177
7.9. Anexo I	181
7.10. Anexo J	185
7.11. Anexo K	190

Índice de figuras

1.1. Tres Imágenes de Entrenamiento con continuidad espacial distinta. (Referencia: [5])	2
1.2. Variogramas Norte - Sur y Este - Oeste para los conjuntos estructurales de la figura 1.1. (Referencia: [5])	2
1.3. Secuencia estándar de una simulación basada en estadísticas multipuntos.	2
2.1. Cuatro diversas derivas sobre una misma variable regionalizada.	7
2.2. Ilustración de un evento condicionante con 3 posibles estados y 4 datos condicionantes	14
2.3. Ejemplo árbol de condicionamiento.	15
2.4. Ejemplo de función <i>cpdf</i>	16
2.5. Ejemplo, determinación del evento condicionante. (a) Vecindad móvil, (b) grilla de simulación, (c) evento condicionante	17
2.6. 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	17
2.7. 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	17
2.8. Imágenes de entrenamiento. (a) T.I. de una variable binaria – porosidad media. (b) T.I. de una variable categórica – Depósito fluvial con cuatro categorías. (c) T.I. de una variable continua – Distribución de la porosidad. Fuente: [27]	20
2.9. Obtención de patrones a partir de una grilla de búsqueda desde una T.I.	20
2.10. 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	21
2.11. Filtros asociados al gradiente. (d) Norte-Sur, (e) Este-Oeste, (f) Elevación. La escala representa los pesos asociados a cada nodo en la grilla	21
2.12. 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	22
2.13. Caracterización de un patrón, asignación de puntaje, al aplicar un filtro en específico	22
2.14. Ilustración del concepto de imagen erodada	23
2.15. 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	23
2.16. Ejemplo de una clase. (a) Prototipo que identifica a la clase, (b) patrones similares dentro de una misma clase	24
2.17. 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	26
2.18. Ilustración de 3 multigrillas de tamaño 5x5 sobre una imagen de entrenamiento	30
2.19. (a) Vecindad causal, (b) vecindad no causal, (c) síntesis textural al utilizar (a), (d) síntesis textural al utilizar (b). [25]	32
2.20. Reconstrucción textural. (a) Extrapolación de una imagen, (b) interpolación de una imagen [25]	33

2.21. Ejemplo de aplicar una grilla de 8x8 sobre una T.I. (izquierda) y aplicar luego un MDS de 2 dimensiones (derecha). Fuente: [14]	34
2.22. Demanda de memoria RAM. Imagen de entrenamiento en 3 dimensiones con 4 categorías y 600.000 nodos [21]	35
2.23. (izq.) T.I. donde los blancos son 0 y negros son 1. (der.) Grilla de búsqueda de 4 nodos	36
2.24. (a) Grilla primitiva y (b) Grilla dual. Fuente: [1]	37
2.25. Implementación de la grilla dual sobre una variable categórica. Fuente: [1]	38
4.1. Filtro media local - Ejemplo en grilla de búsqueda de 15 x 15	45
5.1. Imagen de entrenamiento (izq.) y Ground Truth (der.) – Roca sedimentaria	47
5.2. (a) Histograma de la TI, (b) Variograma experimental de la TI, (c) Histograma del GT, (d) Variograma experimental del GT	48
5.3. Gráfica del valor medio a lo largo del eje Y para cada valor de X en: (arriba) Imagen de entrenamiento y (abajo) Ground truth	48
5.4. Gráfica del valor medio a lo largo del eje X para cada valor de Y en: (arriba) Imagen de entrenamiento y (abajo) Ground truth	49
5.5. Esquema de trabajo - Caso de estudio – Roca sedimentaria	49
5.6. Datos condicionantes. (a) 100 % Muy pocos D.C. (b) 50 % Muy pocos D.C. (c) 100 % Pocos D.C. (d) 50 % Pocos D.C.	50
5.7. Medias inferidas. (a) Muy Pocos D.C. (b) Media inferida desde Muy Pocos D.C. (c) Pocos D.C. (d) Media inferida desde Pocos D.C. (e) Ground Truth. (f) Media inferida desde Ground Truth	51
5.8. Una realización (arriba) de cien. Promedio de las 100 realizaciones (abajo). Pocos datos condicionantes – Deriva interpretada	54
5.9. Variografía en dirección N0E y N90E. Pocos datos condicionantes – Deriva interpretada	55
5.10. Una realización (arriba) de cien. Promedio de las 100 realizaciones (abajo). Pocos datos condicionantes – Deriva impuesta	56
5.11. Variografía en dirección N0E y N90E. Pocos datos condicionantes – Deriva impuesta	57
5.12. Comparación de realizaciones y promedios entre deriva impuesta e interpretada en el caso <i>pocos datos condicionantes</i>	59
5.13. Derivas. (a) 15 DC y sin derivas externas - (b) 200 DC y sin derivas externas	61
5.14. Derivas. (a) 15 DC - (b) 200 DC. Deriva interpretada con $w_0 : 0.8$	62
5.15. Histograma del ground truth. En rojo los valores sobre umbral 0.63	63
5.16. Categorización del Ground Truth sobre umbral 0.63	63
5.17. Mapas de probabilidad - 15 DC - Deriva interpretada	64
5.18. Mapas de probabilidad - 15 DC - Deriva impuesta	64
5.19. Mapas de probabilidad - 200 DC - Deriva interpretada	64
5.20. Mapas de probabilidad - 200 DC - Deriva impuesta	64
5.21. Gráfico del porcentaje de acierto utilizando <i>muy pocos y pocos datos condicionantes</i> al 100 %	67
5.22. Error Cuadrático Medio Global - Sin umbral	68
5.23. Varianza del Error Cuadrático Medio – Sin umbral	69
5.24. Error Cuadrático Medio Global - Sobre umbral 0.63	71
5.25. Varianza del Error Cuadrático Medio – Sobre umbral 0.63	72
5.26. Variogramas experimentales y modelados para <i>muy pocos y pocos datos condicionantes</i>	73
5.27. Kriging ordinario. (Izq) 15 datos condicionantes – (der) 200 datos condicionantes	74
5.28. Kriging simple. (Izq) 15 datos condicionantes – (der) 200 datos condicionantes	74
5.29. Histogramas resultados Kriging Simple y Ordinario para 15 y 200 datos condicionantes	75
5.30. Bandas Rotantes. Una realización. (Izq) 15 D.C. – (der) 200 D.C.	76

5.31. Bandas Rotantes. Promedio de 100 realizaciones. (Izq) 15 D.C. – (der) 200 D.C. .	76
5.32. Bandas Rotantes. Histogramas: (a) una realización con 15DC, (b) promedio de 100 realizaciones con 15DC, (c) una realización con 200DC, y (d) promedio de 100 realizaciones con 200DC	77
5.33. Ubicación del yacimiento Cerro Colorado	79
5.34. Caracterización de las principales zonas litológicas de Cerro Colorado	80
5.35. Perfil transversal alteraciones, ley de cobre y litología - Cerro Colorado	80
5.36. Migración de datos a una grilla regular	81
5.37. Ilustración análisis variográfico para validación de continuidad espacial	82
5.38. Histogramas y estadísticas básicas para cada propiedad en las muestra de sondajes	84
5.39. Gráfico de dispersión CuT vs CuS en los sondajes de exploración	85
5.40. Sondajes compositados a 10 m - Vista en planta. (Izq. superior) UGCuT 2, (Der. superior) UGCuT 3, (Izq. inferior) UGCuT 4, y (Der. inferior) UGCuT 5	86
5.41. Gráfico de dispersión CuT vs CuS en los pozos de tronadura	87
5.42. Histogramas y estadísticas básicas para cada propiedad en los pozos de tronadura	88
5.43. Pozos de tronadura - Vista en planta. (Izq. superior) UGCuT 2, (Der. superior) UGCuT 3, (Izq. inferior) UGCuT 4, y (Der. inferior) UGCuT 5	89
5.44. Histograma de valores en Z - Pozos de tronadura UGCuT 3	90
5.45. Vista en planta Banco 2395 - Pozos de tronadura UGCuT 3	91
5.46. Vista en planta Banco 2395 - Sector 1 - Grilla T.I.	91
5.47. Imagen de entrenamiento obtenida del Sector 1 - Banco 2395	92
5.48. Gráfico Cuantil-Cuantil entre el Banco 2395 - UGCuT y la imagen de entrenamiento	92
5.49. Vista en planta Banco 2385. (Izq) Sondajes - (Der) Pozos de tronadura	93
5.50. Datos condicionantes y pozos de tronadura antes y despues de migrar a grilla de 10x10x10 - Banco 2385	93
5.51. Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2375	94
5.52. Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2365	94
5.53. Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2355	95
5.54. Deriva en la media en el Banco 2395 para las coordenadas X e Y	96
5.55. Derivas externas inferidas desde los sondajes. Bancos: (a) 2385, (b) 2375, (c) 2365 y (d) 2355.	97
5.56. Histograma de las leyes de cobre en la imagen de entrenamiento (izquierda) y de las leyes luego de aplicar la anamorfosis (derecha)	97
5.57. Variogramas experimentales en los bancos: 2385, 2375, 2365 y 2355	98
5.58. Caso base (SGSim) - Una de cien realizaciones y los promedios para los 4 bancos en estudio	101
5.59. Caso F1 - Una de cien realizaciones y los promedios para los 4 bancos en estudio	102
5.60. Caso F2 - Una de cien realizaciones y los promedios para los 4 bancos en estudio	103
5.61. Promedio en los sesgos, banco a banco, para la media y la varianza respecto a los sondajes y pozos de tronadura	104
5.62. Estudio variográfico - Caso base (SGSim)	106
5.63. Estudio variográfico - Caso F1	107
5.64. Estudio variográfico - Caso F2	107
5.65. Gráfico E.C.M. banco a banco para cada escenario de estudio	109
5.66. Error Cuadrático Medio banco a banco	110
5.67. Derivas en la media - Coordenada X - Banco 2355	112
5.68. Mapas de probabilidad en banco 2365	113
5.69. Categorización sobre banco 2365 - Umbral de 85 % de certeza sobre Ley de corte	114
5.70. Curvas Tonelaje-Ley banco 2375. Utilización del caso base, caso F1 y caso F2 . .	115

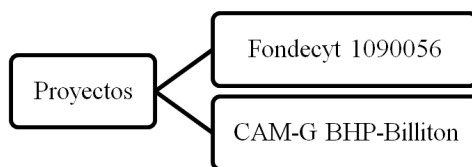
Índice de tablas

2.1. Ventajas y desventajas de los algoritmos de agrupamiento Cross y K-mean [2] . . .	27
5.1. Estadísticas básicas de los datos condicionantes	51
5.2. Parámetros de entrada - Caso de estudio simple	52
5.3. Estadísticas básicas. Pocos datos condicionantes – Deriva interpretada	54
5.4. Estadísticas básicas. Pocos datos condicionantes – Deriva impuesta	56
5.5. Tabla resumen porcentajes de acierto - 100 % datos condicionantes	66
5.6. Parámetros para el cálculo del variograma experimental (15 DC y 200DC)	73
5.7. Sondajes - Estadísticas básicas	85
5.8. Pozos de tronadura - Estadísticas básicas	87
5.9. Grilla asociada a la imagen de entrenamiento	91
5.10. Estadísticas básicas - Datos condicionantes - Bancos 2385, 2375, 2365 y 2355 . . .	95
5.11. Parámetros grillas de simulación	95
5.12. Parámetros para cálculo del variograma experimental	98
5.13. Parámetros Simulación Secuencial Gaussiana	99
5.14. Parámetros de entrada Filtersim sin incluir derivas	99
5.15. Parámetros de entrada Filtersim incluyendo derivas	100
5.16. Media y varianza en las muestras de pozos y sondajes banco a banco	104
5.17. Tabla resumen: promedio en los sesgos para la media y la varianza respecto a los sondajes y pozos	105
5.18. Parámetros para el cálculo del variograma experimental	106
5.19. Errores Cuadráticos Medio por banco para los distintos casos	109
5.20. Parámetros técnico-económicos para el cálculo de la ley de corte	113

Capítulo 1

INTRODUCCIÓN

La presente tesis 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 multipuntos.

- El proyecto Fondecyt 1090056 “*Multiple Point Geostatistics for the Evaluation of Uncertainty in Geological Attributes and Grades*”, el cual culminó (Marzo 2009 a Febrero 2012). No obstante, la presente tesis guarda relación con la temática del proyecto tanto en el contenido como en los recursos (financiamiento).
- El proyecto CAM-G (BHP- Billiton), el cual tiene por objetivo la generación de herramientas de modelamiento geológico semi-automáticas basadas 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

Previo a la realización de una simulación se debe caracterizar la distribución espacial de un atributo relevante para la evaluación de recursos. El variograma es la herramienta que con mayor frecuencia se utiliza para caracterizar correlaciones espaciales, empleada principalmente en la estimación mediante Kriging. El variograma se infiere analizando la correlación entre dos datos separados a una distancia conocida. No obstante, existen variables regionalizadas cuyos comportamientos espaciales no logran ser bien caracterizados utilizando tan sólo dos puntos del espacio. La figura 1.1 [5] presenta un claro ejemplo en el cual 3 conjuntos de estructuras, cuyas proporciones son idénticas, presentan un análisis variográfico similar mientras la continuidad espacial entre ellas difiere notoriamente.

Como se observa en la figura 1.2, el alcance y la meseta para los tres variogramas tanto en dirección Norte-Sur como en Este-Oeste son muy similares a pesar de que sus continuidades espaciales no lo son. Para lograr capturar estos patrones de comportamiento se utilizan las metodologías de simulación basadas en estadísticas multipuntos. Para ello se requiere de una o más imágenes de entrenamiento, **T.I.**, (por sus siglas en inglés, *Training Image*) para proveer la información de los posibles comportamientos estructurales. Estas imágenes de entrenamiento

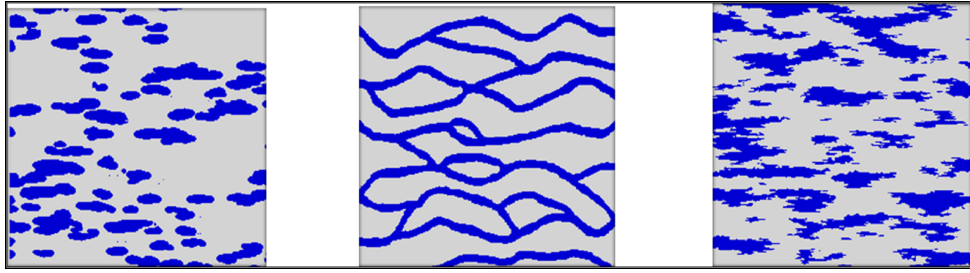


Figura 1.1: Tres Imágenes de Entrenamiento con continuidad espacial distinta. (Referencia: [5])

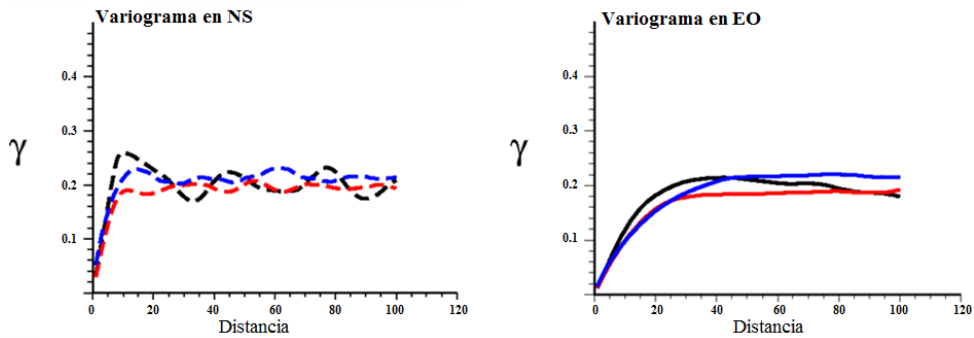


Figura 1.2: Variogramas Norte - Sur y Este - Oeste para los conjuntos estructurales de la figura 1.1. (Referencia: [5])

pueden venir como una interpretación geológica de una zona, los resultados de una exploración geofísica, la información de un banco de tronadura superior, la reconstrucción de una zona ya estudiada similar a una desconocida, etc. Además, y de manera opcional, se puede contar con datos reales medidos en determinados puntos del espacio (por ejemplo sondajes, muestras superficiales, etc.) considerados como datos condicionantes (datos duros). Con ambas informaciones, más la utilización de algún algoritmo (Snesim [22, 23], Filtersim [27], Direct Sampling [18], DisPat [14], Impala [21], Simpat [1]) es posible llevar a cabo una simulación basada en estadísticas multipuntos.

A modo de simplificación, se observa en la figura 1.3 la secuencia lógica de simulación. Se infiere el comportamiento de la variable regionalizada desde la imagen de entrenamiento (T.I.). Se define una grilla a simular. Se opta por una simulación condicional o no condicional y utilizando un algoritmo se generan realizaciones de la variable regionalizada.

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 los datos simulados de manera local. Esto permitiría al usuario, basado en su propia experiencia o datos secundarios, poder modificar la media local en cada una de las realizaciones.

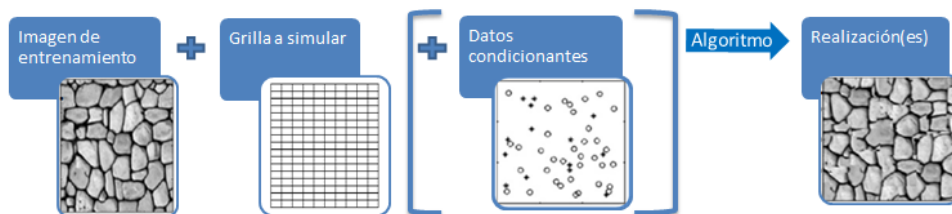


Figura 1.3: Secuencia estándar de una simulación basada en estadísticas multipuntos.

1.2. Objetivos

1.2.1. Objetivo general

Integración de derivas en la media local utilizando el algoritmo de simulación multipuntos Filtersim.

1.2.2. Objetivos específicos

- Presentar el formalismo y la relevancia de generar modelos geológicos.
- Estudiar los algoritmos actuales de simulación multipuntos estableciendo ventajas y desventajas.
- Descripción de la modificación aplicada sobre el algoritmo Filtersim para la incorporación de derivas en la media local.
- Implementación de Filtersim en Python.
- Desarrollar dos casos de estudio. El primero asociado al corte pulido de una roca sedimentaria para cuantificar mejoras o empeoramiento al incorporar las derivas externas. El segundo caso de estudio se realiza sobre bancos de explotación en el yacimiento pórfido-cuprífero Cerro Colorado, propiedad de BHP Billiton. En este se mide el impacto de la técnica en la simulación de leyes de Cobre al utilizar la información de un banco de explotación para simular los bancos inferiores.
- Comparar resultados al utilizar técnicas convencionales de estimación y simulación.

1.3. Alcances

De la presente tesis se establecen como alcances:

- Exponer el estado del arte respecto a la simulación multipuntos.
- Establecer una metodología para incorporar cambios en la media local al utilizar Filtersim generando así una versión modificada del algoritmo.
- Utilizar los parámetros de entrada, o rango de parámetros, propuestos por Ávalos, S. [24] en los casos de estudio: Roca sedimentaria y yacimiento Cerro Colorado.

Capítulo 2

ESTADO DEL ARTE

2.1. Modelamiento de variables en el espacio

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 introducido por primera vez en 1962 por Georges Matheron, describiéndose como la interpretación estadística y probabilística, a través de funciones aleatorias, de fenómenos naturales regionalizados. Una variable regionalizada se extiende en un espacio determinado y presenta cierta continuidad dentro del mismo.

El espacio (campo), D , en el cual se extiende la variable regionalizada por lo general se encuentra parcialmente informado. A partir de los datos conocidos se generan modelos geoestadísticos (representaciones matemáticas) capaces de entregar un posible valor de la variable en puntos del espacio no informados. La posibilidad de generar modelos generales, basados en el modelamiento de procesos, es escasa dada la complejidad de los fenómenos regionalizados y sus procesos de génesis. Para ello se recurre a modelos probabilísticos, los cuales deben asumir ciertas características para dar una base matemática a su construcción.

Sea x un punto en D cuyo valor es conocido. Se asume que este es sólo *una* realización de una variable aleatoria $Z(x)$. El conjunto de variables aleatorias $\{Z(x), x \in D\}$ se denomina función aleatoria Z . Luego una realización de esta función aleatoria Z describiría por completo el espacio D .

La función aleatoria se puede caracterizar dada su distribución espacial resumible en momentos bajo ciertas hipótesis (estacionaridad). Para mayor detalles ver [7].

- Distribución espacial:

Sea una función aleatoria $Z = \{Z(x), x \in D\}$ y un conjunto de k -puntos x_1, \dots, x_k en el campo de extensión D . Se le denomina función de distribución multivariable bajo k -argumentos a:

$$F_{x_1, \dots, x_k} = Prob\{Z(x_1) < z_1, \dots, Z(x_k) < z_k\} \forall z_1, \dots, z_k \quad (2.1)$$

Finalmente, al conjunto total de funciones de distribución para todos los posibles conjuntos entre 1 y k de $x_1, \dots, x_k \in D$ se le denomina distribución espacial [7].

Usual suele ser considerar solo las distribuciones espaciales univariable ($k = 1$) y bivariante ($k = 2$) ya que generalmente sólo se tiene información suficiente para inferir a ese nivel.

- Momentos de primer y segundo orden:

Se toman en consideración las distribuciones univariante y bivariante las cuales pueden, de existir, ser resumidas por:

- Esperanza:

$$E[Z(x)] = m(x) \quad (2.2)$$

Es el valor promedio en un punto dado $x \in D$ alrededor del cual fluctúan los valores de cada realización de la función aleatoria Z .

- Varianza:

$$\sigma^2(x) = var[Z(x)] = E[Z(x)^2] - m(x)^2 \quad (2.3)$$

Es el valor que cuantifica la dispersión que existe entre los valores entregados por las realizaciones de $Z(x)$ entorno al valor esperado $m(x)$.

- Covarianza:

$$C(x_1, x_2) = cov[Z(x_1), Z(x_2)] = E[Z(x_1)Z(x_2)] - m(x_1)m(x_2) \quad (2.4)$$

Cuantifica la relación existente entre los valores de la variable regionalizada en dos puntos x_1 y x_2 .

- Variogramma (semi-variograma):

$$\gamma(x_1, x_2) = \frac{1}{2}var[Z(x_1) - Z(x_2)] \quad (2.5)$$

Es una herramienta para medir la continuidad espacial de la variable. Cuantifica cuán rápido se desestructura la variable regionalizada en el espacio.

- Correlograma:

$$\rho(x_1, x_2) = corr[Z(x_1), Z(x_2)] = \frac{cov[Z(x_1), Z(x_2)]}{\sqrt{var[Z(x_1)]var[Z(x_2)]}} \quad (2.6)$$

El correlograma es un indicador adimensional que toma valores entre $[-1, 1]$ y da cuenta de la dependencia lineal entre dos variables. Los valores extremos -1 y 1 indican una marcada proporcionalidad entre las variables mientras que el valor 0 reconoce la ausencia de correlación lineal.

- Estacionaridad:

Para dar un formalismo probabilístico de la función aleatoria Z se debe introducir el concepto de estacionaridad. Tal como se describió anteriormente los sitios con datos representan tan sólo una realización y son un conjunto finito usualmente menor al dominio de la variable regionalizada. Esto último impide una inferencia global de la función de distribución. Como respuesta a ello el concepto de estacionaridad permite asumir que las características asociadas a un punto son idénticas independiente del punto escogido.

Esto último implica que la distribución espacial de la función aleatoria es invariante por traslación (2.7) y que para un conjunto de datos las propiedades solo guardan relación con la distancia relativa entre ellos y no su posición absoluta. Esta propiedad se conoce como **estacionaridad estricta**.

$$F_{(x_1, \dots, x_k)}(z_1, \dots, z_k) = F_{(x_1+h, \dots, x_k+h)}(z_1, \dots, z_k) \forall h, \forall k \in \mathbb{N}, x_1, \dots, x_k \in \mathbb{R} \quad (2.7)$$

Esto lleva a que la distribución univariable no dependa de la posición del dato considerado y que la distribución bivariable sólo dependa de la distancia entre el par de datos considerados. Los momentos de primer y segundo orden quedan definidos como:

- $F(z_1) = Prob\{Z(x_1) < z_1\}$ & $F_h(z_1, z_2) = Prob\{Z(x+h) < z_1, Z(x) < z_2\}, \forall x \in D$
- $E[Z(x)] = m, \forall x \in D$
- $var[Z(x)] = \sigma^2, \forall x \in D$
- $C(h) = cov[Z(x+h), Z(x)] = E\{Z(x+h)Z(x)\} - m^2, \forall x \in D$

- $\rho(h) = \text{corr}[Z(x+h), Z(x)] = \frac{\text{cov}[Z(x+h), Z(x)]}{\sqrt{\text{var}[Z(x+h)]\text{var}[Z(x)]}}, \forall x \in D$
- $\gamma(h) = \frac{1}{2}E\{[Z(x+h) - Z(x)]^2\}, \forall x \in D$

Verificar la estacionaridad estricta suele ser una tarea bastante ardua. Como solución se plantea verificar las restricciones $E[Z(x)] = m$ y $C(h) = \text{cov}[Z(x+h), Z(x)], \forall x \in D$ que dan cuenta de una **estacionaridad de segundo orden**.

Por otro lado, existe un tipo de estacionaridad denominada **hipótesis de estacionaridad intrínseca**. Consiste en asumir estacionaridad de segundo orden no sobre la variable aleatoria $Z(x)$ sino que sobre su diferencia de primer orden $Z(x+h) - Z(x)$ [19].

Para verificar que la variable $Z(x)$ cumple la hipótesis de estacionaridad intrínseca se debe verificar que:

- $E[Z(x+h) - Z(x)] \forall x \forall h$
- $2\gamma(h) = \text{var}[Z(x+h) - Z(x)] = E\{[Z(x+h) - Z(x)]^2\}$

2.1.1. Derivas

La hipótesis de estacionaridad da sustento matemático a la aplicación de distintas técnicas de estimación. Sin embargo, sobre los fenómenos naturales a modelar es común hallar ciertas tendencias, o cambios sistemáticos, en la variable regionalizada a lo largo del tiempo o espacio. Este cambio sistemático en el valor esperado de la función aleatoria es conocido como deriva. Formalmente, se puede separar la variable regionalizada $Z(x)$ en sus componentes residual $R(x)$ y deriva $m(x)$ (2.8). Dado que la deriva corresponde al valor esperado de $Z(x)$ se plantea entonces que $E\{Z(x)\} = m(x) \forall x$ mientras que para el valor residual el valor esperado a lo largo del dominio es cero, $E\{R(x)\} = 0$

$$Z(x) = R(x) + m(x) \tag{2.8}$$

La deriva puede tener distintas características, siendo unas más sencillas de modelar que otras. En la figura 2.1.a [19] se observa la medición de SO_2 en un periodo de 18 meses. Este gráfico presenta una deriva en el valor esperado siendo distinto dependiendo el mes del año. Si se aplica el criterio de estacionaridad a la diferencia de primer orden del tipo $Z(x+1) - Z(x)$ donde el vector “1” representa una unidad temporal (día) se obtiene el gráfico 2.1.b, que representa los incrementos día a día. Queda así descubierto que el valor esperado de la diferencia antes planteada es idéntico a lo largo del año. Sin embargo, una nueva deriva se presenta en la variación de esta diferencia. Para enfrentar esta situación se plantea aplicar una transformación del tipo logarítmica a los datos originales, $\log[Z(x)]$, obteniendo así el gráfico 2.1.c. Finalmente, al tomar nuevamente los incrementos día a día de la nueva variable transformada se obtiene el gráfico 2.1.d donde se observa que tanto el valor esperado como la varianza de los incrementos logarítmicos es relativamente constante a lo largo del año.

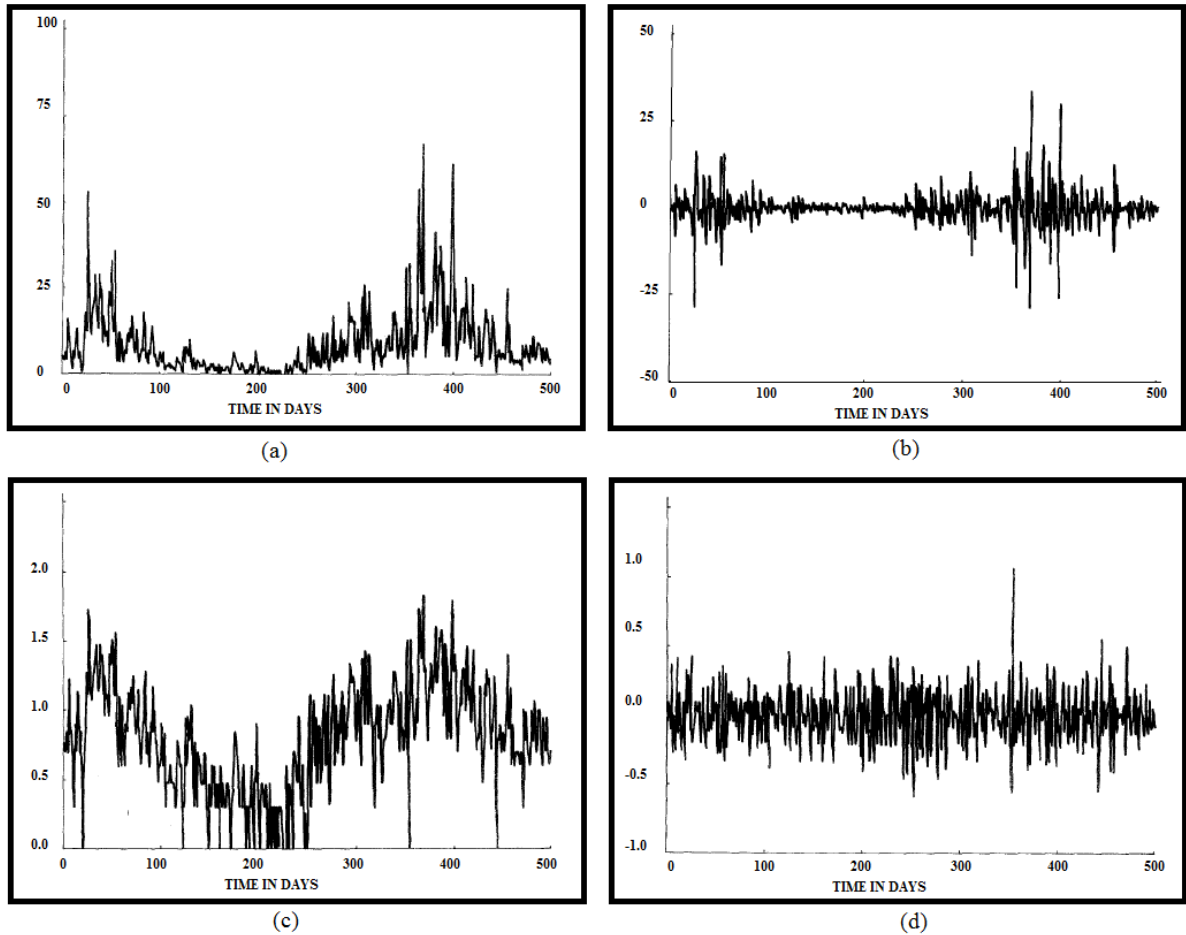


Figura 2.1: Cuatro diversas derivas sobre una misma variable regionalizada.

El ejemplo anterior presenta una manera de tratar las derivas para llevar la variable regionalizada no estacionaria a una nueva variable estacionaria bajo condiciones. Otra manera de analizar las derivas es utilizar la ecuación (2.8) de descomposición, suponer estacionaridad en la componente residual para luego estimarla sobre todo el dominio y finalmente agregar la componente de deriva. Por otro lado, una alternativa es generar diversas unidades de estimación en donde la variable se comporta relativamente estacionaria y asumir alguno de los criterios de estacionaridad.

Finalmente el proceso de inferir, interpretar y modelar una variable regionalizada en presencia de deriva es un tema delicado con diversas formas de enfrentarlo. En el caso particular de la presente tesis se plantea la incorporación y el tratado de derivas externas en la media local sobre la variable regionalizada.

2.2. Modelamiento geológico

El modelamiento geológico busca construir modelos interpretados de la geología, de manera simplificada, a partir de un conjunto limitado de información. La dimensión de los modelos va desde un modelo global a gran escala hasta sectores específicos del dominio a pequeña escala. La extensión de estos guarda estrecha relación con el tipo de información necesaria. Sin embargo, es posible integrar información de diversas escalas para mejorar la reproducción (respecto a la realidad) de los modelos propuestos.

Un modelo geológico se caracteriza por estar definido sobre un dominio específico delimitado en las tres direcciones del espacio y en un instante de tiempo en particular. Dependiendo del

interés por el cual un modelo es creado se define entonces el tamaño de soporte al cual se entrega y se interpreta la información. Es habitual que el soporte de ellos sea en cubos o paralelepípedos de entre 3 metros a 25 metros por lado (para el caso de la minería, el tamaño del soporte de bloques guarda estrecha relación con el tamaño de la Unidad Básica Minera¹). Cada bloque es representado luego por su centroide a través de las coordenadas Este (X), Norte (Y) y Elevación (Z). Dentro de las variables que con mayor frecuencia se integran a los modelos se encuentran la mineralización de los elementos de interés económico, mineralización de los elementos contaminantes y/o nocivos en los procesos posteriores, la densidad de la roca, la litología, el tipo de alteración, etc.

Los modelos se sustentan en diversas fuentes de información. La calidad, cantidad y origen de la información guarda estrecha relación con la etapa del proyecto. En las etapas tempranas se busca detectar ciertas anomalías geológicas y por ello la información tiende a tener un carácter más masivo, es decir, se cuenta con información geofísica, registros de ondas sísmicas, mallas amplias de sondajes, muestreos superficiales, entre otras. A medida que los modelos buscan ser más detallados se procede a generar mallas de sondajes más densas, información proveniente de galerías de exploración, análisis químico sobre soporte más pequeño en cada pozo, análisis de resistencia y/o competencia de la roca in-situ (UCS, ensayos de corte, etc.), entre otras. Finalmente, la creación de los modelos se basa en las técnicas de estimación y/o simulación para predecir, bajo ciertos niveles de incertidumbre, la ocurrencia de los fenómenos geológicos en los sectores del modelo donde no se cuenta con información directa. Lo anterior se sustenta en una o más interpretaciones determinísticas que se basan en el conocimiento geológico del fenómeno modelado.

2.3. Estimación mediante Kriging

Basados en las hipótesis y formalismos que definen la distribución espacial de la variable regionalizada se genera un estimador Z^* capaz de representar esta variable en todo su dominio. Para ello el estimador debe cumplir con las restricciones de linealidad, insesgo y optimalidad.

2.3.1. Linealidad

Se define al estimador Z^* como una combinación lineal ponderada de todos los datos existentes dentro de una vecindad en torno al punto a estimar. Sea x_0 el punto a estimar; $\{x_\alpha, \alpha = 1, \dots, n\}$ el conjunto de datos dentro de una vecindad centrada en x_0 ; $\{\lambda_\alpha(x_0), \alpha = 1, \dots, n\}$ los ponderadores denominados ponderadores del kriging; y a un coeficiente. El estimador queda descrito por la ecuación (2.9):

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_\alpha(x_0) Z(x_\alpha) + a \quad (2.9)$$

En donde a y $\lambda_\alpha(x_0)$ son las incógnitas de kriging que se busca resolver.

2.3.2. Insesgo

Por estimador insesgado se tiene cualquier estimador cuya esperanza del error es nula, es decir, sobre el dominio D estimado el promedio de los errores tiende a cero. De manera formal se tiene la ecuación (2.10):

¹UBM: *Unidad Básica Minera. El tamaño y forma es función de la selectividad esperada en el yacimiento que a su vez depende del método de explotación asociado y habitualmente repercute en los costos de producción y flexibilidad en la operación*

$$E[Z^*(x_0) - Z(x_0)] = 0 \quad (2.10)$$

Cabe destacar que esta propiedad sólo establece que los errores fluctúan en torno a cero, sin embargo, no tiene implicancia en la variabilidad que estos errores pueden alcanzar.

2.3.3. Optimalidad

Con la restricción de optimalidad se obliga al estimador a minimizar la varianza del error, es decir, se busca que la dispersión de los errores de estimación al interior del dominio D sea la menor posible. Expresado matemáticamente en la ecuación (2.11):

$$\text{Minimizar } \text{Var}[Z^*(x_0) - Z(x_0)] \quad (2.11)$$

Finalmente se obtiene un estimador lineal, óptimo e insesgado.

2.3.4. Vecindad del Kriging

Por vecindad del kriging se entiende al espacio dentro del dominio que circunda la posición del valor a estimar dentro de la cual se buscan los datos condicionantes. Esta vecindad puede ser *única* o *móvil*.

Vecindad única

Es el caso particular cuando se decide utilizar todo el dominio para estimar los puntos en su interior. Esto puede llevar a procesos de cómputo bastante costosos y resultados que no difieren mayormente al utilizar una vecindad móvil.

Vecindad móvil

Se define como vecindad móvil a un sector definido geoméricamente y centrado en el punto a estimar que sólo toma en consideración los valores cercanos y que están al interior de esta vecindad. Se le denomina móvil ya que al ir variando el punto a estimar varían también los datos considerados en la estimación. La forma y tamaño de la vecindad es decisión del usuario.

2.3.5. Kriging simple

Se habla de kriging simple cuando el valor de la media es conocido, y bajo la hipótesis de estacionaridad, es idéntico en todo el dominio.

$$E[Z(x)] = m, \forall x \in D \quad (2.12)$$

Donde D es el dominio total a estimar. Resolviendo las restricciones de linealidad, insesgo y optimalidad se lleva a que los ponderadores del kriging quedan definidos por el sistema de ecuaciones descrito de forma matricial en la ecuación (2.13).

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ C(x_n - x_1) & \cdots & C(x_n - x_n) \end{pmatrix} \begin{pmatrix} \lambda_1^{KS} \\ \vdots \\ \lambda_n^{KS} \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \end{pmatrix} \quad (2.13)$$

Finalmente, Z^* queda definido como:

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{KS} Z(x_{\alpha}) + \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{KS}\right) m \quad (2.14)$$

2.3.6. Kriging ordinario

Como contraparte es posible utilizar un kriging ordinario, que da un grado más de libertad al sistema de ecuaciones ya que asume la media m como un parámetro desconocido. Al resolver las restricciones de linealidad, insesgo y optimalidad se llega al sistema de ecuaciones expuesto matricialmente en la ecuación (2.15):

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(x_n - x_1) & \cdots & C(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1^{KS} \\ \vdots \\ \lambda_n^{KS} \\ \mu \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \\ 1 \end{pmatrix} \quad (2.15)$$

Finalmente, el estimador queda definido tal como se plantea en la ecuación (2.16):

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{KS} Z(x_{\alpha}) \quad (2.16)$$

Para más detalles, demostraciones y definiciones ver [7].

2.4. Simulación

Los métodos de estimación mediante kriging antes expuestos no logran caracterizar las propiedades de los valores reales de la variable regionalizada. En efecto, el kriging lleva a resultados con suavizamiento, es decir, la dispersión obtenida en los valores estimados es menor a la dispersión de los valores reales. La variabilidad de los valores estimados no es uniforme dentro del campo estimado, es decir, disminuye en vecindades del kriging con poca presencia de datos. El intervalo de los valores estimados es siempre menor que el intervalo real lo que no logra caracterizar valores extremos, valores que en situaciones son el objetivo de investigación.

Como respuesta a lo anterior, se infiere el comportamiento de la variable regionalizada a partir de los datos y se modela su distribución espacial para la generación de la función aleatoria, la cual en cada realización genera una variable regionalizada ficticia similar a la variable real. Lo interesante de las simulaciones es que *una* sola realización puede ser considerada como una posible realidad. En particular, cada realización cumple con:

- Reproduce el histograma a priori de los valores reales.
- Varianza de estimación mayor a la varianza del kriging (ya que no suaviza).
- Reproduce el efecto pepita presente en el variograma de la variable real. Esto no se da en la estimación de kriging debido nuevamente al suavizamiento.
- No es un estimador preciso.

Dado que *una* sola realización no estima la variable regionalizada si no que sólo da cuenta de una posible realidad de esta misma, es que para lograr estimar la variable, o el desempeño de esta frente a un proceso, se debe contar con un número adecuado (suficientemente grande) de realizaciones. Al considerar un gran número de realizaciones se cumple con:

- El promedio de las realizaciones se asemeja a la estimación por kriging simple.
- Para un punto dado, es posible construir una distribución de los valores obtenidos y dar cuenta de cuan precisa es la estimación mediante: intervalos de confianza, umbrales de tolerancia, etc.

- Se pueden procesar y analizar conjuntos de puntos para así obtener modelos de su variabilidad conjunta. En minería se puede mencionar el análisis de un sector sobre una ley de corte o bien el cambio de soporte, entre otros ejemplos.

Las simulaciones (en su mayoría) independiente de la técnica empleada pueden ser llevadas a cabo de forma condicional o no condicional. Se entiende por una *simulación no condicional* cuando se busca repetir la distribución espacial de la función aleatoria sobre un dominio en particular sin contar con la restricción de valores específicos sobre sitios determinados. Por su parte, sea $x_\alpha, \alpha = 1, \dots, n$ un conjunto de sitios informados, se denomina *simulación condicional* al modelo matemático capaz de restituir el valor real al momento de simularlo. Cabe notar que no vale tan sólo con reemplazar el valor real en el sitio estimado ya que esto generaría artefactos en el resultado. Para ello debe ser considerada toda la vecindad y respetada la distribución espacial en torno al dato condicionante.

La simulación puede ser llevada a cabo tanto en variables continuas como categóricas. Ciertas técnicas deben llevar variables continuas a un conjunto finito de variables categóricas dividiendo el universo en intervalos, mientras que otras pueden ser aplicadas en ambos casos sin trabajo previo sobre los datos.

2.4.1. Simulación de variables categóricas

La caracterización de la función aleatoria, en particular su función de distribución, en el caso de variables categóricas se plantea como una función escalonada que caracteriza la probabilidad de ocurrencia de una categoría en relación al resto. Es posible que exista una dependencia entre una categoría y el resto (orden jerárquico) lo que obliga a caracterizar esta distribución con más de una función de distribución (usualmente, una para cada variable condicionada por orden jerárquico).

Citando alguna de las técnicas existentes se tiene (para más detalle ver [10]):

- Simulación de indicadores
- Modelo Gaussiano truncado
- Modelo plurigaussiano
- Simulación multi-puntos

2.4.2. Simulación de variables continuas

La confección de la función de distribución espacial resulta ciertamente más fácil para variables continuas. Sin embargo, varias técnicas requieren que la función aleatoria cumpla con la hipótesis de multi-gaussianidad. Para cumplir esto es posible aplicar una *anamorfosis* sobre la función aleatoria que la lleva de una distribución particular a una distribución marginal gaussiana estándar. Luego, un método en particular es aplicado sobre la variable transformada y los resultados son destransformados por la función de anamorfosis inversa. La función de anamorfosis puede obtenerse de manera práctica mediante una equivalencia de las funciones de distribución acumulada entre la variable inicial y una gaussiana estándar, o bajo un modelamiento matemático al utilizar una serie de polinomios de Hermite [10].

Citando alguna de las técnicas existentes se tiene (para más detalle ver [10]):

- Descomposición matricial
- Método espectral continuo

- Método secuencial Gaussiano
- Método de bandas rotantes

2.4.3. Simulaciones no paramétricas

Hasta ahora se ha comentado que las simulaciones requieren generar por completo la función aleatoria caracterizando su función de distribución a través de un set de parámetros dependiendo del modelo a utilizar. Sobre ciertas variables y situaciones resulta difícil esta caracterización dado que los modelos no logran ajustarse a la variable en estudio. Como respuesta a ello se puede caracterizar un sector parcial de la distribución espacial o bien inferir el comportamiento directamente de una imagen de referencia cuyas variables internas caracterizan a la variable en estudio. De ello deriva el nombre de *simulaciones no paramétricas* dado que no requiere especificar parámetros internos a los modelos.

Dentro de la simulación no paramétrica dos son las técnicas más utilizadas (solo serán citadas con una breve descripción, para más detalle ver [9]):

- Simulación secuencial de indicadores.
Consiste en aplicar un kriging de indicadores, sobre umbrales definidos, con la finalidad de determinar la distribución condicional de la variable regionalizada y luego simular un valor a partir de ella que se usa para condicionar los puntos siguientes.
- Simulación estocástica multipuntos.
Considerando un patrón de datos finitos, k , posiblemente condicionado a un set de p datos ($p < k$), se infiere la distribución de probabilidad condicional a partir de una imagen de referencia.

Dado que el tema de la tesis se centra en la utilización de un algoritmo específico dentro de la simulación estocástica multipuntos se llevará a cabo una descripción más detallada de este tipo de simulaciones y los actuales algoritmos que se han generado en su desarrollo.

2.5. Simulación estocástica multipuntos

Guardiano y Srivastava [12], introducen por primera vez el concepto de la estadística multipuntos. En ella es posible incorporar la relación espacial que existe entre 3 o más puntos. Para ello la función aleatoria Z es generada a partir de una previa interpretación de la realidad: interpretación geológica determinística, imágenes de entrenamiento (T.I.), bases de datos de zonas análogas, etc. y no de un modelo matemático teórico. Guardiano y Srivastava proponen la simulación de valores a partir de la generación de la función aleatoria Z como inferencia probabilística al recorrer una T.I. 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 T.I. nuevamente. Strebelle ([22], [23]) 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 T.I. disminuyendo el tiempo del proceso. El problema surge cuando la T.I. 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. [27] desarrollan una nueva técnica, denominada Filtersim (*filter-based simulation*) para simplificar el almacenamiento de los patrones estructurales a partir de imágenes de entrenamiento. La técnica ocupada es el uso de filtros que recorren la T.I. obteniendo diversos patrones de comportamiento y almacenando el valor de la media, la gradiente y la curvatura en cada dirección del espacio (2D o 3D) sobre cada patrón. De esta manera hay un ahorro sustancial en el almacenamiento de información, sin verse mayormente afectado por el tamaño de la T.I. Su aplicación es posible tanto para variables discretas como continuas.

Nuevas técnicas o modificaciones a las existentes han sido propuestas, tales como Direct Sampling [18], DisPat (*distance-based pattern modeling*) [14], Impala (*Improved Parallel Multipoint Algorithm Using a List Approach*) [21], Simpat [1], etc. Estas técnicas logran incorporar más de una variable (co-simulaciones), disminuyen la demanda de memoria o bien la anulan, logran ser paralelizables en procesos internos puntuales aumentando con ello la velocidad de cómputo, entre otras ventajas.

2.5.1. Algoritmo Snesim

Snesim (*Single Normal Equation Simulation*), fue desarrollado por Strebelle [23]. Se caracteriza por ser un algoritmo secuencial, es decir, trabaja sobre una secuencia de nodos predefinida. Tiene la capacidad de simular 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, con $K \in \mathbb{N}$.
- Simple de ser generado bajo algún lenguaje de programación.
- Presente en el software SGeMS ².
- Demandante en recursos: gran cantidad de memoria RAM para almacenar las condiciones previa confección de la función distribución de probabilidad condicionada.
- Lento. Si la imagen de entrenamiento es de gran tamaño entonces el proceso demanda mucho tiempo.

²SGeMS: *Stanford Geostatistical Modeling Software*

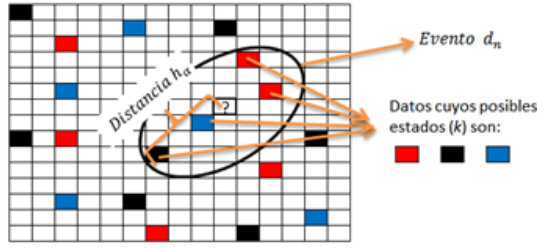


Figura 2.2: Ilustración de un evento condicionante con 3 posibles estados y 4 datos condicionantes

- Dificultad de incorporar la media local, ya que el algoritmo es rígido en estructura.

Terminología

Previo a una explicación detallada de la estructura del algoritmo, es necesario definir varios conceptos y terminologías [23].

- Se denota por S una característica de la variable y por k los posibles estados de esta característica $\{S_k \text{ con } k = 1 \dots K\}$, 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 2.2 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 $S(u_\alpha) = S(u + h_\alpha)$.

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 2.3 presenta un ejemplo de confección de árbol de búsqueda en donde, en 2.3.a, una vecindad de 4 nodos recorre la imagen de entrenamiento para ir luego en 2.3.b almacenando la cantidad de veces que se presenta el valor Blanco (W) o Negro (B) en el centro de la vecindad condicionado a los valores de los cuatro nodos. 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).

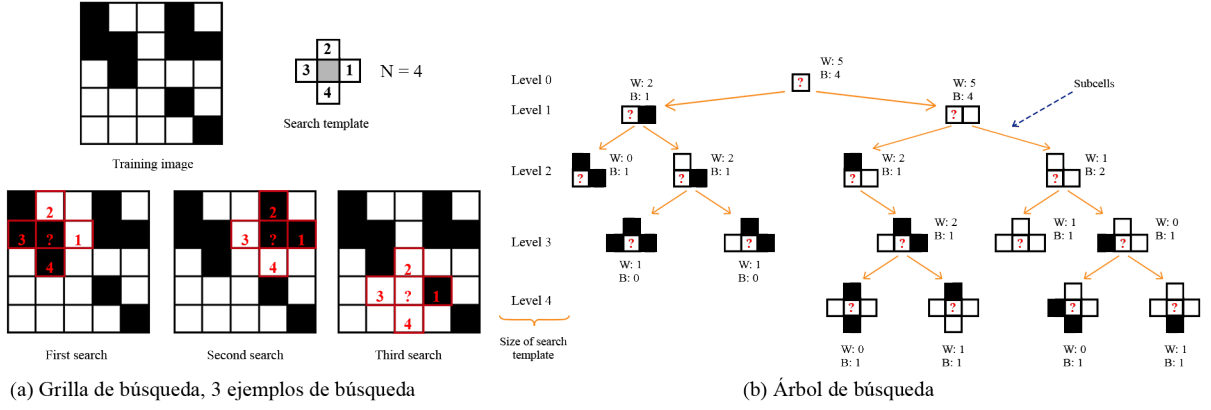


Figura 2.3: Ejemplo árbol de condicionamiento.

Utilizando la terminología antes expuesta se explica a continuación 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. La ecuación (2.17) presenta la analogía entre el concepto probabilístico de la ocurrencia del evento (izquierda) y la cuantificación desde el árbol de búsqueda (derecha).

$$Prob\{S(u) = s_k \mid d_n\} = f(u; k \mid d_n) \quad \text{con } k = 1, 2, \dots, K \quad (2.17)$$

Snesim establece como indicadores binarios A_k y D tal como se aprecia en la ecuación (2.18) y (2.19) respectivamente:

$$A_k = \begin{cases} 1 & \text{si } S(u) = S_k \\ 0 & \text{si no} \end{cases} \quad (2.18)$$

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.

$$D = \begin{cases} 1 & \text{si } S(u_\alpha) = S_{k_\alpha} \forall \alpha = 1 \dots n \\ 0 & \text{si no} \end{cases} \quad (2.19)$$

- c) Se establece que para confeccionar la curva de distribución de probabilidad condicional (*cpdf*, conditional probability distribution function) se tiene que las probabilidades de la ecuación (2.16) pueden ser obtenidas utilizando en la ecuación (2.20).

$$f(u; k \mid d_n) = Probabilidad\{A_k = 1 \mid D = 1\} \quad (2.20)$$

Más aun, esta última equivale a la relación probabilística expuesta en la ecuación (2.21) (Teorema de Bayes).

$$\frac{Probabilidad\{A_k = 1, D = 1\}}{Probabilidad\{D = 1\}} \quad (2.21)$$

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 2.3.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:

$$f(u; k | d_n) = \frac{c_k(d_n)}{c(d_n)} \quad (2.22)$$

- d) Finalmente se obtiene la función *cpdf*, tal como expone el ejemplo de la figura 2.4. A través de esta función se procede a realizar el proceso de simulación asociado al valor desconocido en u .

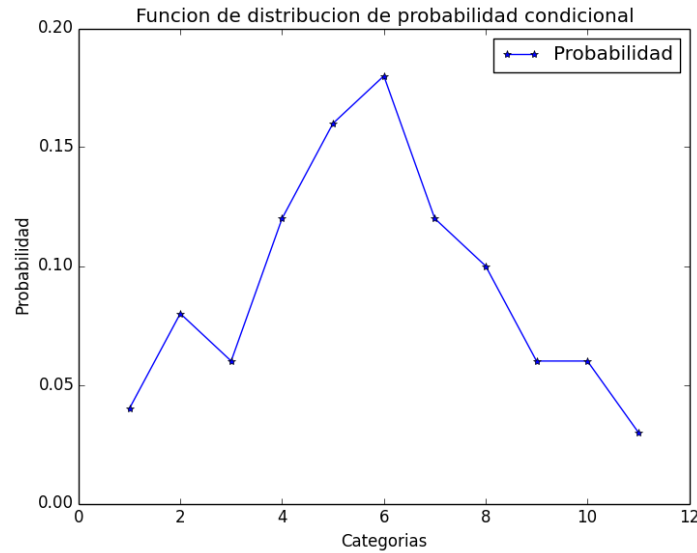


Figura 2.4: Ejemplo de función *cpdf*

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 2.5.b.
- c) Se establece una ruta aleatoria que recorrerá todos los nodos sin información.
- d) Se establece una vecindad móvil, figura 2.5.a, que recorrerá la ruta previamente definida determinando los eventos condicionantes, figura 2.5.c. Si dentro de la vecindad no se encuentran nodos con información condicionantes un valor se escoge de manera aleatoria a partir de la función de distribución a priori.
- 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 2.6.
- f) Se realiza el proceso e) hasta que toda la grilla de simulación queda completamente informada. La figura 2.7 es un ejemplo de realización completo.

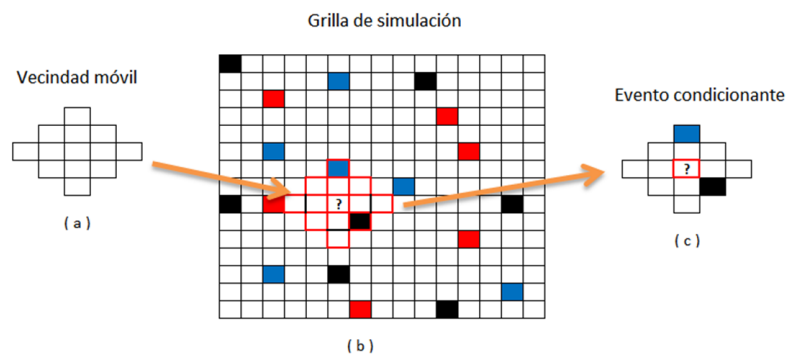


Figura 2.5: Ejemplo, determinación del evento condicionante. (a) Vecindad móvil, (b) grilla de simulación, (c) evento condicionante

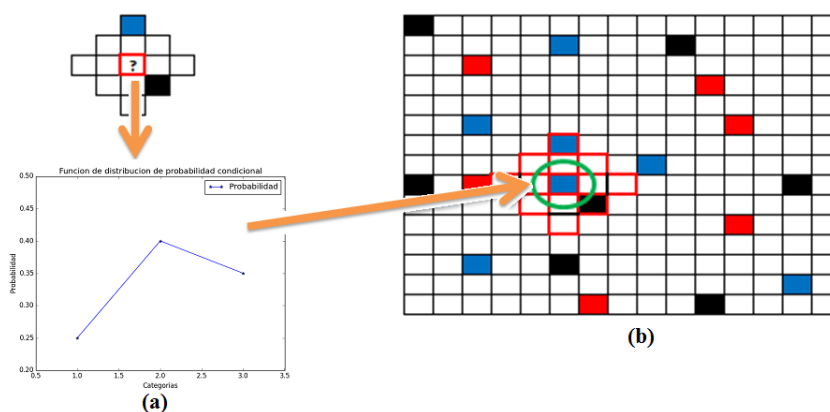


Figura 2.6: Determinación del valor asociado al nodo central. (a) Uso de la cpdf dado el evento condicionante, (b) asignación del valor al nodo central

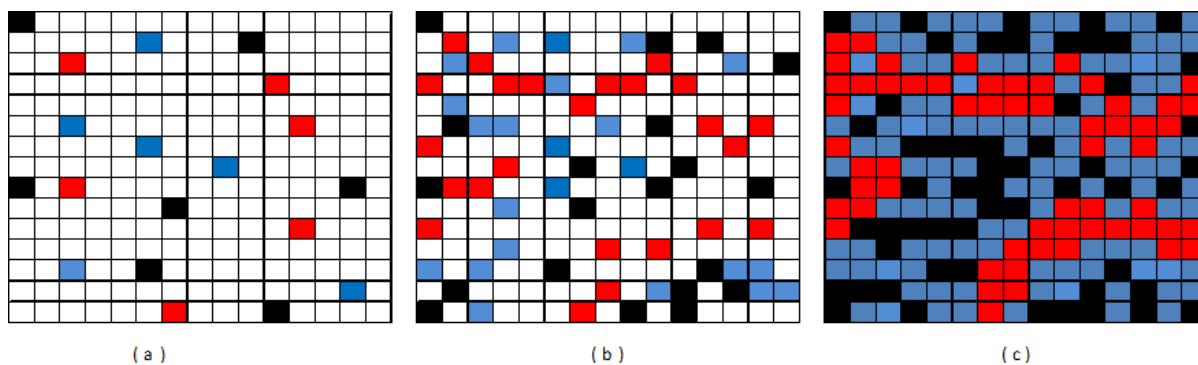


Figura 2.7: Ejemplo de una realización con Snesim (caso ficticio). (a) Grilla a simular con datos condicionantes, (b) mitad del proceso de simulación, (c) grilla de simulación completa

2.5.2. Algoritmo Filtersim

El algoritmo Filter-Based Simulation, Filtersim fue desarrollado por Zhang [27]. Filtersim, como metodología para la simulación basada en estadísticas multipuntos, presenta como características:

- Es aplicable tanto para variables categóricas (litologías, unidades geológicas, etc.) como continuas (leyes de mineralización, porosidad, densidad, concentración, etc.) por lo que permite su utilización en campos como la geología, evaluación de recursos mineros, industria del petróleo, 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 un patrón de búsqueda (grilla) de $10 \times 10 \times 10$ 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 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 determinística, 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 vecindad de búsqueda móvil. Sus dimensiones son definidas por el usuario siendo usualmente implementadas vecindades rectangulares (2 dimensiones) o paralelepípedas (3 dimensiones) sin desmedro de vecindades (grillas) asimétricas.
3. La aplicación de los F -filtros sobre un mismo sector de la T.I. llevan a caracterizar cada patrón local con un vector de F -dimensiones. Ver figura 18.
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 un patrón de entrenamiento específico. Cabe destacar que es posible encontrar puntos de igual valor representando 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 las secciones siguientes. *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.
8. Sobre la grilla a simular se genera en 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. Son los procesos aleatorios internos de escoger el patrón dentro del prototipo (o prototipo hijo) y la definición aleatoria del camino a seguir 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 qué 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.

Obtención de patrones a partir de una imagen de entrenamiento

Por imagen de entrenamiento (T.I.) se entiende al conjunto de patrones representativos de una o más estructuras categóricas o 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 2.8 presenta 3 tipos de T.I., en 2 dimensiones, cuyas variables son binarias, categóricas y continuas, respectivamente.

Un patrón es un sector de la imagen de entrenamiento bajo una grilla de búsqueda específica. Sea una T.I. de 20×20 y una grilla de búsqueda móvil de 7×7 es posible encontrar 196 patrones al desplazar la grilla por toda la T.I. La figura 2.9 da cuenta de 2 patrones obtenidos para el ejemplo del caso anterior.

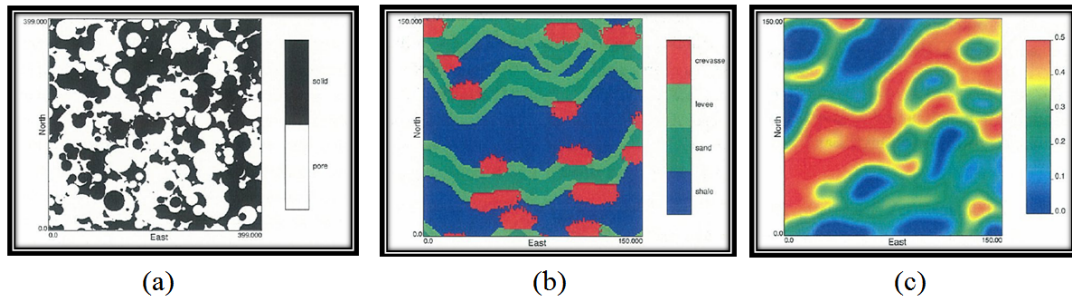


Figura 2.8: Imágenes de entrenamiento. (a) T.I. de una variable binaria – porosidad media. (b) T.I. de una variable categórica – Depósito fluvial con cuatro categorías. (c) T.I. de una variable continua – Distribución de la porosidad. Fuente: [27]

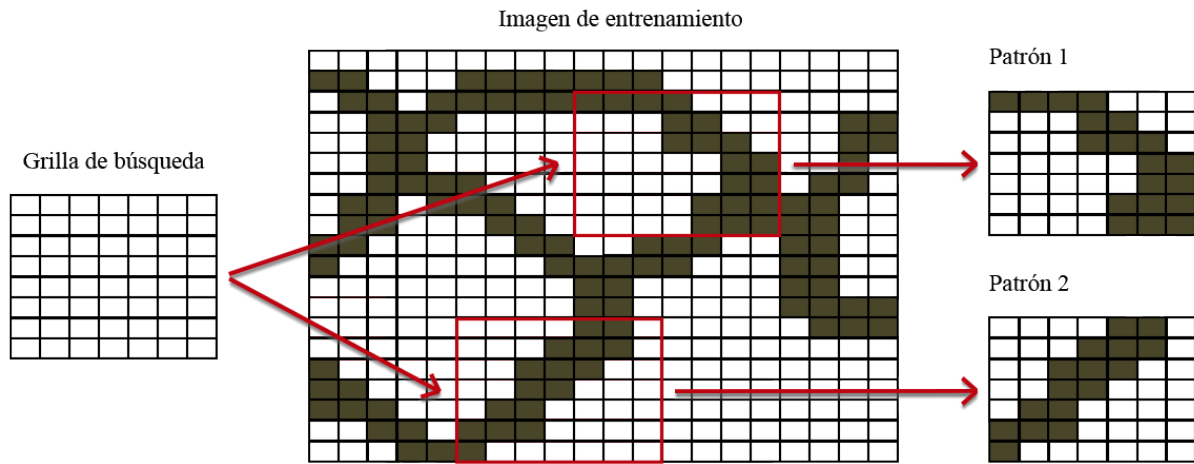


Figura 2.9: Obtención de patrones a partir de una grilla de búsqueda desde una T.I.

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 subsección *Generación del espacio de puntuación (score space)*.

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.

Las dimensiones de la grilla de búsqueda guardan relación con la forma del patrón de comportamiento que se desea inferir. Esta puede ser regular o bien asimétrica. Independiente de su forma ésta debe hacer referencia a un nodo específico que puede estar en su interior (convexa) o en su exterior (no convexa). A modo de implementación del algoritmo se decide utilizar grillas cuadradas o rectangulares. Es necesario que los lados de la grilla de búsqueda 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). En el ejemplo de la figura 2.9 se utilizó una grilla de 49 subdivisiones que provienen de lados $(2n + 1)$ de largo y $(2n + 1)$ de ancho, con $n = 3$.

Filtros

El algoritmo presenta 3 filtros predefinidos, con los cuales se capturan 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 2.10, 2.11 y 2.12 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 $(2n+1) \cdot (2n+1) \cdot (2n+1)$, los filtros quedan definidos como:

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

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

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

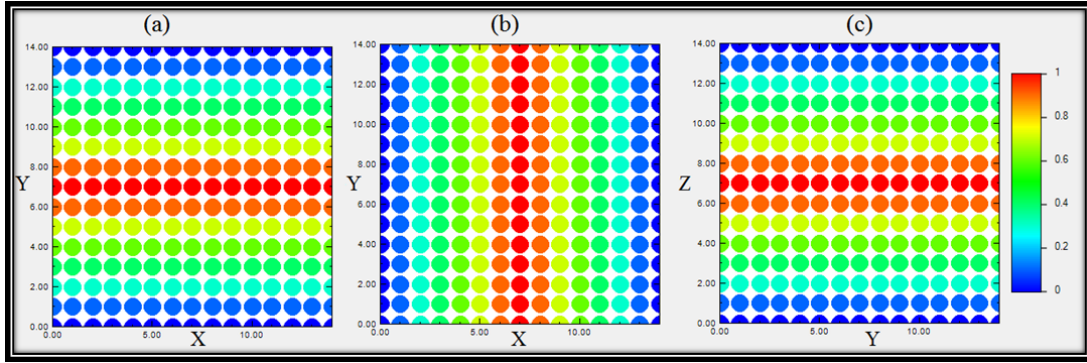


Figura 2.10: Filtros asociados a la media. (a) Norte-Sur, (b) Este-Oeste, (c) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

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, $f_6 : \frac{w}{n}$, $w = -n, \dots, n$

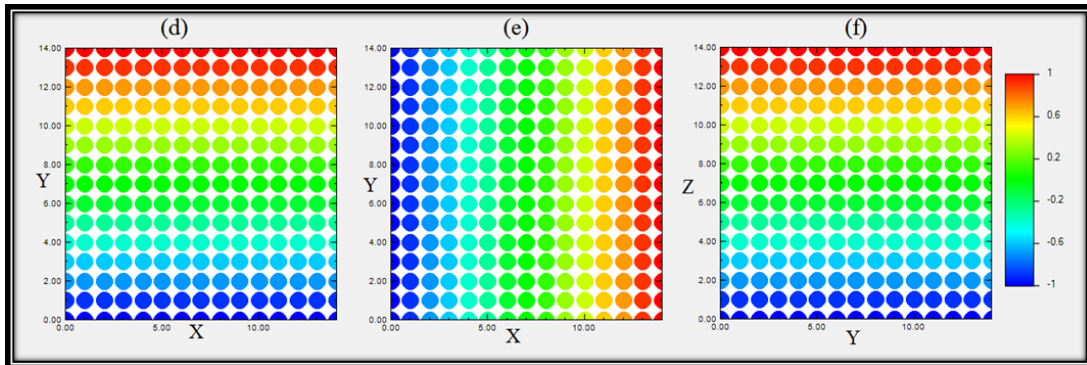


Figura 2.11: Filtros asociados al gradiente. (d) Norte-Sur, (e) Este-Oeste, (f) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

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

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

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

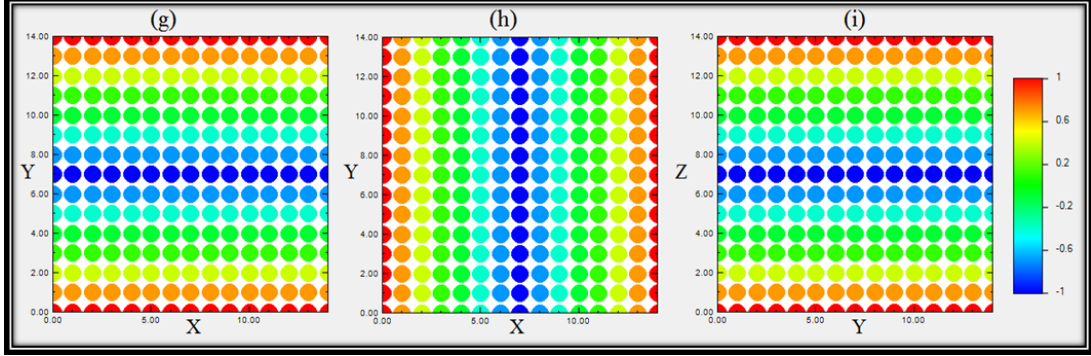


Figura 2.12: Filtros asociados a la curvatura. (g) Norte-Sur, (h) Este-Oeste, (i) Elevación. La escala representa los pesos asociados a cada nodo en la grilla

A continuación se presenta la aplicación de los filtros dentro de una grilla de búsqueda móvil sobre una T.I. en 2 dimensiones [27] 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 de búsqueda cuadrada de tamaño $(2n + 1) \cdot (2n + 1)$. Se tiene entonces que el puntaje $S_f(i, j)$ asociado al punto (i, j) bajo el filtro f viene dado por:

$$S_f(i, j) = \sum_{v=-n}^n \sum_{u=-n}^n f(u, v) \cdot X(i + u, j + v) \quad (2.23)$$

Donde $f(u, v)$ representa uno de los 6 filtros predefinidos o algún otro filtro generado por el usuario. La figura 2.13 presenta como ejemplo la aplicación del segundo filtro sobre el patrón 2 que se obtuvo en la figura 2.9, 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 (2.22) puede ser idéntica para distintas combinaciones de datos.

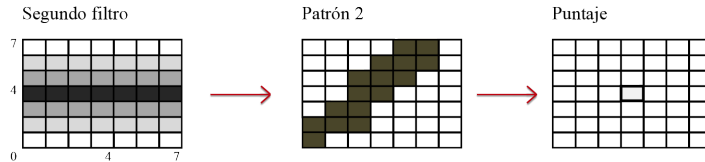


Figura 2.13: Caracterización de un patrón, asignación de puntaje, al aplicar un filtro en específico

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 $(2n + 1) \cdot (2n + 1)$ y una imagen de entrenamiento de tamaño $(p) \cdot (q)$ entonces el tamaño del mapa con los valores de la puntuación obtenida tendrá un tamaño de $(p-2n) \cdot (q-2n)$. Tal como se dijo en secciones anteriores, se le denomina imagen de entrenamiento

erodada a la sección de $(p - 2n) \cdot (q - 2n)$ obtenida desde la imagen de entrenamiento siendo ambos centros coincidentes. La figura 2.14 presenta la imagen erodada del ejemplo expuesto en la figura 2.9.

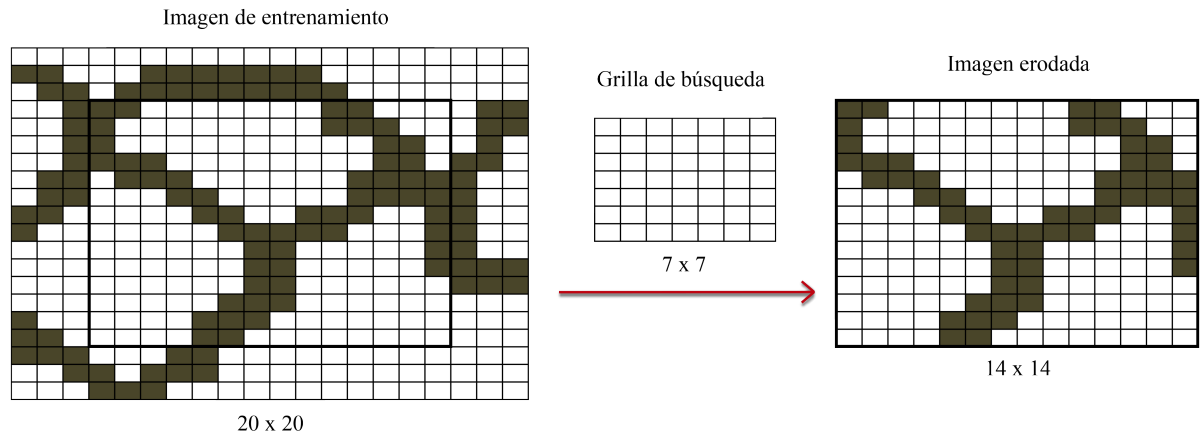


Figura 2.14: Ilustración del concepto de imagen erodada

Es posible encontrar un ejemplo de aplicación de filtros y mapas de puntuación en [27].

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 $F \cdot 1$ en los cuales cada elemento representa el valor en el mapa de puntuación. La figura 2.15 presenta un esquema de lo antes mencionado.

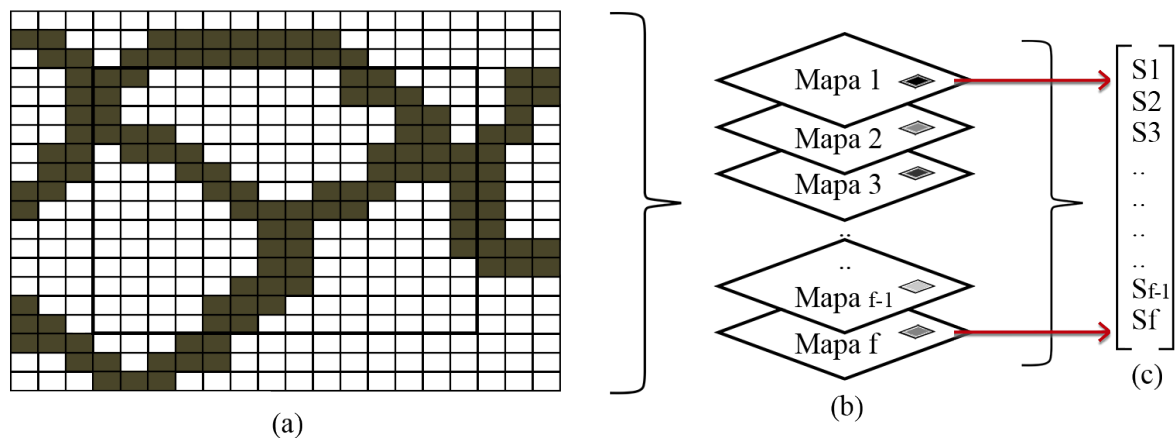


Figura 2.15: Ubicación de un patrón en el espacio de puntuación de F dimensiones. (a) Imagen de entrenamiento, (b) F mapas de puntuación, (c) vector representativo de la ubicación de la puntuación del patrón en el espacio de F dimensiones

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 (evitando analizar cada patrón en la búsqueda de similitud al evento condicionante). Para ello se proponen 2 herramientas para generar el Clustering: a través de *cross partition* [27] o bien *k-mean partition* [13]. A continuación se detallan ambos procesos de agrupamiento.

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 [27]:

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 celdas 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 corresponde al promedio punto a punto de todos los patrones al interior de la clase. La figura 2.16 presenta un ejemplo ilustrativo de 6 patrones similares pertenecientes a la misma clase y el prototipo representativo e identificador de la misma.

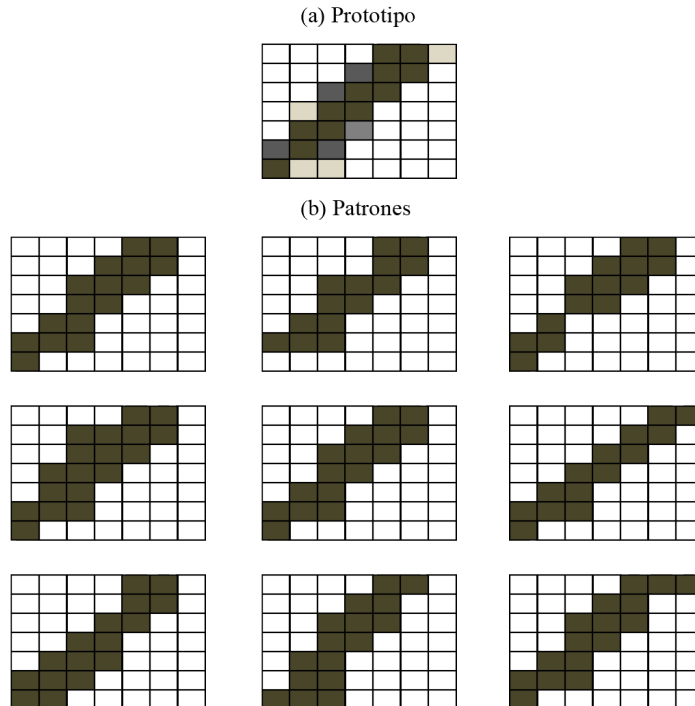


Figura 2.16: Ejemplo de una clase. (a) Prototipo que identifica a la clase, (b) patrones similares dentro de una misma clase

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*) [27] para las variables continuas:

$$SI = 1 - \frac{\left(\frac{1}{n} \sum_{j=1}^n \sigma_j^2\right)}{\sigma^2} \quad (2.24)$$

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 es descrita en esta sección ya que es similar y se prioriza la explicación detallada de las variables continuas dada su mayor aplicación en área de la minería. No obstante, para más detalle en casos categóricos o binarios ver [27].

El índice SI , ecuación (2.24), 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 vendría 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 $SI \geq 0.95$ (alto nivel de homogeneidad), entonces no se subdivide.
 - c) En caso contrario se subdivide la clase en 2 de la siguiente forma:
 - 1) Para los rangos que entrega cada filtro, se propone dividir este por la mitad utilizando la mediana de los valores.
 - 2) Calcular el índice SI para ambas subclases.
 - 3) 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.

K-mean partition

El algoritmo de agrupamiento *K-mean*, propuesto por Hartigan [13], 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 2.17 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.

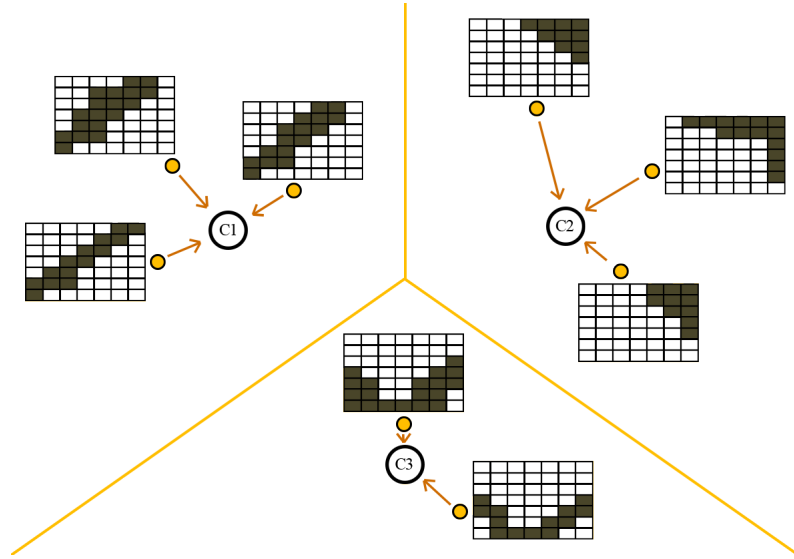


Figura 2.17: Ejemplo del proceso de agrupamiento utilizando el método K-mean. Se observan 3 centroides C1, C2 y C3 que llevan a generar las 3 líneas continuas que establecen los puntos equidistantes a los centroides más cercanos

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 [17]:

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 (2.25):

$$ESS = \sum_{k=1}^K \sum_{C(i)} (x_i - \bar{x}_k)^2 \quad (2.25)$$

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 se tiene el resumen de la tabla 2.1 [2]:

Tabla 2.1: Ventajas y desventajas de los algoritmos de agrupamiento Cross y K-mean [2]

Algoritmo de agrupación	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 cuya cardinalidad (cantidad de elementos) es menos variable	Altamente demandante de velocidad (CPU)

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 [27]:

1. Generación de una grilla de simulación. En ella se ubican los datos duros condicionantes (muestras), 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 utilizado 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_0 , 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 estos no son reemplazados por los nuevos valores. De existir presencia de datos previamente simulados estos si son reemplazados. 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 están 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:

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. Cabe mencionar que los filtros no están directamente utilizados en el cálculo de la distancia (ecuación (2.26)), si no que fueron utilizados en la definición de los patrones ($prot^{(l)}$). 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 (2.26):

$$D_l(u_o) = \sum_{d=1}^3 w(d) \sum_{i=1}^{n_d} \left(dat_d(u_o + u_i) - prot^{(l)}(u) \right) / n_i \quad (2.26)$$

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 (2.27):

$$D_l(u_o) = \sum_{i=1}^F \left| S_{dat}^{(i)}(u_o) - S_{prot^{(l)}}^{(i)}(u) \right| \quad (2.27)$$

Dónde:

- F , es el número de filtros.
- $S_{dat}^{(i)}$ es el valor del filtro i sobre el evento condicionante.
- $S_{prot^{(l)}}^{(i)}$ es el valor del filtro i sobre el prototipo l .

- u_0 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.

Grillas múltiples

Para capturar los comportamientos a gran escala de las estructuras se utiliza el concepto de multigrillas [27] dentro del proceso de simulación. Se asume una grilla de búsqueda, τ , de tamaño n la cual será re-escalada una cantidad G ($G \in 1..g$) de veces. Cada grilla de búsqueda re-escalada (g) tendrá una cantidad n_τ^g de nodos cuya cantidad es la misma en todo momento ($n = n_\tau^g$), 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 2.18 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 $g = G$. Los procesos de simulación 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 condicionantes para la ($g - 1$)-ésima grilla. Esto se logra migrando los nodos simulados de la grilla superior a la inmediatamente inferior. Cabe destacar que estos nodos condicionantes no son fijos y deben ser re-simulados en la ($g - 1$)-ésima grilla aportando información condicionante antes y después de ser simulado.

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

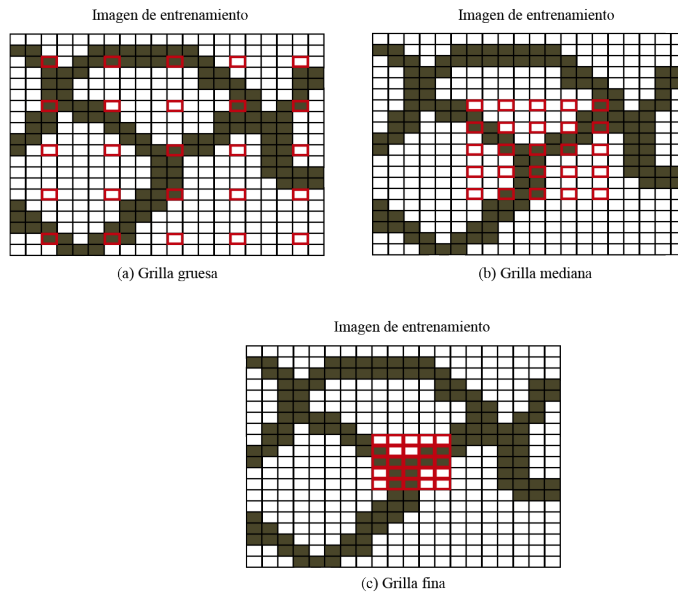


Figura 2.18: Ilustración de 3 multigrillas de tamaño 5×5 sobre una imagen de entrenamiento

2.5.3. Algoritmo Direct Sampling

La técnica de simulación multipuntos Direct Sampling (DS) [18] tiene como gran ventaja la no utilización de árboles de búsqueda (como es en el caso de Snesim) o mapas de puntuación (como es en el caso de Filtersim) inferidos a partir de la imagen de entrenamiento. Esto lleva a una disminución considerable en la demanda de memoria por parte del proceso global. Por su estructura puede ser utilizado tanto para variables categóricas como continuas entregando además la posibilidad de realizar co-simulaciones, lo cual concede gran versatilidad y flexibilidad al algoritmo.

Como toda técnica MPS requiere como input:

- Una imagen de entrenamiento (T.I.) desde la cual inferir el comportamiento de la variable regionalizada, denotando por y cada uno de los nodos al interior de una T.I o la posición espacial de cada dato en un conjunto de datos condicionantes.
- Una grilla a simular cuyo tamaño es definido por el usuario y cuyos nodos se denotan por x .
- (Opcional) Un conjunto de N datos condicionantes ubicados al interior de la grilla a simular.

Se procede de la siguiente manera ([18]):

- Los datos condicionantes son asignados al nodo más cercano dentro de la grilla a simular. Cabe la posibilidad que 2 o más datos condicionantes estén relacionados al mismo nodo, en ese caso el nodo simplemente toma el valor del dato condicionante más cercano.
- Sobre el resto de los nodos no informados, dentro de la grilla a simular, se establece un camino aleatorio a recorrer en el proceso de simulación. Este camino se recorre unilateralmente, es decir, se sigue un orden regular en una dirección definida a priori por el usuario.
- Para cada uno de los x nodos no informados se tiene:
 1. Dado un número máximo de vecinos (n) se determina la vecindad sobre el nodo x como el conjunto de nodos informados (datos condicionantes o nodos previamente

simulados) más cercanos (x_1, x_2, \dots, x_n) . Cabe destacar que si la vecindad está totalmente vacía el algoritmo toma aleatoriamente un punto y en la T.I. cuyo valor $Z(y)$ es copiado directamente en $x(Z(y) = Z(x))$.

2. Se define el vector de distancias L como $L = \{h_1, h_2, \dots, h_n\} = \{x - x_1, x - x_2, \dots, x - x_n\}$, de esta forma la vecindad en torno a x queda definida como $N(x, L) = \{x + h_1, x + h_2, \dots, x + h_n\}$.
3. Se define como evento al vector que contiene el valor de cada variable de interés dentro de la vecindad. Se expresa de la forma $d_n(x, L) = \{Z(x + h_1), Z(x + h_2), \dots, Z(x + h_n)\}$.
4. Sobre la T.I. se define una sección denominada ventana de búsqueda (símil a la imagen erodada en Filtersim) dentro de la cual se ubican todos los puntos y que cumplen con la restricción que los nodos de la vecindad estén dentro de la imagen de entrenamiento $N(y, L) \in T.I.$
5. Sobre la ventana de búsqueda aleatoriamente se establece un nodo y desde el cual se escanea sistemáticamente ésta en alguna dirección. Sobre cada nodo y se realizan los siguientes procesos:
 - a) Se define el evento $d_n(y, L)$ sobre el nodo y .
 - b) Se calcula la distancia entre ambos eventos (grilla a simular, $d_n(x, L)$, e imagen de entrenamiento, $d_n(y, L)$). El cálculo depende de la naturaleza de la variable en análisis, por lo que difiere entre variables continuas y categóricas.
 - 1) **Variables categóricas:** se comparan los valores nodo a nodo entre ambos eventos utilizando el indicador a_i . Si tienen idéntica categoría el indicador a_i toma el valor cero, mientras que si difieren entre ellas toma el valor uno.

$$d\{d_n(x), d_n(y)\} = \frac{\sum_{i=1}^n a_i \|h_i\|^{-\delta}}{\sum_{i=1}^n \|h_i\|^{-\delta}}, \quad (2.28)$$

donde $a_i = \begin{cases} 0 & \text{si } Z(x_i) = Z(y_i) \\ 1 & \text{si } Z(x_i) \neq Z(y_i) \end{cases}$

Aplicando una función potencia de orden δ sobre el vector distancia h_i entre el nodo a estimar y el nodo informado.

- 2) **Variables continuas:** al igual que para variables categóricas se compara el valor nodo a nodo con la salvedad que en este caso se utiliza una función de distancia euclidiana definida en la ecuación (2.29).

$$d\{d_n(x), d_n(y)\} = \sqrt{\sum_{i=1}^n \alpha_i [Z(x_i) - Z(y_i)]^2}, \quad (2.29)$$

donde $\alpha_i = \frac{\|h_i\|^{-\delta}}{d_{\text{máx}}^2 \sum_{j=1}^n \|h_j\|^{-\delta}}$, $d_{\text{máx}}^2 = \max_{y \in TI} Z(y) - \min_{y \in TI} Z(y)$

Cabe mencionar que existe una variedad de funciones que pueden ser utilizadas para el cálculo de distancia [18].

- c) Se almacena tanto la posición y , como el valor en el nodo $Z(y)$ y la distancia obtenida $d\{d_n(x, L), d_n(y, L)\}$.
- d) Si la distancia $d\{d_n(x, L), d_n(y, L)\}$ presenta un valor menor a un cierto umbral t definido por el usuario entonces el valor $Z(y)$ es copiado y asignado al valor en $x(Z(x) = Z(y))$.
- e) De no cumplirse la condición umbral t se repite el proceso en el nodo y siguiente. De superarse una cierta fracción de la T.I., definida por el usuario, y no haber logrado una distancia menor a t se procede a escoger el valor $Z(y)$ cuya distancia sea la menor entre todas las estudiadas.

6. Se repite el paso de a hasta e hasta que la grilla a simular está completamente informada.

2.5.4. Algoritmo de síntesis textural

Un algoritmo de síntesis textural es una herramienta capaz de generar diversas texturas considerando diversas fuentes de información textural. Como definición directa se tiene: “Dada una muestra textural, sintetizar una nueva textura que, al ser percibida por un observador humano, parece ser generada por el mismo proceso subyacente” ([25]).

Wei Li-Yi desarrolló un algoritmo capaz de reproducir los comportamientos texturales al ser inferidos desde una imagen textural de referencia. Trabajos similares fueron desarrollados tomando técnicas como Simulaciones Físicas (modelos teóricos directos de los procesos de generación), Campos Aleatorios de Márkov (Markov Random Field, MRF) y Muestradores de Gibbs, entre otros. Wei basado en los principios de los MRF propone, globalmente, el siguiente algoritmo [20]:

1. Generar la distribución condicional, en un punto, de la variable en estudio (independiente de su naturaleza categórica o continua) desde una imagen de referencia textural dada una cierta vecindad.

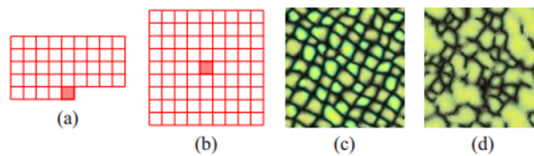


Figura 2.19: (a) Vecindad causal, (b) vecindad no causal, (c) síntesis textural al utilizar (a), (d) síntesis textural al utilizar (b). [25]

La vecindad debe tener un carácter causal, es decir, que los datos internos fueron simulados y condicionan el nuevo punto a simular. En la figura 2.19.a se observa una vecindad causal mientras que en la figura 2.19.b una no causal, esta última es idéntica a la grilla de búsqueda utilizada por el algoritmo Filtersim. Se observa en la figura 2.19.c y 2.19.d como la textura queda bien caracterizada al utilizar una vecindad causal y lo contrario al implementar una no causal.

2. Definir una grilla a simular la cual puede o no tener nodos previamente informados. Sin embargo, estos serán luego reemplazados.
3. Ir visitando cada nodo no informado de manera ordenada y fijando siempre la misma vecindad con la finalidad que los nodos previamente informados sirvan como condicionamiento para el nodo a ser simulado. Existe la posibilidad, sobre todo al comienzo del proceso, que nodos dentro de la vecindad de búsqueda queden fuera de la grilla a simular, en este caso los nodos no son tomados en consideración. Si no existen nodos informados dentro de la vecindad un valor cualquiera es escogido dentro de la imagen textural de referencia.
4. Completar el camino hasta que la grilla quede completamente informada.

El método carece de poder condicionar cada una de las realizaciones. Sin embargo, Parra y Ortiz proponen una variación de la técnica para lograr texturas condicionadas [20].

Una aplicación interesante es la reconstrucción de imágenes, tal como se ilustra en la figura 2.20. Es posible la interpolación de la imagen, es decir, los contornos de la grilla a simular deben estar informados y el centro de ella es luego reconstruido. El proceso anterior es conocido como *importing*. La extrapolación de la imagen es el caso contrario donde los contornos son simulados

mientras que el centro está previamente informado.

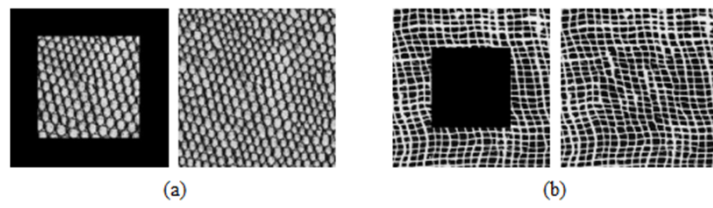


Figura 2.20: Reconstrucción textural. (a) Extraplación de una imagen, (b) interpolación de una imagen [25]

Entre otras aplicaciones destacan la renderización de imágenes y la restauración de imágenes dañadas.

2.5.5. Algoritmo DisPat

Los algoritmos basados en objetos, en específico Simpat [1] y Filtersim [26, 27], infieren el comportamiento de la variable regionalizada a partir de la extracción de patrones desde una imagen de entrenamiento. Ambos algoritmos son bastante sencillos a la hora de agrupar los patrones que presentan características comunes. Más aún, para el caso de Filtersim el concepto de estocasticidad viene dado por una elección aleatoria del patrón representativo dentro de una clase previamente determinada. Como respuesta a la simpleza de agrupación Honarkhah y Caers proponen el algoritmo *DisPat (Distance-Based Pattern Modeling)* [14], técnica que utiliza métodos basados en distancias para lograr procesos de agrupación y modelación de los patrones computacionalmente más eficiente. Para ello se plantea la representación de los patrones en un espacio cartesiano utilizando funciones de distancia para analizar la variabilidad que existe entre todos los posibles patrones a partir de la imagen de entrenamiento.

La función distancia requiere como variables de entrada dos patrones, no necesariamente distintos, y genera un valor numérico que representa la similitud entre ambos. Un valor bajo, cercano a cero, representa una alta similitud mientras que valores altos representan una alta diferencia. Sea T la grilla de búsqueda de patrones sobre la T.I. y sea n_{Pat_T} la cantidad de patrones obtenidos desde la T.I., es posible definir la matriz de disimilitud D de $n_{Pat_T} \times n_{Pat_T}$ en donde cada elemento $\sigma_{i,j}$ representa el valor numérico obtenido con la función distancia entre los patrones i -ésimo y j -ésimo. Al observar el set completo de valores al interior de la matriz es posible inferir relaciones entre los patrones. Estas relaciones pueden ser reveladas a través de la técnica de escalamiento multidimensional.

El escalamiento multidimensional (MDS, *multidimensional scaling*) es un conjunto de técnicas que buscan representar una configuración de datos en un espacio de muchas dimensiones en uno de pocas dimensiones preservando lo más posible la distancia entre los elementos. En efecto, el usuario tiene la capacidad de definir la dimensionalidad del espacio final. Sea n (\mathbb{R}^n) el espacio original y d (\mathbb{R}^d) ($d < n$) el espacio final al que se desea escalar el problema, se observa empíricamente que tomar un valor alto de d (cercano a n) lleva a muy bajos errores pero una alta demanda del tiempo de cómputo, mientras que un valor bajo de d (cercano a 1) lleva a bajos tiempos de cómputos pero no logra representar correctamente las estructuras/relaciones entre los elementos. Se observa en la figura 2.21 una aplicación de la técnica MDS que lleva una imagen de entrenamiento a un espacio de 2 dimensiones en donde puntos cercanos representan patrones similares.

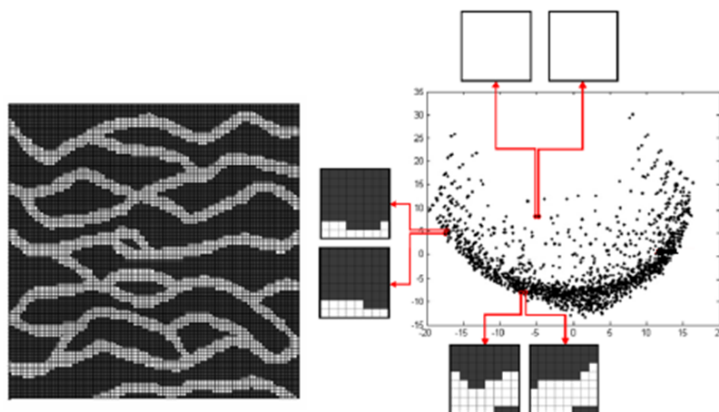


Figura 2.21: Ejemplo de aplicar una grilla de 8x8 sobre una T.I. (izquierda) y aplicar luego un MDS de 2 dimensiones (derecha). Fuente: [14]

Dado que el objetivo principal de DisPat es la propuesta de una técnica capaz de representar las características de una variable regionalizada, sin la utilización de filtros (como en el caso de

Filtersim), se propone directamente una técnica para la clasificación de patrones. Específicamente, al tener los patrones representados en un espacio d -dimensional luego de aplicar la MDS es posible generar una clasificación de patrones similares. Una de las técnicas más utilizadas es *k-mean partition*, sin embargo adolece de no lograr caracterizar de manera adecuada relaciones no lineales entre los patrones. Como respuesta, Honarkhah y Caers proponen el Kernel del conjunto de patrones para su agrupación, el cual puede ser utilizado directamente al conjunto o al resultado de aplicar MDS previamente. Esto último depende de las características de los patrones. Para más detalles ver [14].

Las ventajas de utilizar DisPat por sobre los métodos convencionales, como el caso de Filtersim, son expuestas en la publicación antes mencionada y a modo de resumen se tiene que: caracteriza de mejor manera la continuidad espacial, extrae relaciones espaciales más complejas (ejemplo, presencia de 4 o más estructuras en una variable categórica), entre otras.

2.5.6. Algoritmo Impala

Como se ha mencionado anteriormente, todos los algoritmos de simulación multipuntos requieren de una T.I. desde la cual inferir el comportamiento de la variable regionalizada. Esta información, en específico las estadísticas multipuntos, son o bien extraídas directamente sin un previo almacenamiento (Direct sampling [18]) o bien almacenadas (Snesim [23], Filtersim [26, 27],) para generar las distribuciones de probabilidad condicional a la hora de simular un valor. Diversas técnicas de almacenamiento han sido propuestas o bien mejoras a las mismas. Tal es el caso de DisPat [14] que busca mejorar la demanda de memoria computacional al aplicar técnicas de MDS sobre la base de patrones. En el caso del algoritmo Snesim [23] las estadísticas multipuntos son almacenadas en un árbol de búsqueda, sin embargo, esta metodología adolece de ser demandante en memoria computacional de manera exponencial al aumentar el tamaño de la T.I. o al número de categorías dentro de la misma. El presente algoritmo propone modificar la forma de almacenar la información dejando de lado los árboles de búsqueda y pasando a arreglos de listas, logrando una mejora considerable en la demanda de almacenaje. Debido a este propósito es que recibe el nombre de Impala (*An Improved Parallel Multi-point Algorithm Using a List Approach*) [21] y dentro de sus bondades principales se tiene que:

- Da flexibilidad en el uso de grillas de búsqueda para la extracción de patrones. Con la metodología de árboles de búsqueda un aumento en el tamaño de la grilla de búsqueda llevaba a un aumento considerable en la demanda de almacenamiento, más aun si se está trabajando en tres dimensiones. Con la propuesta de listas el tamaño de la grilla y la dimensionalidad de la T.I. no tienen mayor impacto en la demanda de almacenaje. Esta idea es ilustrada en la figura 2.22.

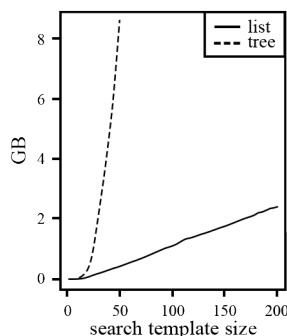


Figura 2.22: Demanda de memoria RAM. Imagen de entrenamiento en 3 dimensiones con 4 categorías y 600.000 nodos [21]

- Es capaz de incluir información adicional sin necesidad de modificar la información actual ni rehacer los procesos anteriores. Esta propiedad toma vital importancia cuando se debe recurrir al uso de imágenes de entrenamiento no estacionarias.
- El algoritmo es paralelizable. En particular, no paraleliza la simulación en diversos nodos para llevar a cabo una simulación más rápida, sino que paraleliza el proceso de cómputo a la hora de determinar la función de distribución condicional para cada nodo en la grilla a simular. Esto lleva a que no solo sea menos demandante de memoria sino que más veloz.

Con el apoyo del ejemplo teórico en [21] se detalla el proceso de almacenaje. Se define una grilla de búsqueda de N nodos a aplicar sobre la imagen de entrenamiento. En la figura 2.23 se observa una T.I. de 6×6 con dos categorías, la categoría 0 representada por color blanco y la categoría 1 por negro, además de una grilla de búsqueda de 4 nodos en la cual el nodo central, v , es el nodo referencial de búsqueda. Sea $d = (s_1, \dots, s_N)$ la representación de un evento condicionante en particular, se define entonces el vector $c = (c_0, \dots, c_{M-1})$ el cual indica para cada c_i la cantidad de veces que el evento condicionante $d(v)$ está presente en la imagen de entrenamiento y el nodo referencial v es idéntico a la categoría i . Para cada configuración de evento condicionante, d , se obtiene un vector c almacenado finalmente como un vector (d, c) . A la derecha de la figura 2.23 se observa la lista final de vectores procedentes de la T.I. y grilla de búsqueda de la misma figura.

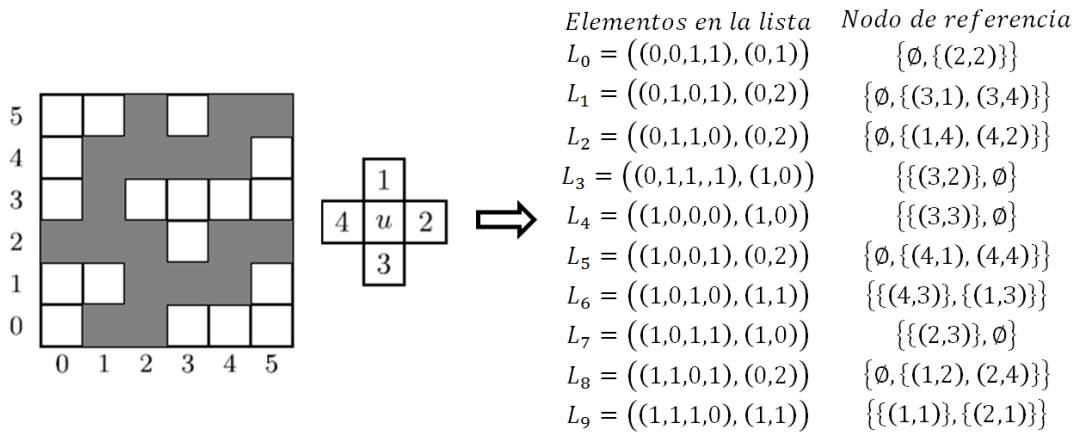


Figura 2.23: (izq.) T.I. donde los blancos son 0 y negros son 1. (der.) Grilla de búsqueda de 4 nodos

Una comparación directa en el uso de memoria RAM (demanda computacional), la descripción de la paralelización del método en el cálculo de la distribución de probabilidad condicional para cada nodo, la integración de una nueva variable para imágenes de entrenamiento no estacionarias, y un caso de estudio se encuentran desarrolladas con precisión en la publicación de Straubhaar et. al. [21].

2.5.7. Algoritmo Simpat

En los algoritmos multipuntos determinar sobre un conjunto de imágenes de entrenamiento cual es la más adecuada para caracterizar el dominio a simular resulta habitualmente difícil. El principal problema es la inferencia de los patrones y la capacidad de ellos de interpretar las propiedades a pequeña y gran escala de la variable regionalizada. Además, los patrones a gran escala están usualmente relacionados con los patrones a pequeña escala siendo relevante capturar esa información. El presente algoritmo propuesto por Arpat y Caers [1] denominado *Simpat* (*Simulation with Patterns*) propone incorporar esta información y servir como un complemento favorable para el resto de los algoritmos multipuntos en la etapa de inferencia de los patrones.

Sea T la grilla de búsqueda, definida por los vectores h_α con $\alpha = 0, \dots, n_{T-1}$, sobre la imagen de entrenamiento \mathbf{ti} (misma utilizada en la grilla a simular \mathbf{G}). Se obtienen todos los patrones pat_T^k posibles desde la \mathbf{ti} generando una base de datos de dimensionalidad n_{pat} . Los patrones vienen definidos por:

$$pat_T^k = \{pat_T^k(h_0), pat_T^k(h_1), \dots, pat_T^k(h_{n_{T-1}})\} \text{ con } k = 0, \dots, n_{pat}$$

La metodología usual es definir la grilla de búsqueda que recorra la imagen de entrenamiento para obtener la base de posibles patrones. Luego se establece una grilla a simular sobre la cual un camino aleatorio es definido sobre los nodos y se visita cada uno una sola vez extrayendo el evento condicionante $dev_T(u)$. Este evento condicionante incorpora información de datos duros y nodos previamente simulados mientras que los nodos no informados no son tomados en consideración. A través de una función distancia se calcula la diferencia entre el evento condicionante y cada uno de los patrones con la finalidad de encontrar el mínimo valor de esta diferencia y establecer así el patrón más cercano al evento condicionante. Una vez determinado este patrón sus valores son congelados sobre los nodos correspondientes teniendo la precaución de no reemplazar los datos duros. El proceso sigue sobre el camino aleatorio hasta terminar por completo la grilla a simular. El problema se genera cuando las estructuras a inferir son de gran tamaño, en este caso se opta por utilizar grillas de búsqueda más grandes lo cual resulta costoso en memoria y tiempo al proceso completo de simulación. Como respuesta a esta problemática se plantea el uso de técnicas de multigrillas y es en este proceso del algoritmo que *Simpat* presenta una nueva alternativa denominada *Grilla dual*.

En el proceso de simulación utilizando multigrillas los valores simulados en la grilla más gruesa son luego migrados y congelados a la grilla subsiguiente. La idea propuesta en *Simpat* es no congelar estos valores sino que los deja afectos de cambio. Además incorpora la relación entre patrones de gran y pequeña escala. Para ello se supone la utilización de una grilla de búsqueda T en la grilla más gruesa sobre la imagen de entrenamiento, asociado a T se tiene una grilla de búsqueda más fina T' cuya área (o volumen en tres dimensiones) abarca la misma área que T . Esta nueva grilla T' es denominada *grilla dual* mientras la grilla original T es denominada *grilla primitiva* (ver figura 2.24).

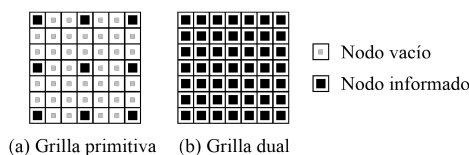


Figura 2.24: (a) Grilla primitiva y (b) Grilla dual. Fuente: [1]

Al extraer los patrones desde la \mathbf{ti} utilizando la grilla primitiva se enlazan estos a su correspondiente patrón bajo la grilla dual. Comenzando el proceso de simulación sobre la grilla más gruesa utilizando una grilla T una segunda grilla T' en la grilla de simulación más fina

simulará un sector de la misma área (o volumen) que T . El proceso condicionante es la búsqueda del patrón más cercano en la grilla gruesa, sea pat_T^k el patrón bajo la grilla T entonces se utiliza su patrón enlazado $pat_{T'}^k$, proveniente de la grilla T' . Suponiendo la simulación de una variable categórica 0 o 1 utilizando multigrillas de orden 2 se tendría, tal como se observa en la figura 2.25, que el evento condicionante es extraído desde la grilla gruesa. Este es luego comparado con la base de datos de patrones estableciendo el patrón más cercano. Este patrón viene asociado con un patrón dual único cuyo valor en los nodos es copiado sobre la grilla fina a simular.

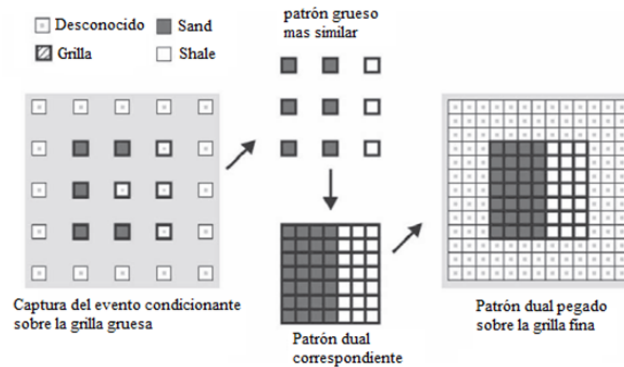


Figura 2.25: Implementación de la grilla dual sobre una variable categórica. Fuente: [1]

El proceso de interacción de los distintos niveles ocurre cuando se procede a simular la grilla subsiguiente. Sobre la nueva grilla a simular (más fina) migran los valores de la grilla fina hacia esta sobre los nodos no informados. Suponiendo 3 niveles de grillas, al simular el nivel de la grilla gruesa queda automáticamente simulada la grilla fina. Luego previo a simular sobre la grilla media, los valores de la grilla fina son migrados a la grilla media en los nodos no informados re-simulando la grilla fina.

Finalmente esta técnica no es excluyente del resto de los algoritmos de MPS y puede ser implementada como variación de las técnicas de multigrillas.

2.5.8. Comentarios sobre el estado del arte y elección del algoritmo a utilizar - Filtersim

Luego de haber planteado los principales algoritmos de simulación multipuntos se presenta la siguiente sección en donde se comenta brevemente cada uno de ellos. Lo anterior sirve como argumento para la elección de Filtersim como algoritmo a incorporar derivas en la media. El orden de los algoritmos sigue el orden de aparición en la presente sección.

Snesim (2002): fue uno de los primeros algoritmos en implementar las técnicas MPS. Se caracteriza por ser un algoritmo del tipo secuencial. Tiene aplicación sobre variables categóricas y si se desea utilizar sobre variables continuas éstas deben estar subdivididas en intervalos finitos. Adolece de ser costoso en memoria de cómputo dada su estructura de almacenaje en forma de árbol de búsqueda. Es sensible al tamaño de la imagen de entrenamiento, es decir, a mayor tamaño de ésta, mayor será el tiempo utilizado en la creación del árbol de búsqueda (árbol de condiciones posibles).

Filtersim (2006): se caracteriza por utilizar filtros sobre la imagen de entrenamiento para la extracción de patrones estructurales. Dado lo anterior reduce la dimensionalidad del problema, es económico en la demanda de memoria y no es sensible al tamaño de la imagen de entrenamiento. Brinda la posibilidad de aplicación tanto para variables categóricas como continuas. Además existe información bibliográfica suficiente respecto a la estructura interna del algoritmo

lo cual facilita su implementación y futura modificación. Sin embargo, adolece de ser sensible a las características de los filtros implementados a la hora de agrupar los patrones comunes previo proceso de simulación.

Direct Sampling (2010): una de las últimas, si no la última, técnica de MPS. Su principal característica consiste en desligarse completamente de la demanda de memoria en el proceso de inferencia de patrones desde la imagen de entrenamiento. No utiliza árbol de condiciones (caso de Snesim) ni filtros para generar mapas de puntuación (Filtersim), simplemente se establece el tamaño y forma de la grilla de búsqueda muestreando con ella la imagen de entrenamiento en búsqueda del evento condicionante sin necesidad de construir la función *cpfd*. La desventaja de esto, es que no se respeta a priori la probabilidad de ocurrencia de un valor/patrón. Por construcción el algoritmo puede ser aplicado tanto para variables categóricas como continuas brindado además la posibilidad de generar co-simulaciones.

Algoritmos de síntesis textural (2001): esta técnica es capaz de generar texturas considerando diversas fuentes de información textural. Puede ser aplicada tanto a texturas continuas como discretas. Adolece de no poder condicionar las realizaciones. Sin embargo, variaciones a la técnica han sido propuestas. Respecto a la grilla de búsqueda (vecindad) obliga a que ésta sea del tipo causal, es decir, las celdas internas fueron simuladas (o presentan datos condicionantes) y condicionan el nuevo punto a simular. Presenta ventajas respecto al resto de los algoritmos MPS cuando se busca reconstruir el sector de una imagen. Por reconstrucción textural se tiene tanto la extrapolación como interpolación de ciertos sectores en específico.

DisPat (2010): esta técnica propone la utilización de herramientas más complejas en el proceso de agrupación de patrones en los algoritmos MPS. Utiliza una matriz de distancias en la que cada elemento δ_{ij} representa el resultado de una función distancia sobre el par de patrones *i-ésimo* y *j-ésimo*. Un valor alto representa una alta disimilitud mientras que entre más cercano a cero mayor similitud. Luego, se interpreta la matriz a través de la técnica de escalamiento multidimensional (MDS). Tiene la capacidad de caracterizar relaciones no lineales entre los patrones. Además tiene la ventaja de extraer relaciones espaciales más complejas y mejora la caracterización de la continuidad espacial respecto a métodos como *k-mean partition* y *cross partition*. Como desventaja se tiene el alto tiempo empleado en llevar a cabo todo el proceso antes descrito, y la necesidad de implementar/programar la técnica al interior de los algoritmos actuales.

Impala (2011): técnica que propone una modificación en la manera en que la información proveniente de la imagen de entrenamiento es almacenada (árbol de búsqueda, mapas de puntuación, etc.). Para ello propone la utilización de arreglos de listas. Con ello se disminuye considerablemente la demanda de memoria y es casi invariante al tamaño tanto de la imagen de entrenamiento como de la grilla de búsqueda. Tiene la capacidad de integrar información adicional sin alterar la información actual, solamente se debe agregar un nuevo elemento al arreglo. Por su construcción es completamente paralelizable disminuyendo así el tiempo de cómputo. Como desventaja es que patrones similares, pero no idénticos, quedan agrupados en listas aparte aumentando la demanda de memoria RAM, por lo tanto, al tener una imagen de entrenamiento de gran tamaño las listas llega a ser demasiado grande.

Simpat (2005): presenta una alternativa para integrar las relaciones de pequeña y gran escala. Al postular una grilla dual se aumenta el costo del algoritmo respecto a la demanda de memoria. Sin embargo los resultados presentados por Arpat y Caers [1] muestran mejoras consistentes en la continuidad espacial de los cuerpos. Es posible aplicar esta técnica en los diversos algoritmos MPS.

Se observa que los principales algoritmos MPS son Snesim, Filtersim y Direct Sampling (sin cerrarse a la posibilidad de modificaciones de los mismos) mientras que el resto son técnicas

o herramientas que mejoran procesos puntuales de cualquiera de ellos. Finalmente se opta por implementar cambios en la media local en el algoritmo Filtersim ya que simula sobre variables continuas sin necesidad de discretizarlas en intervalos, es invariante al tamaño de la T.I., reduce la dimensionalidad del problema ahorrando memoria RAM, y su estructura interna está bien documentada facilitando su implementación.

2.5.9. Aplicaciones de los algoritmos MPS

Las secciones previas tienen un énfasis en la descripción teórica de los algoritmos multipuntos (MPS). A continuación se detallan brevemente 2 aplicaciones en la industria y la ciencia de los recursos naturales. Durante los últimos años el esfuerzo ha estado enfocado en el estudio teórico de las técnicas MPS dando paso a su utilización como herramienta de simulación. Sus aplicaciones van desde pequeña hasta gran escala. En cualquiera sea, obtener una imagen de entrenamiento lo suficientemente representativa resulta crucial. En la industria del petróleo, la hidrología de subsuelos, y otras áreas las TI provienen de la interpretación experta de un grupo reducido de muestras o bien variables secundarias. En la industria minera, la necesidad de cuantificar los recursos lleva a campañas de sondajes más densa y continuas en el tiempo. Esto lleva a tener distintas imágenes de entrenamiento informadas. Sin embargo, dada la génesis del yacimiento y la posterior alteración, la distribución espacial del elemento de interés tiende a no presentar patrones disminuyendo la eficacia de las técnicas MPS. A pesar de lo anterior, se presenta como primera aplicación la simulación de los contactos entre unidades geológicas en un yacimiento de hierro.

Simulación de los contactos entre unidades geológicas

Determinar la forma y la posición de los contactos entre unidades geológicas es una tarea que demanda tiempo y un conocimiento experto por parte del geólogo, quien toma en consideración la información de sondajes, la geología estructural, los principios metalogénicos que dan creación a los yacimientos, entre otras variables. Sin embargo, esto lleva a generar un único modelo geológico. Boucher, et al [3] utilizan una variación de los algoritmos, llamada *contactsim*, para simular el contacto entre unidades geológicas, basado en el modelo geológico inicial. La técnica consiste en:

1. Dado el modelo geológico inicial el usuario define una zona de incertidumbre en torno al contacto entre las unidades geológicas.
2. La zona a simular y la imagen de entrenamiento corresponden a todo el modelo geológico comprendido al interior de la zona previamente definida.
3. Los patrones son agrupados e inferidos separándose previamente en cada conjunto de intersección existente entre unidades geológicas.
4. El condicionamiento proviene de los sondajes de exploración presentes solo al interior de la imagen de entrenamiento. El condicionamiento no es una condición necesaria.
5. Se determina el patrón más cercano a través de una función distancia.
6. Se inicia el proceso de simulación. *contactsim* perturba el sector a simular definiendo la zona de contacto continua en el espacio respetando los patrones inferidos y el condicionamiento de forma similar al resto de los algoritmos multipuntos.

El algoritmo es aplicado sobre un yacimiento de Hierro en el cuadrilátero férreo de Minas Gerais, Brasil. Cuatro litologías son consideradas: lastre, itabirita dura, itabirita friable y hematita. Un radio de 30 metros es considerado desde la zona de contacto en el modelo geológico inicial. Once particiones son generadas por la intersección de las litologías. Luego una sería de

realizaciones entregan las posibles posiciones de la zona de contacto. El post proceso de las realizaciones permite obtener el e-type y los mapas de probabilidad. Los resultados aseguran que *contactsim* reproduce la proporción de las 4 litologías simuladas.

Finalmente, la expertiz de quien modela la zona de contacto recae en definir la zona de incertidumbre dentro de la cual se llevará a cabo la simulación. La calidad de los resultados dependen de la calidad de la agrupación de patrones, ya que una alta variabilidad por clase lleva a contactos poco definidos y con alta variabilidad. La técnica es deficiente cuando las litologías son pequeñas, ya que es probable que la zona de contacto no haya sido muestreada por campañas de sondajes u otras técnicas. La dependencia de un modelo geológico preexistente reduce sus campos de aplicación principalmente a la Minería (campañas de sondajes) y la industria del petróleo (uso de la sismología para la generación del modelo geológico).

Estratos de arcillas – Conductividad, permeabilidad y coeficiente de absorción

Huysmans et al [16] aplican la simulación multipuntos en un contexto hidrológico. Determinan la permeabilidad de una zona específica en Bélgica. Para ello estudian el paso de trazadores (sodium naphthionate) a través de una zona con presencia de estratos de arcilla. Estos estratos son delgados pero abundantes, y dado su alto grado de absorción y baja permeabilidad resulta crucial su modelación en la determinación de la permeabilidad. Utilizando técnicas de MPS modificadas [15] generan posibles escenarios del comportamiento de los estratos asumiendo heterogeneidad y la inferencia del comportamiento a partir de un modelo convencional (imagen de entrenamiento) creada artificialmente (representando el comportamiento real). Se generan modelos en 2 dimensiones a lo largo de la vertical y luego para obtener un modelo en 3 dimensiones utilizan el algoritmo SNESIM. Como resultado se tiene que la técnica de simulación multipuntos es adecuada y eficiente para la modelación de estructuras geológicas complejas y heterogéneas. En desmedro de su utilización está el elevado requerimiento computacional para soportar los modelos que incluyen pequeñas y grandes escalas así como el tiempo requerido para realizar cada una de las simulaciones, siendo un impedimento para sensibilizar los parámetros.

Capítulo 3

METODOLOGÍA

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

1. Incorporación de derivas en la media local en el algoritmo Filtersim.

Se presenta un conjunto de posibilidades que llevan a cabo esta incorporación y se define cuál de ellas es la más viable y que a priori entrega mejores resultados y mayor flexibilidad al usuario. Habiendo definido la estrategia anterior se implementa el algoritmo Filtersim en Python brindando la posibilidad de integrar las modificaciones. Cabe destacar que esta etapa es crucial para los desarrollos siguientes y demandó bastante tiempo respecto al periodo completo de la tesis (un 40 % del tiempo, por lo menos). Se finaliza con una descripción detallada del proceso llevado a cabo por Filtersim incorporando derivas en la media local.

2. Primer caso de estudio: corte en roca sedimentaria.

Sobre dos cortes pulidos de roca sedimentaria se estudia la continuidad de estratos (las dimensiones y procedencias se detallan en el caso de estudio). El primero sirve como imagen de entrenamiento y el segundo como *Ground Truth*. Desde la primera se infieren los patrones del comportamiento estructural y de la segunda los datos duros condicionantes. Dos conjuntos de datos condicionantes son extraídos de forma aleatoria: el primero cuenta con 15 puntos informados (denominado *muy pocos datos condicionantes*) y el segundo con 200 (denominado *pocos datos condicionantes*). Además se subdivide en subconjuntos de 50 % y 100 % de datos lo que lleva a tener 4 conjuntos de datos con 7, 15, 100 y 200 puntos informados respectivamente. Dos tipos de derivas en la dirección Este-Oeste (coordenada X) son utilizadas: deriva impuesta e interpretada. La impuesta proviene de una inferencia acabada de la deriva en el *Ground Truth* y la interpretada se infiere a partir de los datos condicionantes (por lo que varía dependiendo del conjunto de datos condicionantes seleccionado). Luego, sobre un dominio equivalente al *Ground Truth*, se realizan las simulaciones con Filtersim convencional y modificado el cual incorpora la deriva a distintos niveles de influencia, pasando de 0 % a 100 %. Lo anterior se realiza para cada par *Datos condicionantes - Deriva externa*.

Se definen y fijan los parámetros asociados a la simulación por Filtersim. Se generan 100 realizaciones en cada simulación. Los resultados son presentados visualmente, tanto realizaciones como el promedio entre las 100. Para validarlos se analizan las estadísticas básicas, la variografía y la continuidad espacial visualmente. Una vez validado el modelo se analizan las derivas resultantes comparando entre las derivas del *Ground Truth*, de la deriva interpretada/impuesta y el promedio entre las 100 derivas asociadas a las realizaciones. Se compara así la calidad de los resultados entre las distintas combinaciones *Datos condicionantes - Deriva externa* mediante el Error Cuadrático Medio sobre las derivas obtenidas respecto al *Ground Truth*. Además se establecen los mapas de probabilidad sobre un valor umbral y se analiza el porcentaje de acierto de la simulación a distintos niveles de influencia

de la deriva externa. Finalmente se contrastan los resultados con los mapas de valores esperados al utilizar estimación por Kriging y los resultados de simular utilizando el algoritmo de Bandas Rotantes.

3. Segundo caso de estudio: yacimiento Cerro Colorado - BHP Billiton.

Llevando el algoritmo a una aplicación más ligada al rubro minero, en especial a la minería cuprífera se decide estudiar un sector de la mina Cerro Colorado, propiedad y operada por BHP Billiton. Se analiza sólo las leyes de Cobre Total dentro de una unidad geológica previamente definida. Se establece un banco y un sector acotado de este desde el cual extraer la imagen de entrenamiento basada en las leyes de pozos de tronadura. Los cuatro bancos inferiores subsiguientes cuentan con información de pozos de tronadura y sondajes de exploración. Para cada uno de ellos se extraen los datos condicionantes a partir de los sondajes de exploración migrados a una grilla regular. Se extraen las derivas de cada banco a partir de los datos de sondajes en la dirección Este-Oeste. Se simula utilizando Filtersim convencional e incorporando derivas y se contrasta con los resultados obtenidos mediante Simulación Secuencial Gaussiana.

Se generan 100 realizaciones para cada simulación. Al incorporar derivas sólo se analiza el caso con una influencia óptima (balanceada) entre la influencia de la deriva y del condicionamiento en la grilla de simulación. Este porcentaje de influencia se obtiene a partir del primer caso de estudio. La validación de los resultados se realiza mediante verificación visual (gráfica), verificación de continuidad espacial (variografía) al comparar el variograma de la imagen de entrenamiento, los pozos de tronadura, los sondajes de exploración y las realizaciones, y finalmente un análisis del sesgo tanto en las medias como en las varianzas obtenidas.

El análisis y métricas sobre los resultados son: Error Cuadrático Medio entre los valores simulados y los reales (pozos de tronadura), las derivas en la media en dirección Este-Oeste, mapas de probabilidad sobre umbral fijado mediante Ley de Corte, y curvas Tonelaje-Ley contrastando los resultados obtenidos con la realidad (pozos de tronadura). En todos los casos se mide la calidad de los resultados dada su cercanía al comportamiento real observado en los pozos de tronadura.

Capítulo 4

FILTERSIM INCORPORANDO DERIVAS

4.1. Formas de integrar deriva

La integración de la deriva en la media local dentro del algoritmo Filtersim puede ser implementada en más de una forma, sin embargo, se debe definir cuál (a criterio del autor) es la más adecuada. Dado los procesos que lleva a cabo el algoritmo las posibles formas de integrar la media son: modificación de la función distancia *evento condicionante* \leftrightarrow *prototipo*, nuevo filtro (media local) que modifica la representación vectorial previo agrupamiento de patrones en el proceso de *clustering*, y finalmente una combinación de ambas agregando un nuevo filtro sin modificación al proceso de *clustering* pero que modifica la función distancia *evento condicionante* \leftrightarrow *prototipo* basado en una imposición externa de la media local esperada (deriva impuesta).

4.1.1. Modificación función distancia evento condicionante \leftrightarrow prototipo

Se busca mantener las metodologías de agrupamiento, considerando la aplicación de los 3 filtros (para el cálculo del promedio, gradiente y curvatura) en las principales direcciones del espacio (2 en $2D$ y 3 en $3D$), pero el cambio sustancial se realiza en la determinación del prototipo más cercano a la condición de evento. Se agrega un ponderador, w_0 , 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.

Basado en los resultados expuestos por Ávalos [24] se decide utilizar la metodología *template pixels* como función para determinar la distancia al prototipo más cercano. Dado lo anterior, la ecuación que define la incorporación de la media local viene dada por (4.1).

$$D_l(u_0) = \left(\sum_{d=1}^3 w(d) \sum_{i=1}^{n_d} \left(dat_d(u_0 + u_i) - prot^{(l)}(u) \right) / n_i \right) (1 - w_0) + w_0 \left(\left| dat_m - prot_m^{(l)} \right| \right), \text{ con } w_0 \in [0, 1] \quad (4.1)$$

Donde:

- $D_l(u_0)$: distancia del evento condicionante centrado en u_0 al prototipo l .
- n_d : número de nodos informados con categoría d .
- u_i : vector entre el nodo central u_0 y el nodo en análisis.
- $w(d)$: ponderador al tipo de datos d .

- $dat_d(u_0)$: valor del dato del tipo d ubicado en $(u_0 + u_i)$ en el evento condicionante.
- $prot^{(l)}(u)$: valor del nodo ubicado en u_i en el prototipo l .
- dat_m : valor de la media local en el evento condicionante.
- $prot_m^{(l)}$: valor de la media local del prototipo l .

4.1.2. Nuevo filtro y modificación del proceso de clustering

Esta alternativa incorpora la media local en los patrones iniciales de Filtersim, es decir, agrega un nuevo filtro en el proceso de *clustering*. Este nuevo filtro es directamente la media de los datos dentro de la grilla de búsqueda aplicada sobre la imagen de entrenamiento, en otras palabras, es el promedio de los valores al interior de la grilla de búsqueda.

Recordando que la grilla de búsqueda tiene un tamaño de $(2n + 1) \times (2n + 1)$, el nuevo filtro viene dado por:

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

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 incluye más de una vez el mismo filtro recién expuesto.

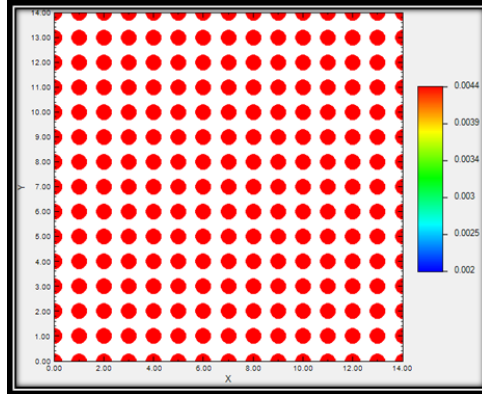


Figura 4.1: Filtro media local - Ejemplo en grilla de búsqueda de 15 x 15

4.1.3. Nuevo filtro e imposición de una deriva externa en la media local

Esta técnica toma ciertas ideas de ambas técnicas anteriores. Se propone utilizar el mismo nuevo filtro que pondera la media de los datos dentro de cada patrón inferido desde la T.I. pero cuyo resultado es utilizado en la modificación de la función distancia al prototipo más cercano. La modificación de la función distancia es idéntica a la expuesta en (4.1) salvo que dat_m no es el promedio de los datos del evento condicionante sino un valor de media local impuesto externamente por el usuario.

A criterio del autor esta última alternativa será la utilizada dado que entrega una flexibilidad al usuario de imponer diversas derivas sin necesidad de intervenir en el proceso interno del algoritmo modificado.

La implementación detallada de Filtersim en Python se encuentra en el Anexo A.

Capítulo 5

APLICACIONES Y ANALISIS

La presente sección tiene por objetivo aplicar el algoritmo de Filtersim modificado a dos casos de estudio. A través de estas aplicaciones se valida el algoritmo y a la vez analizan las mejoras/desventajas de los resultados al incorporar derivas en la media local. El primer caso de estudio tiene un carácter más bien simple utilizando la imagen de un corte pulido de roca sedimentaria. El segundo caso de estudio tiene un carácter más complejo siendo aplicado sobre un sector específico del yacimiento Cerro Colorado, propiedad de BHP Billiton.

5.1. Primer Caso de estudio – Corte roca sedimentaria

5.1.1. Presentación del caso de estudio

El presente caso de estudio tiene relación con la continuidad espacial de los estratos de una roca sedimentaria asociados a su color en escala de grises (0 a 255) y luego normalizado. La variable regionalizada en estudio es el valor de este elemento. Para ello se cortó y pulió la muestra para finalmente ser digitalizada. Se cuenta con una imagen de entrenamiento (figura 5.1) en la que se conoce el valor del elemento a lo largo de un plano XY y cuyo valor fluctúa entre $[0 - 0.897]$. La inferencia del comportamiento espacial (obtención de patrones) de la variable regionalizada se hace sobre esta imagen de entrenamiento. Por otro lado, se desea simular la continuidad espacial del elemento en otro sector de la roca. Este sector es denominado Ground Truth (figura 5.1), el cual está completamente informado para la variable en análisis. Lo interesante, y que guarda relación con la metodología de trabajo, es que a partir del Ground Truth se extraen los datos condicionantes que son luego migrados a la grilla a simular con la misma disposición espacial entre ellos.

Una de las principales dificultades de los algoritmos multipuntos es definir una imagen de entrenamiento adecuada. Sin embargo, definir el cambio sistemático en la media local a imponer sobre el modelo resulta igualmente delicado en relación a los resultados. Dado que se tiene conocimiento del sector a simular, Ground Truth, se decide tener un tipo de deriva en la media denominada *impuesta*, la cual proviene de ésta. Por otro lado se tienen derivas denominadas *inferidas*, las cuales vienen directamente del conjunto de datos condicionantes asociados a la grilla a simular.

Respecto a la cantidad de datos condicionantes, el número de repeticiones, el porcentaje de influencia de la deriva en la media, entre otras variables, estas son detalladas en la sección "*Metodología*".

Imagen de entrenamiento y Ground Truth

Tal como fue explicado con anterioridad, la imagen de entrenamiento (T.I.) será la responsable de entregar la información sobre el comportamiento espacial que presenta el elemento de interés, el cual es llevado a una escala de grises en el intervalo $[0 - 255]$ y normalizado a $[0-1]$ y desplegada luego a color. En este caso de estudio se cuenta con una T.I. de 240×250 píxeles en la cual se observa continuidad espacial solo de las zonas de valores altos (mayores a 0.8, ver histograma en figura 5.2.a) sumergidos en una matriz de valores bajo los 0.8. Se observa además que el ancho de las estructuras de valores altos (color rojo en la figura 5.1) tiene un promedio aproximado de unos 5 píxeles siendo de 1 píxel el sector más delgado y unos 11 píxeles la parte más gruesa. A pesar de que el variograma no logra caracterizar bien estructuras sinusoidales es posible aplicarlo con un sentido de cuantificación, es decir, para obtener una métrica de cuán rápido se descorrelaciona la variable en estudio. En efecto, se observa en la figura 5.2.b los variogramas en las dos direcciones principales: en la dirección $N90^\circ E$ la variable se desestructura extremadamente rápido teniendo un alcance de 5 a 6 píxeles para asumir que ya pares de puntos ubicados a una distancia mayor no tienen relación significativa. Sin embargo, y como es de esperar, al analizar la dirección perpendicular $N0^\circ E$ se observa una mayor continuidad espacial entregando un alcance de 20 píxeles. De lo anterior se desprende que la variable en estudio presenta una anisotropía geométrica Norte-Sur (En el eje Y).

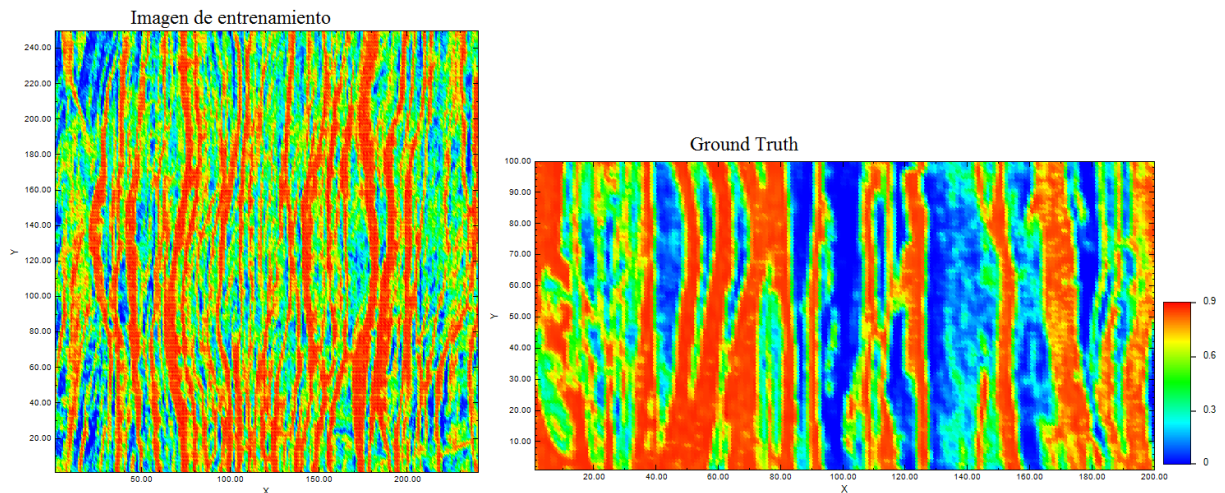


Figura 5.1: Imagen de entrenamiento (izq.) y Ground Truth (der.) – Roca sedimentaria

Respecto a las derivas en la media si estas se analizan a lo largo del eje Y para cada valor en X, tanto en la Imagen de Entrenamiento como en el Ground Truth, se observa claramente la presencia de derivas a lo largo del eje X. En efecto, se observa en la figura 5.3 que el rango de valores para la T.I. va entre 0.26 y 0.79 mientras que para el G.T. va entre 0.02 y 0.87. Por otro lado y tal como se observa en la figura 5.4 no sucede lo mismo si se analiza la media a lo largo del eje X para cada valor en Y. En efecto, para la T.I. el valor de la media va entre 0.43 y 0.62 mientras que en el G.T. el valor fluctúa entre 0.43 y 0.59. Por lo tanto, para la definición posterior de las medias locales impuestas, éstas sólo tendrán derivas a lo largo del eje X.

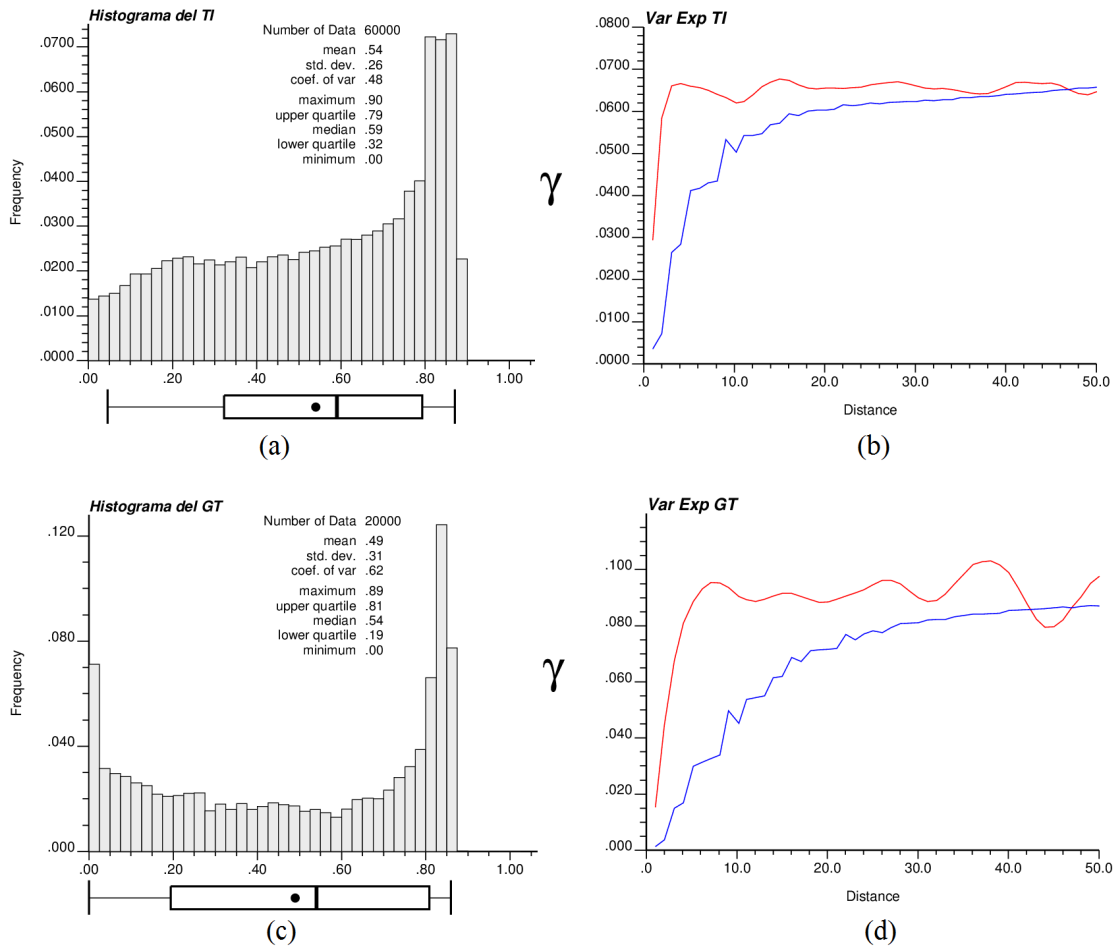


Figura 5.2: (a) Histograma de la TI, (b) Variograma experimental de la TI, (c) Histograma del GT, (d) Variograma experimental del GT

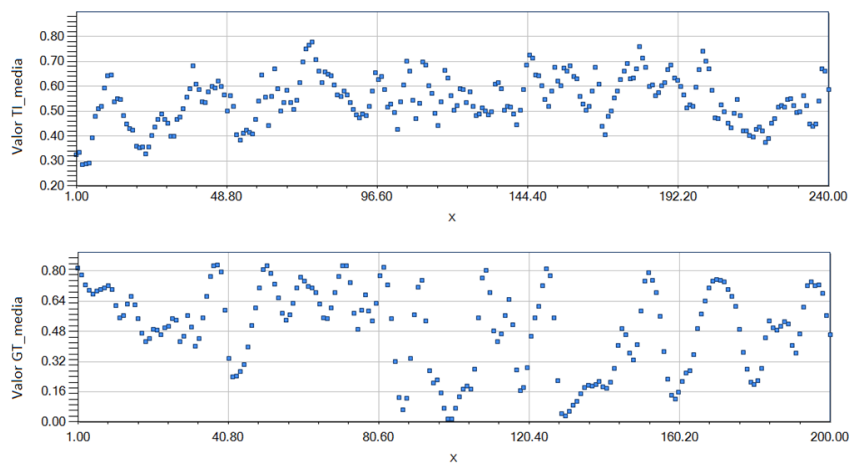


Figura 5.3: Gráfica del valor medio a lo largo del eje Y para cada valor de X en: (arriba) Imagen de entrenamiento y (abajo) Ground truth

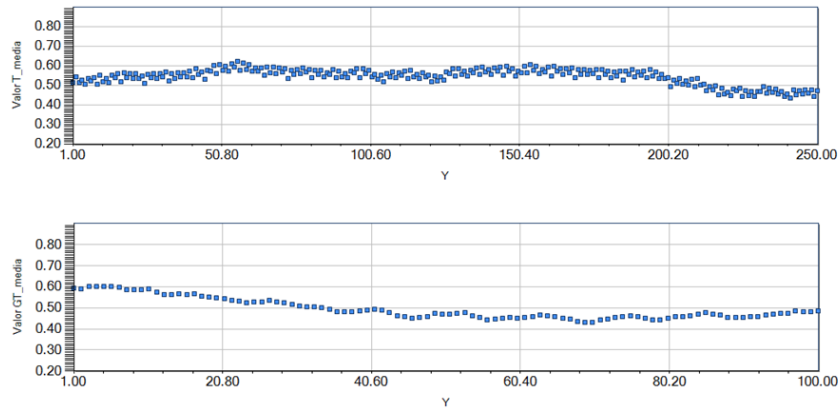


Figura 5.4: Gráfica del valor medio a lo largo del eje X para cada valor de Y en: (arriba) Imagen de entrenamiento y (abajo) Ground truth

5.1.2. Metodología

Dado que el objetivo del presente caso de estudio es cuantificar la mejora en la caracterización de las estructuras espaciales al incorporar derivas en la media local se propone la siguiente metodología (ver figura 5.5).

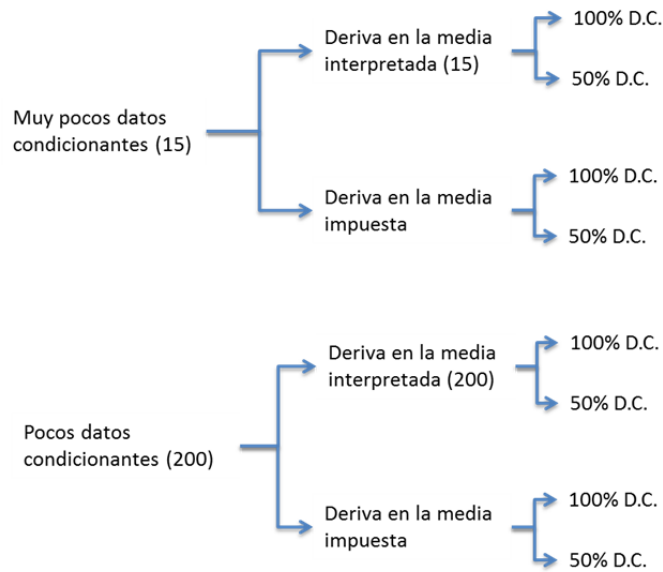


Figura 5.5: Esquema de trabajo - Caso de estudio – Roca sedimentaria

1. Sobre el Ground Truth se extraen aleatoriamente dos conjuntos de datos. El primero contiene 15 datos condicionantes (figura 5.6.a) y el segundo 200 (figura 5.6.c). Al primero se le denomina set con “Muy pocos datos condicionantes” y al segundo “Pocos datos condicionantes”.
2. Una deriva en la media local es inferida desde el Ground Truth denominada deriva en la media impuesta (figura 5.7.f). Esta se obtiene al promediar todos los valores en Y para cada posición en X y luego ese valor es asignado a lo largo de toda la columna en Y.
3. Derivas en la media local son inferidas a partir de los datos condicionantes. Estas derivas son denominadas derivas interpretadas. Para obtenerlas se aplica el mismo procedimiento descrito en el paso 2, pero dado que hay posiciones en X que no poseen ningún valor a lo largo del eje Y, se debe asumir un ancho de banda para la obtención de estos promedios.

El ancho de banda corresponde a la separación máxima entre dos datos condicionantes a lo largo del eje X con la finalidad de asegurar al menos un dato para cada posición en X. En el caso de Muy pocos datos condicionantes se asume un ancho de banda igual a 25 unidades, es decir, ± 13 unidades a partir de la posición en X (figura 5.7.b). Para el caso de Pocos datos condicionantes se asume un ancho de banda de 5 unidades, por lo que cada banda tiene asociado un rango de ± 3 centrado en la posición en X.

4. Sobre cada conjunto de datos condicionantes se simula utilizando como deriva en la media tanto la media impuesta como la media interpretada.
5. Para cada combinación de “Datos condicionantes – Deriva en la media” se simula utilizando el algoritmo Filtersim (implementado y modificado en Python).
6. Con la finalidad de cuantificar la importancia de los datos condicionantes se decide repetir las simulaciones. La única diferencia es que se utiliza sólo el 50 por ciento de los datos condicionantes elegidos aleatoriamente. Para el caso de Muy pocos datos condicionantes se obtienen 7 datos finales (figura 5.6.b) y para el caso de Pocos datos condicionantes se obtienen 100 datos finales (figura 5.6.d).
7. La validación de las simulaciones se realiza mediante estudio de variogramas e histogramas (en forma de diagramas de caja). Se verifica que se respetan las distribuciones y estadísticas globales. Dado que el variograma no caracteriza adecuadamente estructuras sinusoidales se añade una validación visual al proceso.
8. Se analizan los resultados a través de una comparación de Errores Cuadráticos Medios (entre las derivas externas y resultantes) a nivel tanto global como sobre un umbral de corte (valor 0.63). Se despliegan los mapas de probabilidad asociados al mismo umbral y se analizan los porcentajes de aciertos sobre el valor esperado. Se comparan las derivas en la media entre la T.I., el G.T. y el promedio de las 100 realizaciones.
9. Se comparan los resultados con los que se obtendrían al utilizar estimación por Kriging (Simple y Ordinario) y simulación por Bandas Rotantes.

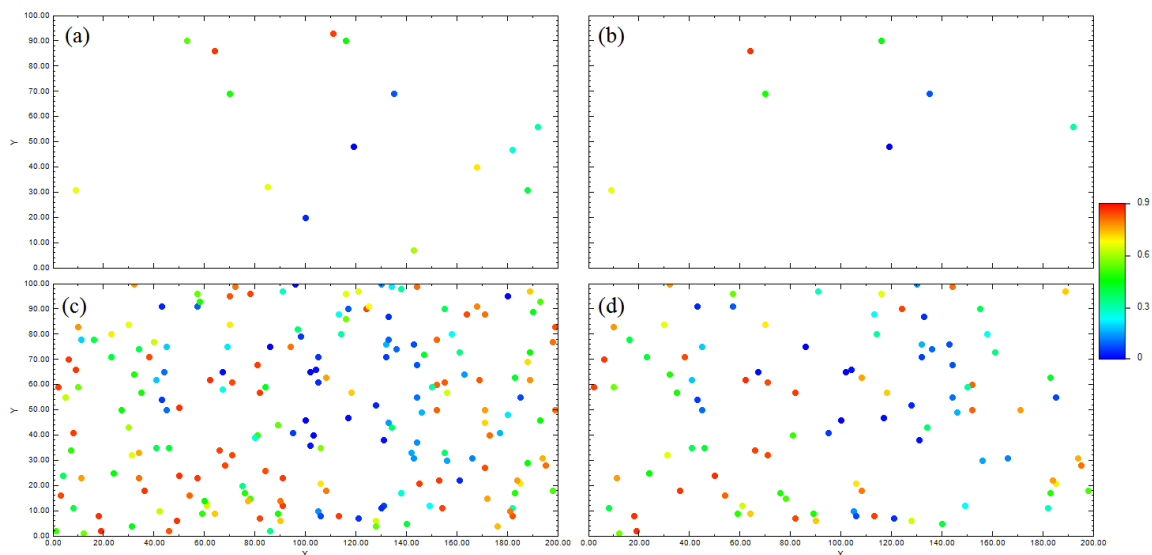


Figura 5.6: Datos condicionantes. (a) 100 % Muy pocos D.C. (b) 50 % Muy pocos D.C. (c) 100 % Pocos D.C. (d) 50 % Pocos D.C.

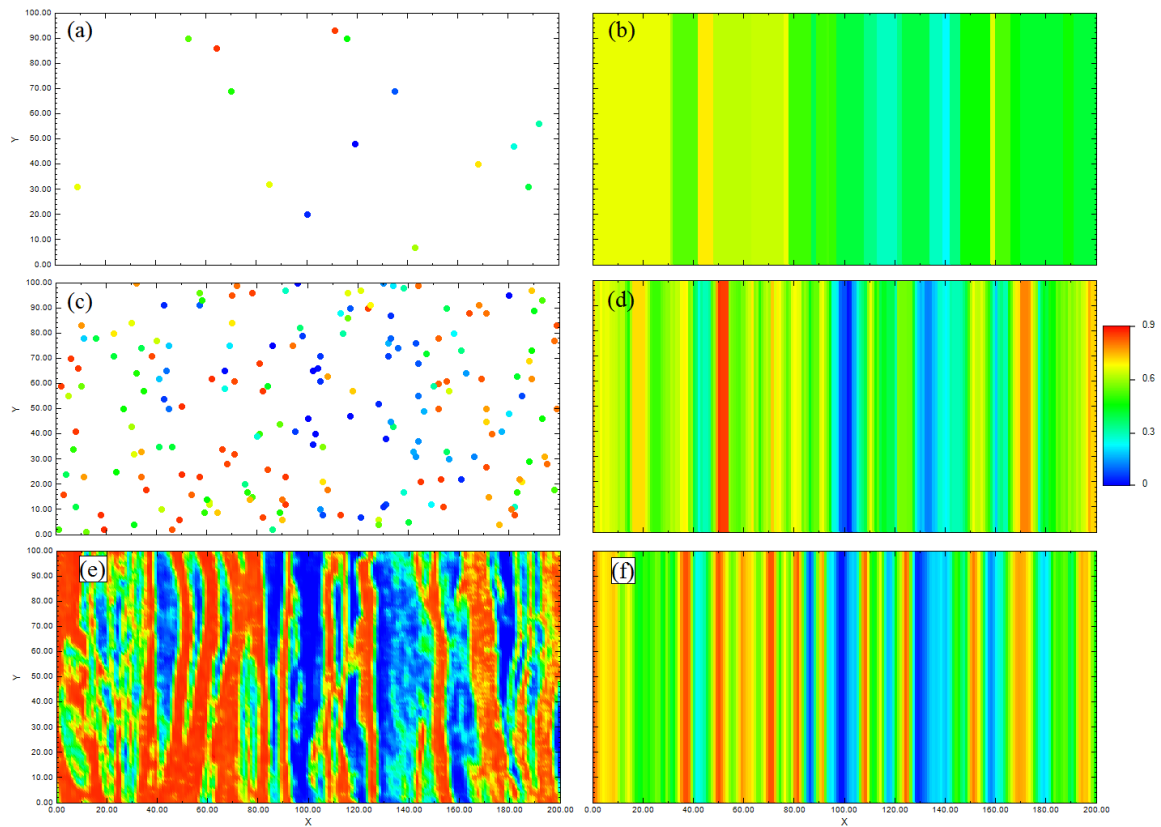


Figura 5.7: Medias inferidas. (a) Muy Pocos D.C. (b) Media inferida desde Muy Pocos D.C. (c) Pocos D.C. (d) Media inferida desde Pocos D.C. (e) Ground Truth. (f) Media inferida desde Ground Truth

Respecto a los datos condicionantes, la tabla 5.1 presenta las estadísticas básicas de los datos condicionantes.

Tabla 5.1: Estadísticas básicas de los datos condicionantes

Parámetros	15 DC (50%)	15 DC (100%)	200 DC (50%)	200 DC (100%)
Cantidad	7	15	100	200
Mínimo	0	0	0	0
Media	0.40	0.45	0.45	0.49
Máximo	0.85	0.85	0.86	0.86
Varianza	0.08	0.07	0.09	0.09

5.1.3. Parámetros de simulación

Habiendo expuesto la metodología de la simulación se procede a detallar los parámetros de entrada al algoritmo para llevarla a cabo. Estos parámetros guardan relación con la grilla de búsqueda, la grilla interna, el nivel de multigrillas, la ponderación de las derivas en la media, entre otros factores. La tabla 5.2 presenta en resumen los parámetros y sus valores. La definición de estos se fundamenta en las conclusiones expuestas por Ávalos, S. en su trabajo de título

Tabla 5.2: Parámetros de entrada - Caso de estudio simple

Parámetros	Valor
Número de realizaciones	100
Promedio sobre las repeticiones	Si
Grilla de búsqueda eje X	9
Grilla de búsqueda eje Y	9
Grilla interna: eje X	9
Grilla interna: eje Y	9
Número de multigrillas	2
Mínima cantidad de D.C.	5 – 1
Ponderador de deriva en la media, w_0	[0.0, 0.2, 0.4, 0.6, 0.8, 1.0]
Ponderador datos condicionantes, w_1	0.6
Ponderador datos simulados, w_2	0.25
Ponderador datos previamente simulados, w_3	0.15
Número de clusters	200

Desglosando la tabla 5.2 se observa que para cada combinación “Datos condicionantes – Deriva en la media” los parámetros fijos son: el tamaño de la grilla de búsqueda e interna, los ponderadores w_1 , w_2 y w_3 , el número de multigrillas, el número de clusters y la cantidad mínima de datos en los eventos condicionantes (5 para *pocos datos condicionantes* y 1 para *muy pocos datos condicionantes*). El único parámetro que varía es la ponderación, w_0 , asociada a la deriva en la media (ver ecuación (7.1)) el cual toma 6 posibles valores 0.0, 0.2, 0.4, 0.6, 0.8 o 1.0. Finalmente, para cada combinación se generan 100 realizaciones obteniendo además un promedio sobre las mismas. Cabe recordar que los datos condicionantes son considerados tanto en su 100 por ciento como en un 50 por ciento de ellos.

5.1.4. Resultados - Visualización, estadísticas básicas y variografía

Los resultados principales son las realizaciones. Dado que la cantidad de combinaciones de parámetros y conjuntos de datos condicionantes llega a 48 (4 conjuntos de datos condicionantes, 6 valores de w_0 y 2 derivas externas) y cada combinación cuenta con 100 realizaciones se decide presentar los resultados en la sección **Anexo B**. En ella se visualizan 3 realizaciones, escogidas de manera aleatoria, y el promedio respectivo sobre las 100 realizaciones. La presente sección presenta una sola realización y el promedio asociado para cada simulación. Se presenta el caso de 200 datos condicionantes al 100 % incorporando derivas interpretadas e impuestas, siendo analizados los 6 posibles valores de w_0 .

La motivación de presentar los resultados es la entrega de información para una adecuada validación de los modelos que generan las realizaciones. La validación tiene varias aristas. La más rápida pero no por eso menos importante es la validación gráfica o visual. En ella se busca comprobar que las estructuras (sectores de valores altos) presenten una continuidad espacial marcada y respeten los comportamientos sinusoidales. Una validación más detallada se logra al extraer los variogramas sobre cada realización para establecer cuanto se logra caracterizar la correlación espacial entre dos puntos del espacio simulado. En otras palabras, se busca validar que las realizaciones provienen de funciones aleatorias ergódicas, entendiéndose por ello que las estadísticas experimentales convergen a las estadísticas del modelo teórico cuando estas son calculadas sobre un dominio suficientemente grande. Dado que el dominio está restringido, en este caso 200 x 100 píxeles, las estadísticas presentan fluctuaciones respecto al modelo teórico. En particular, al representar de manera conjunta los variogramas de cada realización se tiene la amplitud en el origen, en la zona media y en el alcance representando así las fluctuaciones antes descritas. Dado el gran número de simulaciones los variogramas asociados a cada una son expuestos en la sección **Anexo C**. Se añaden los variogramas tanto de la deriva en la media como de la Imagen de Entrenamiento y el Ground Truth para lograr las respectivas comparaciones y análisis. En lo que respecta a la presente sección se presentan los análisis variográficos sobre 20 de las 100 realizaciones por simulación dado que un número mayor ensucia la representación gráfica. La última forma de validar es a través de las estadísticas básicas. Para ello se calculan las estadísticas básicas para cada realización y luego se obtiene el valor promedio sobre las 100 realizaciones de cada estadística dada las fluctuaciones presentes en las realizaciones [8]. Dadas las mismas razones antes expuestas, las estadísticas básicas promedio son presentadas en la sección **Anexo D**. En la presente sección se presentan las estadísticas básicas promedio (sobre las 100 realizaciones).

Pocos datos condicionantes (100 %) – Deriva interpretada

En la figura 5.8, se presentan tanto el promedio de las realizaciones como una de ellas escogida de manera aleatoria. En este caso se utiliza una deriva externa en la media interpretada a partir del conjunto de 200 datos condicionantes, figura 5.7.d.

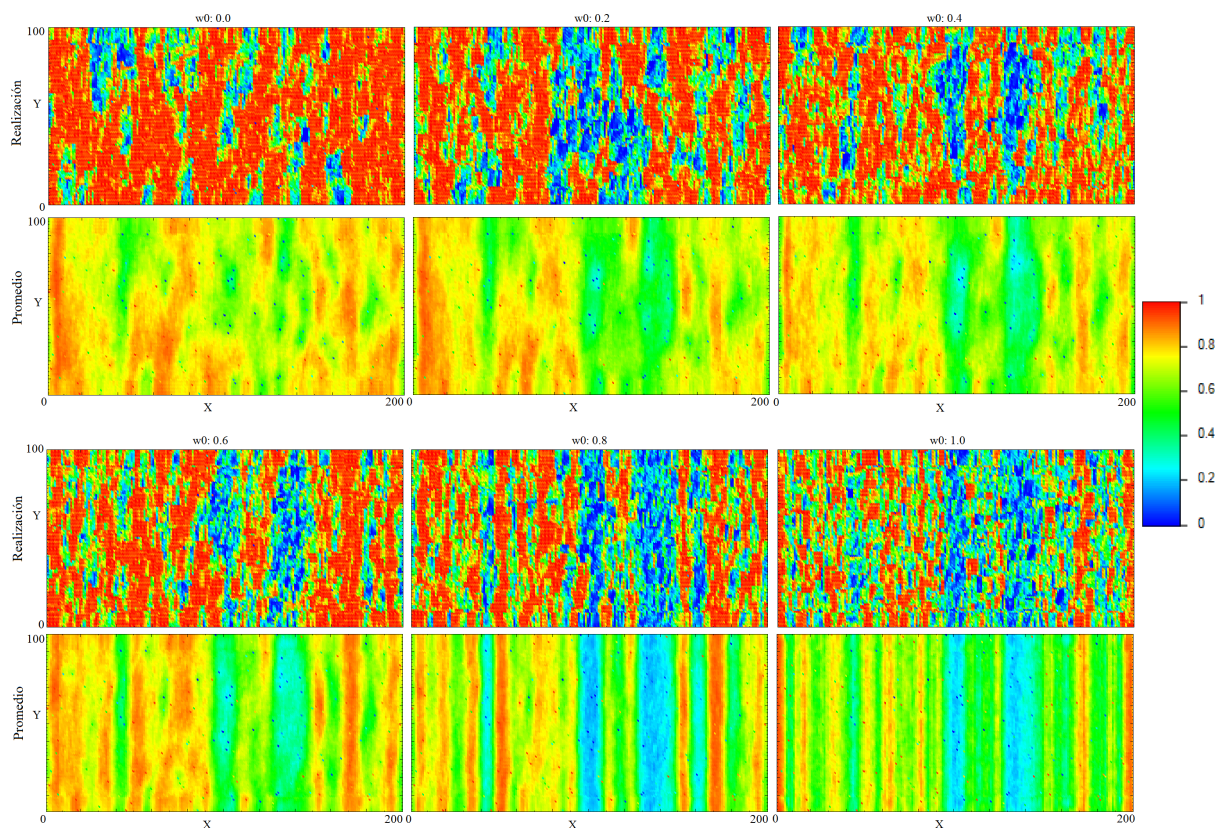


Figura 5.8: Una realización (arriba) de cien. Promedio de las 100 realizaciones (abajo). Pocos datos condicionantes – Deriva interpretada

La tabla 5.3 presenta las estadísticas básicas de la Imagen de Entrenamiento, el Ground Truth y la Deriva externa en la media local. Además se presentan las estadísticas promedio sobre las 100 realizaciones para los casos w_0 0.0, 0.2, 0.4, 0.6, 0.8 y 1.0.

Tabla 5.3: Estadísticas básicas. Pocos datos condicionantes – Deriva interpretada

Indicador	T.I.	G.T.	Deriva Media	w0:0.0	w0:0.2	w0:0.4	w0:0.6	w0:0.8	w0:1.0
Promedio	0.544	0.494	0.504	0.629	0.587	0.589	0.598	0.530	0.516
Mínimo	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000
Primer cuartil	0.322	0.194	0.383	0.415	0.329	0.352	0.377	0.245	0.250
Mediana	0.591	0.541	0.552	0.782	0.713	0.696	0.706	0.594	0.557
Tercer cuartil	0.793	0.808	0.631	0.849	0.843	0.833	0.835	0.818	0.798
Máximo	0.897	0.888	0.858	0.897	0.897	0.897	0.897	0.897	0.897
Varianza	0.069	0.094	0.033	0.080	0.084	0.077	0.075	0.087	0.080

La variografía es aplicada sobre la Imagen de Entrenamiento (**azul**), el Ground Truth (**negro**), la deriva externa en la media (**verde**) y sobre 20 realizaciones (**rojo**) aleatoriamente escogidas. Se utilizan 50 pasos de largo igual a 1 analizados en las direcciones N0°E y N90°E. Ancho de banda horizontal y vertical igual a 20. Los resultados son expuestos en la figura 5.9.

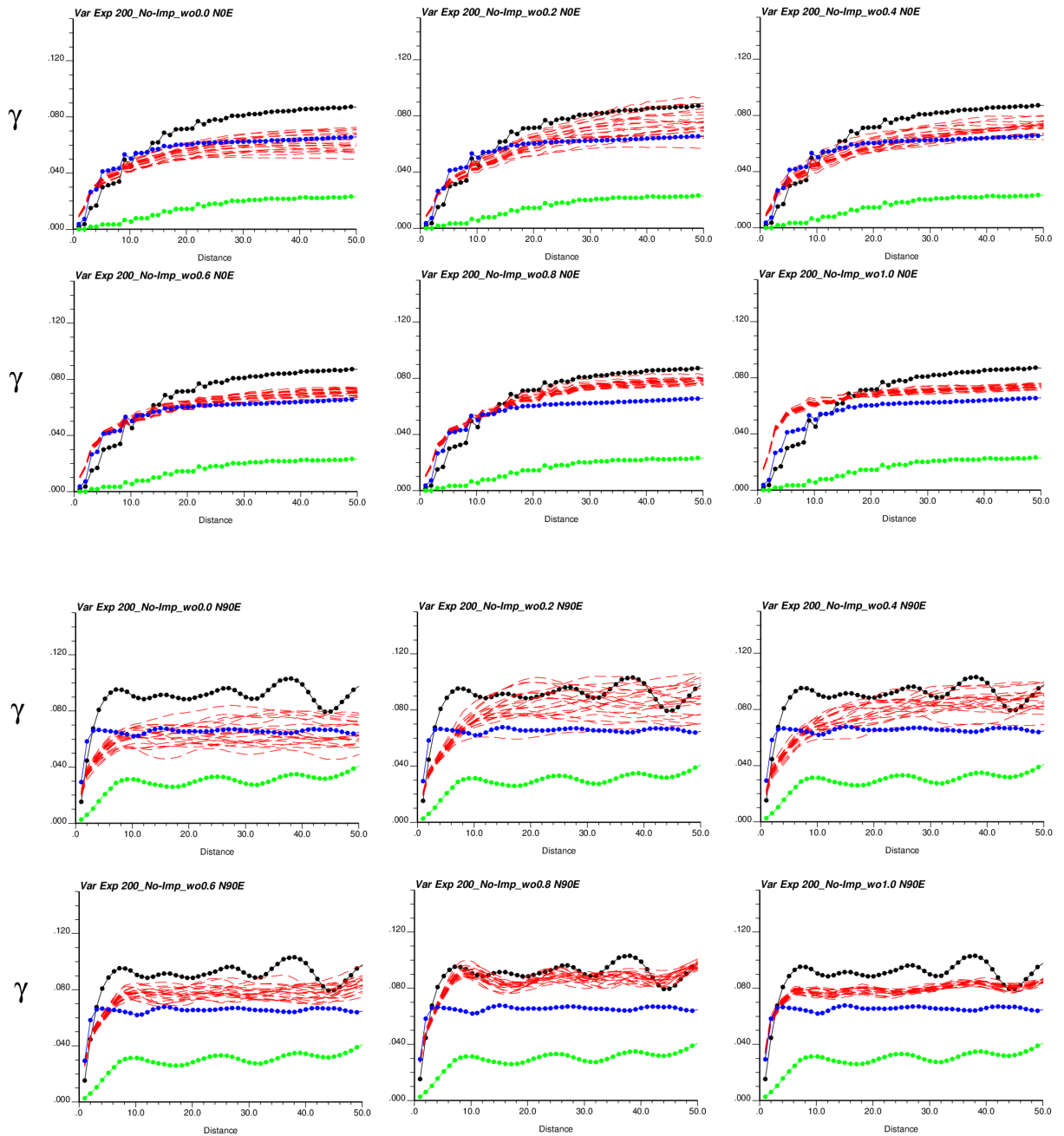


Figura 5.9: Variografía en dirección N0E y N90E. Pocos datos condicionantes – Deriva interpretada

El análisis tanto de la visualización 2D de las realizaciones, las estadísticas básicas y la variografía se desarrolla en la sección siguiente.

Pocos datos condicionantes (100%) – Deriva impuesta

La figura 5.10 presenta tanto el promedio de las realizaciones como una de ellas escogida de manera aleatoria. En este caso se utiliza una deriva externa en la media impuesta a partir de la información entregada en el Ground Truth, figura 5.7.f.

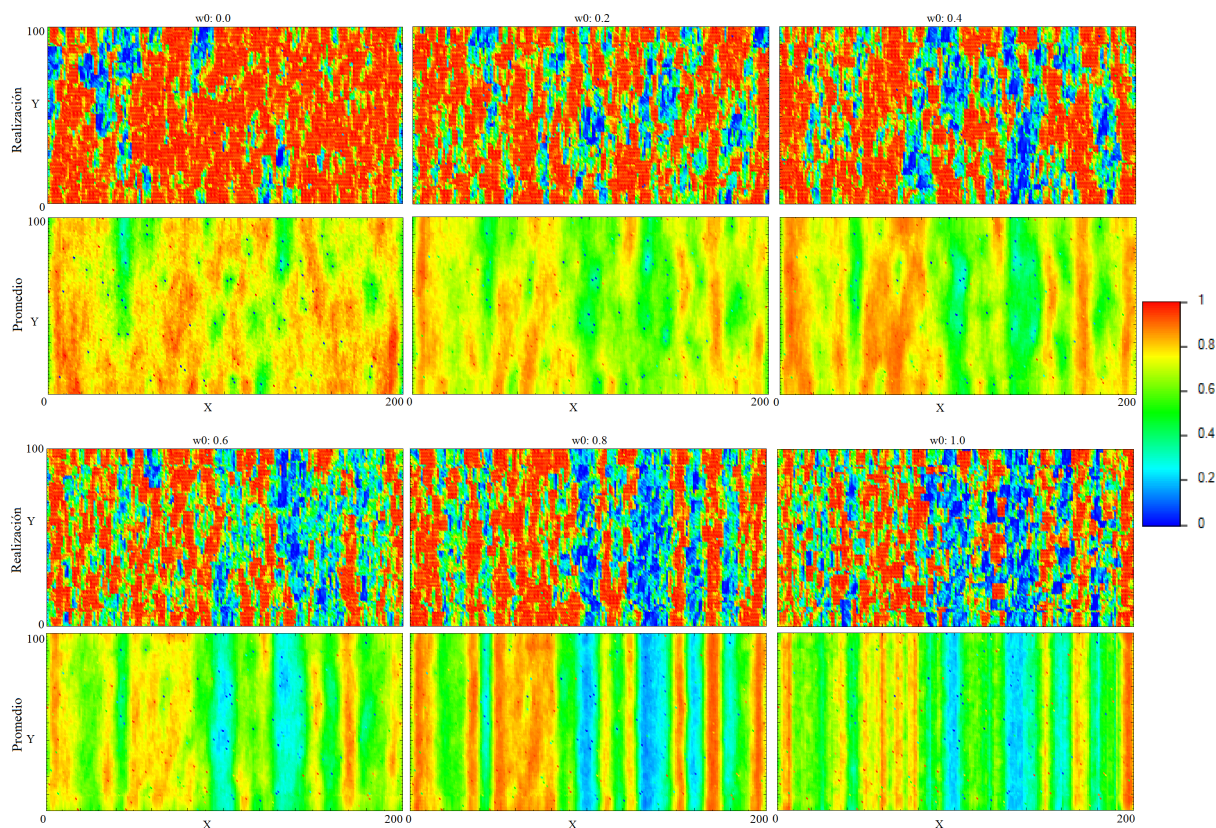


Figura 5.10: Una realización (arriba) de cien. Promedio de las 100 realizaciones (abajo). Pocos datos condicionantes – Deriva impuesta

La tabla 5.4 presenta las estadísticas básicas de la Imagen de Entrenamiento, el Ground Truth y la Deriva externa en la media local. Además se presentan las estadísticas promedio sobre las 100 realizaciones para los casos w_0 0.0, 0.2, 0.4, 0.6, 0.8 y 1.0.

Tabla 5.4: Estadísticas básicas. Pocos datos condicionantes – Deriva impuesta

Indicador	T.I.	G.T.	Deriva Media	w0:0.0	w0:0.2	w0:0.4	w0:0.6	w0:0.8	w0:1.0
Promedio	0.544	0.494	0.494	0.623	0.601	0.659	0.573	0.536	0.504
Mínimo	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000
Primer cuartil	0.322	0.194	0.286	0.440	0.365	0.507	0.332	0.239	0.211
Mediana	0.591	0.541	0.538	0.733	0.743	0.799	0.658	0.624	0.549
Tercer cuartil	0.793	0.808	0.690	0.837	0.843	0.849	0.827	0.826	0.809
Máximo	0.897	0.888	0.831	0.897	0.897	0.897	0.897	0.897	0.897
Varianza	0.069	0.094	0.049	0.067	0.083	0.065	0.077	0.092	0.091

La variografía es aplicada sobre la Imagen de Entrenamiento (**azul**), el Ground Truth (**negro**), la deriva externa en la media (**verde**) y sobre 20 realizaciones (**rojo**) aleatoriamente escogidas. Se utilizan 50 pasos de largo igual a 1 analizados en las direcciones N0°E y N90°E. Ancho de banda horizontal y vertical igual a 20. Los resultados son expuestos en la figura 5.11.

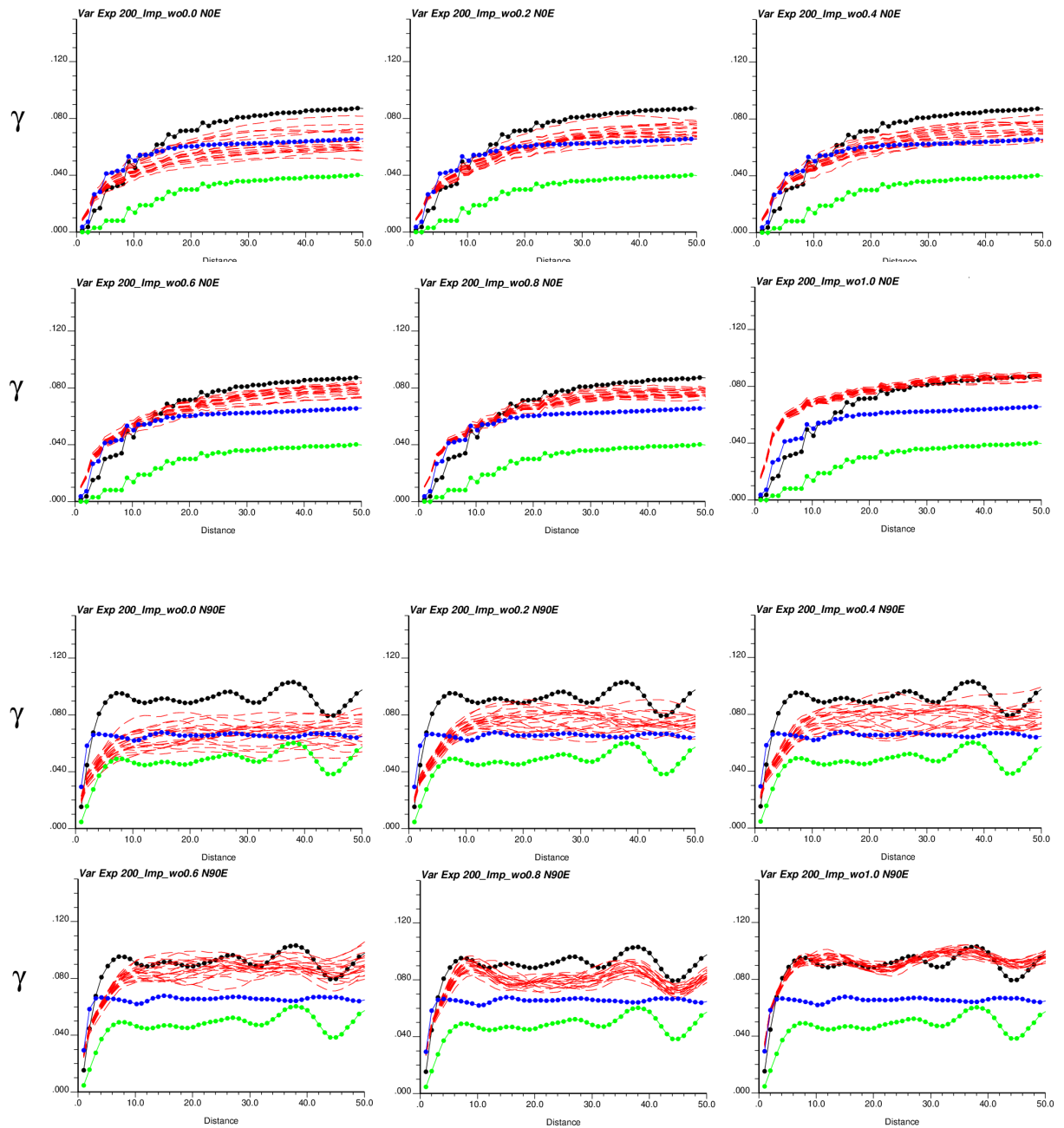


Figura 5.11: Variografía en dirección N0E y N90E. Pocos datos condicionantes – Deriva impuesta

El análisis tanto de la visualización 2D de las realizaciones, las estadísticas básicas y la variografía se desarrolla en la sección siguiente.

5.1.5. Validación y análisis de resultados

En la sección anterior se presentó de manera completa el caso *Pocos datos condicionantes*. La presente sección considera estos resultados más los expuestos en los diferentes anexos para los casos *Muy pocos datos condicionantes* los cuales son procesados, analizados y comentados. Un primer análisis se realiza directamente sobre las realizaciones de manera visual. Luego se comparan estos resultados contra los obtenidos mediante métodos convencionales. Se obtienen los Errores Cuadráticos Medios (globales y sobre un cierto umbral) como medida de dispersión de las realizaciones y el Ground Truth. Se comparan las derivas en la media entre las realizaciones, el Ground Truth y la deriva interpretada/impuesta influyente. Se analizan los mapas de probabilidad para el mismo umbral definido en el error cuadrático medio. Se comparan los porcentajes de acierto a distintos w_0 y finalmente se comparan los resultados respecto a la estimación por Kriging y simulación por Bandas Rotantes.

Realizaciones y promedio

Se aplica un análisis sobre la información entregada por la validación de los resultados en sus tres aristas: visual, variográfica y estadística. Dado que varios son los parámetros en combinación se analiza la repercusión de variar cada uno de ellos dejando fijo todo el resto (*ceteris paribus*).

Análisis visual

Al analizar solo el conjunto de *pocos datos condicionantes* y fijando w_0 en 0.0 se observa que los promedios locales obtenidos son similares entre si. Esto está en concordancia a lo esperado dado que al asumir w_0 : 0.0 las derivas no tienen influencia en la simulación. Además se observan sectores de valores altos en una matriz de valores medios sin presencia de estructuras sinusoidales de alto valor. Por otra parte, las realizaciones entregan estructuras continuas de valores altos cuyo ancho aproximado supera los 40 píxeles en contraste con los 5 y 14 píxeles de ancho promedio en la T.I. y el G.T. respectivamente. Recordando que se busca tener realizaciones “similares” a la del GT basado en la T.I., los resultados no son satisfactorios de forma visual. Al incorporar deriva en la media, es decir, al aumentar el valor de w_0 se generan realizaciones con estructuras (en valores altos) cuyo ancho promedio disminuye mejorando además la continuidad espacial en la vertical. Esto mejora la calidad en la reproducción de la estructura. Respecto a los promedios se observa, para w_0 : 0.4 al incorporar la deriva impuesta, una estructura delgada y continua en el sector izquierdo que no se aprecia al incorporar una deriva interpretada (ver figura 5.12). Esto se debe a la ausencia de valores bajos en ese sector que acote el ancho de la estructura. En las realizaciones se caracteriza de buena manera el sector medio de bajos valores tanto en la deriva impuesta como en la interpretada (ver figura 5.12).

Al fijar la influencia de la deriva en w_0 : 0.8 se aprecia tanto en las realizaciones como en los promedios estructuras delgadas y continuas de valores altos. Sin embargo, se pierde la propiedad sinusoidal de la variable regionalizada alejando a las realizaciones de reproducir la realidad. Del Anexo B se interpreta que una influencia entre 0.4 y 0.6 mejora considerablemente la reproducción de las estructuras. Al fijar w_0 : en 1.0 sólo se toma en consideración el aporte de la deriva influyente y no el aporte de los datos condicionantes, sin embargo, se reproduce el valor puntual de cada dato condicionante. Se obtienen resultados visualmente distorsionados ya que se ajustan a la deriva interpretada/impuesta y no a los datos condicionantes (a pesar de que tienen correlación ya que estos fueron sacados del GT o bien fueron generadores de la deriva interpretada). Se entiende por “distorsionados” al hecho de generar estructura netamente verticales con poca variabilidad en la dirección y en el ancho.

Un análisis similar se puede aplicar sobre el conjunto de *muy pocos datos condicionantes*. Resulta interesante la comparación de los resultados al fijar w_0 en 0.4 y deriva interpretada al utilizar 15 DC y 200 DC. Al pasar de 15DC (0.075 % del dominio informado) a 200DC (1 %

del dominio informado) la deriva interpretada se vuelve más rica en información (incluso similar a la impuesta) llevando al promedio de las realizaciones a poseer estructuras más delimitadas.

Una desventaja transversal a todas las configuraciones se visualiza en los pequeños quiebres de las estructuras al utilizar grillas de búsqueda/internas cuadradas. Esto queda en evidencia en las realizaciones pero se difumina al promediarlas.

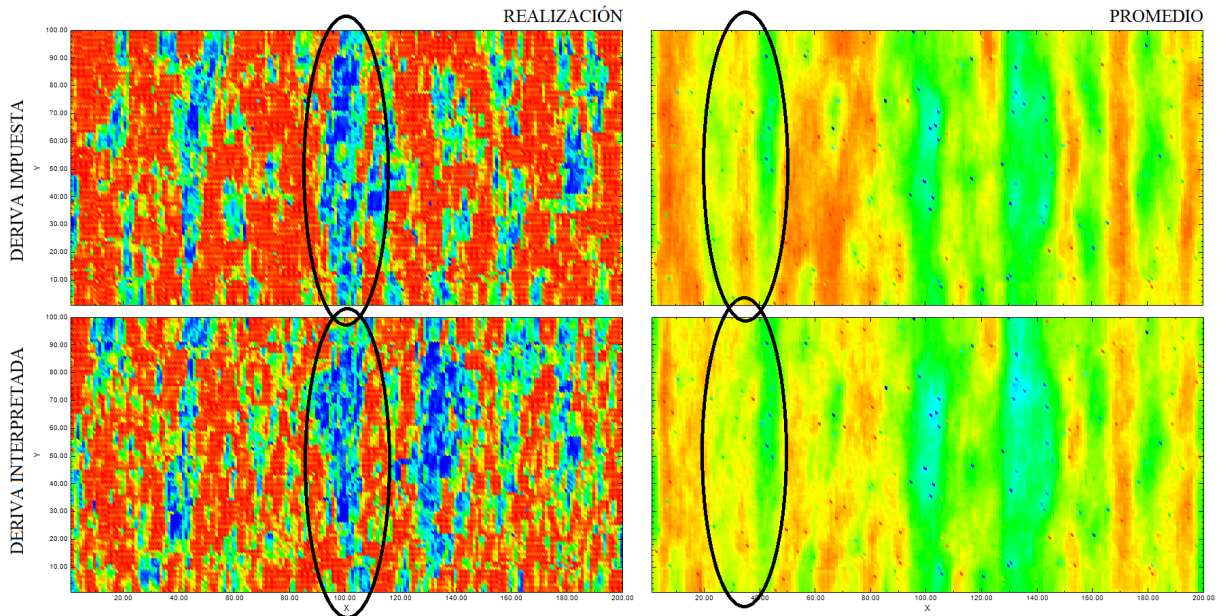


Figura 5.12: Comparación de realizaciones y promedios entre deriva impuesta e interpretada en el caso *pocos datos condicionantes*

Análisis variográfico

En los cuatro casos estudiados, y tanto en la dirección N0E como en la N90E, el variograma experimental de las realizaciones se aleja totalmente del comportamiento variográfico de la deriva impuesta. Lo anterior es un resultado esperado dado que la deriva es una función regular y de poca varianza y que el algoritmo no busca generar escenarios similares a la deriva si no más bien utilizar su información para enriquecer las realizaciones. En ambas direcciones de estudio y independiente de la cantidad de datos condicionantes y tipo de deriva externa se observa que al aumentar el porcentaje de influencia por parte de la deriva externa los variogramas experimentales de las realizaciones disminuyen la amplitud de las diferentes respuestas, independiente de la distancia de análisis. En otras palabras, se vuelve menos errático el variograma de las realizaciones.

Al aumentar la cantidad de datos condicionantes de un 0.075 % a un 1 % y sin considerar derivas en la media ($w_0 : 0.0$) los variogramas de las realizaciones se vuelven menos erráticos disminuyendo la amplitud entre ellos. Además, se verifica con ello la tendencia a la similitud del variograma de la imagen de entrenamiento. Lo anterior se observa para ambas direcciones principales (N0E y N90E). Si se aumenta la cantidad de datos condicionantes y la deriva externa se extrae de los mismos a una influencia entre el 60 % y el 80 % se observa claramente cómo los variogramas de las realizaciones se asemejan al variograma del Ground Truth.

Condicionando la simulación con un 1 % del dominio conocido (200 datos condicionantes) y una influencia del 80 % de la deriva externa ($w_0 : 0.8$), se tiene que los variogramas de las realizaciones se asemejan más al presentado por el Ground Truth al ser inferidos desde los datos condicionantes que directamente del Ground Truth. Esto último se debe a que al calcular la

media a lo largo del eje Y para cada coordenada en X, desde el Ground Truth, se promedian 100 valores acercando este valor a la media de los datos. Por su parte la deriva interpretada considera un número pequeño pero variable de datos entregando valores menos suavizados.

Resumiendo, la variografía de las realizaciones tiende a la variografía de la Imagen de entrenamiento y mientras más influencia se le asigna a la deriva externa más se asemejan al variograma del Ground Truth. De todas maneras, se valida variográficamente que las simulaciones representan de buena manera el comportamiento esperado en el Ground Truth.

Análisis estadístico

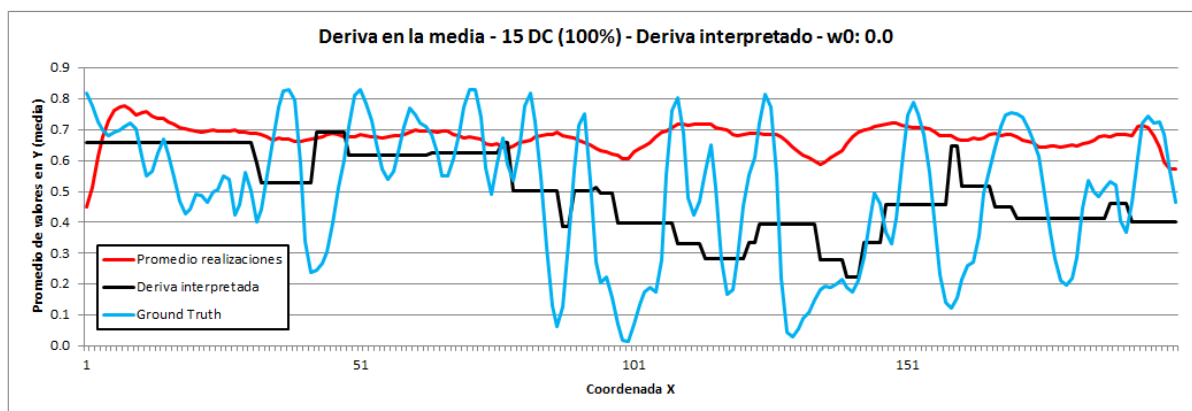
Las estadísticas expuestas en la sección anterior corresponden al valor promedio de los 100 valores obtenidos para cada métrica y no el valor de la métrica sobre el promedio de las realizaciones. Estadísticamente se tiene que los valores máximos y mínimos son los mismos que los presentes en la Imagen de entrenamiento. Los resultados de Filtersim sin incorporar derivas entregan un valor promedio por sobre el valor promedio de la T.I., del G.T. y de las derivas. Sin embargo, al aumentar la influencia de la deriva se disminuye el valor promedio hasta un valor similar a la media de la Imagen de entrenamiento. Situación similar se observa en el primer cuartil, que disminuye drásticamente al incorporar derivas. Esto último lleva a inferir que los valores simulados aumentan su varianza al aumentar el porcentaje de influencia. Por último, se observa en los diagramas de caja del Anexo D que al aumentar la influencia de la deriva externa (aumentar el w_0) las distribuciones globales de las realizaciones pasan de tender a la distribución de la Imagen de entrenamiento a la distribución del Ground Truth.

Derivas

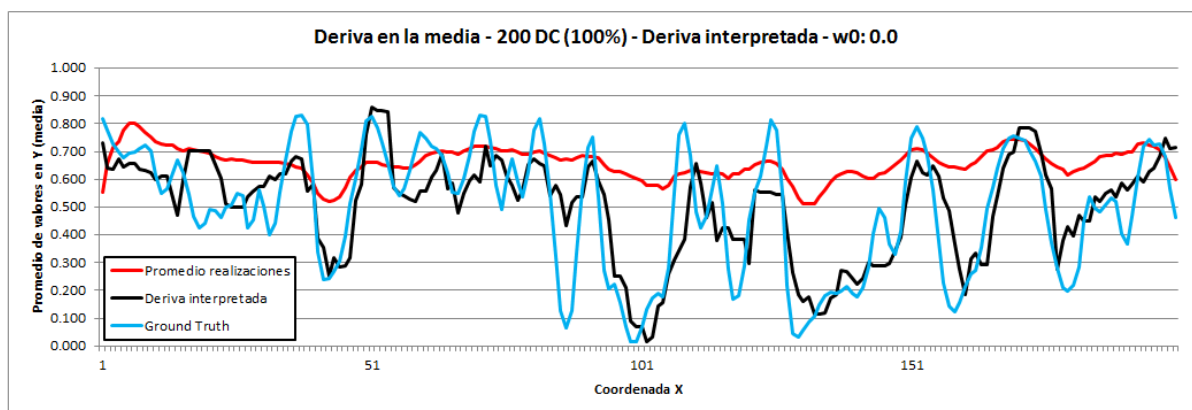
El cálculo de las derivas se realiza a lo largo del eje X promediando, para cada coordenada X, los valores a lo largo del eje Y. Se presenta el gráfico de las derivas en la sección de Anexos E. Se grafica en conjunto las derivas para el Ground Truth (azul), la deriva externa (negro) y el promedio de las 100 realizaciones (rojo).

Al analizar las realizaciones generadas por Filtersim sin incorporación de derivas se observa que:

- Al contar con muy pocos datos condicionantes (15 DC) las derivas obtenidas son altamente erráticas sin presentar un cambio sistemático claro, tal como se observa en la figura 5.13.a.
- Al contar con pocos datos condicionantes (200 DC) no se observa una mejora en las derivas resultantes, salvo coordenadas específicas en las que datos condicionantes de bajo valor llevan a derivas de bajo valor. No obstante, no hay presencia de una deriva clara en los resultados, tal como se observa en la figura 5.13.b.



(a)



(b)

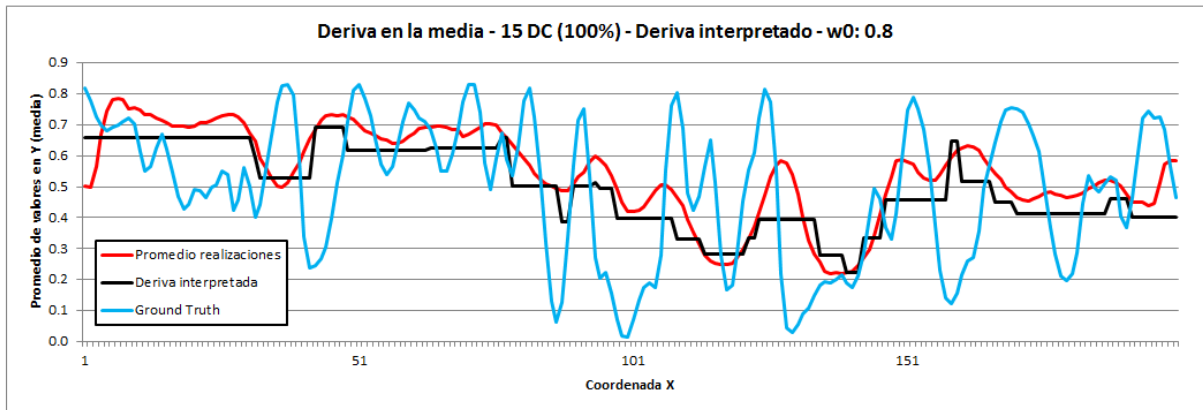
Figura 5.13: Derivas. (a) 15 DC y sin derivas externas - (b) 200 DC y sin derivas externas

Al analizar las realizaciones generadas por Filtersim incorporando derivas se observa que:

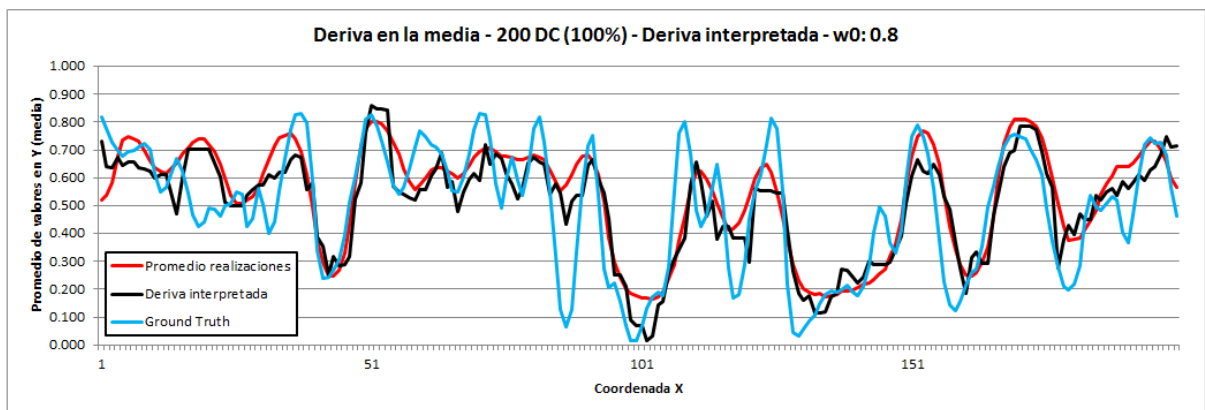
- Al aumentar la influencia de la deriva las realizaciones tienden al comportamiento presentado por la Deriva externa, en concordancia con lo esperado, y no así con la deriva del Ground Truth.
- Si se fija la influencia en un 80% y se analiza sólo la deriva interpretada (figura 5.14), se observa la importancia que tiene la deriva externa en los resultados. Lo anterior se ve

claramente en la figura 5.14 donde las realizaciones, sea con 15 DC o 200 DC, fluctúan al rededor de la deriva externa y no del Ground Truth. Esto se debe a que la deriva no está debidamente interpretada, entre más se asemeja a la realidad (G.T.) más se acercará al G.T.

- Cabe recordar que las estructuras mejor caracterizadas visualmente se obtienen con una influencia entorno al 40 % y 60 % y analizando ambos casos, para todos los escenarios simulados, la reproducción de la deriva es bastante clara y como es de esperar mejor con un 60 %. Sin embargo ya con 40 % ($w_0 : 0.4$) se obtienen resultados considerablemente mejores que al no haberlas considerado ($w_0 : 0.0$).



(a)



(b)

Figura 5.14: Derivas. (a) 15 DC - (b) 200 DC. Deriva interpretada con $w_0 : 0.8$

Mapas de probabilidad

Los mapas de probabilidad (MP) son una representación visual de la probabilidad que el valor simulado en un punto x_0 esté sobre un umbral Z_c al considerar n realizaciones. La ecuación (7.1) es la expresión matemática de lo antes mencionado.

$$P(Z \geq Z_c) = \frac{\sum_{i=1}^n I^i(x_0; Z_c)}{n} \quad (5.1)$$

Donde $I^i(x_0; Z_c)$ se define como:

$$I^i(x_0; Z_c) = \begin{cases} 1, & \text{si } Z(x_0) \geq Z_c \text{ en la realización } i \\ 0, & \text{si } Z(x_0) < Z_c \text{ en la realización } i \end{cases}$$

Tal como se explicó al comienzo del estudio son las zonas de valores altos las que presentan mayor continuidad espacial mientras que los sectores de valores medios y bajos tienden a estar más diseminados. Dado que el objetivo es caracterizar y replicar la continuidad de las zonas de alto valor, se decide utilizar un valor umbral (Z_c) igual a 0.63. El histograma de la figura 5.15 presenta en rojo la frecuencia de valores sobre el umbral 0.63. Se observa que representan un porcentaje significativo del total (44.7% aproximadamente). No se utiliza el histograma de la imagen de entrenamiento puesto que la simulación pretende representar el Ground Truth.

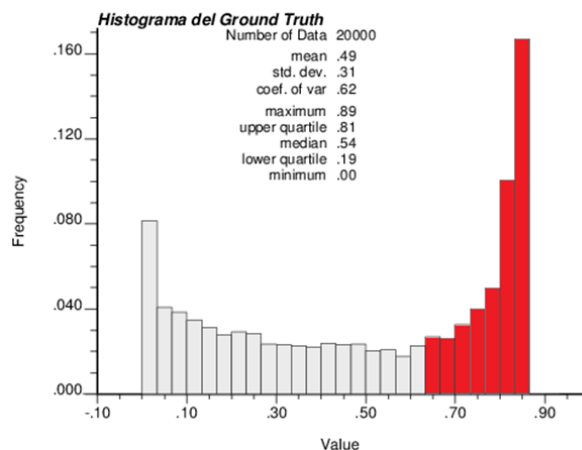


Figura 5.15: Histograma del ground truth. En rojo los valores sobre umbral 0.63

Se categoriza como 1 los valores sobre el umbral 0.63 y como 0 los valores inferiores. La figura 5.16 grafica visualmente la categorización.

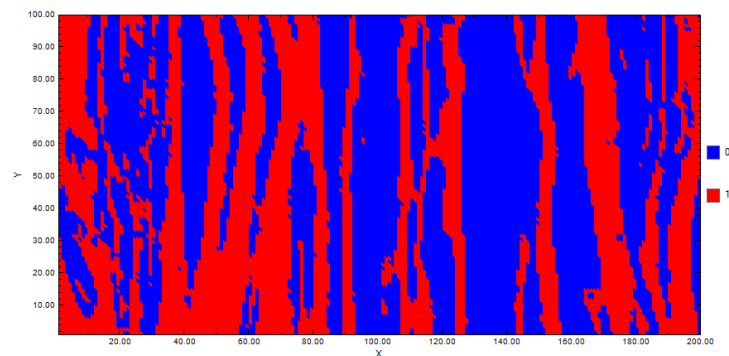


Figura 5.16: Categorización del Ground Truth sobre umbral 0.63

A continuación se presentan los MP para los 4 conjuntos Datos condicionantes – Deriva para los valores de w_0 0.0, 0.4 y 0.8.

Muy pocos datos condicionantes (100 %) – Deriva interpretada

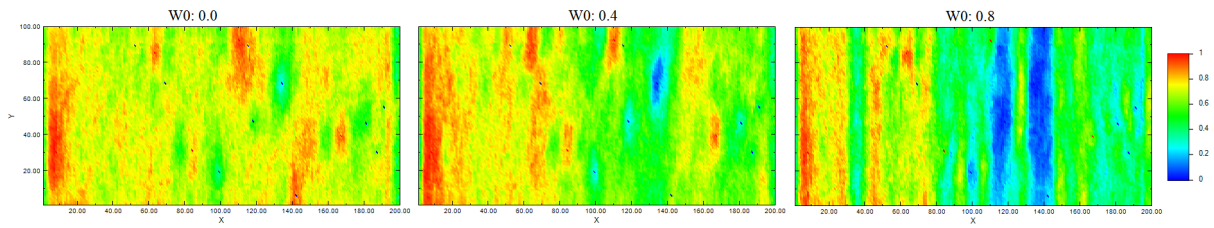


Figura 5.17: Mapas de probabilidad - 15 DC - Deriva interpretada

Muy pocos datos condicionantes (100 %) – Deriva impuesta

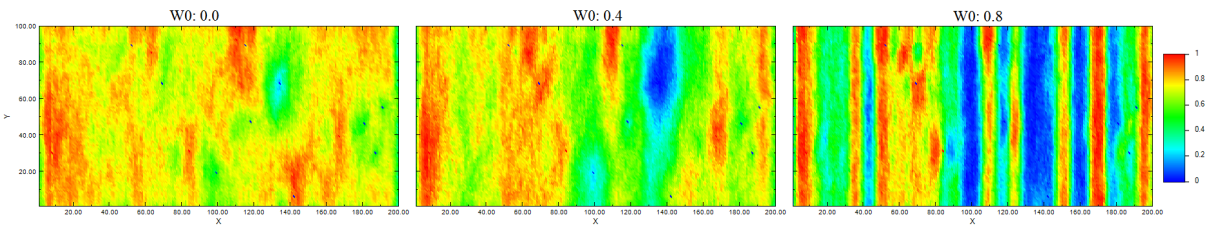


Figura 5.18: Mapas de probabilidad - 15 DC - Deriva impuesta

Pocos datos condicionantes (100 %) – Deriva interpretada

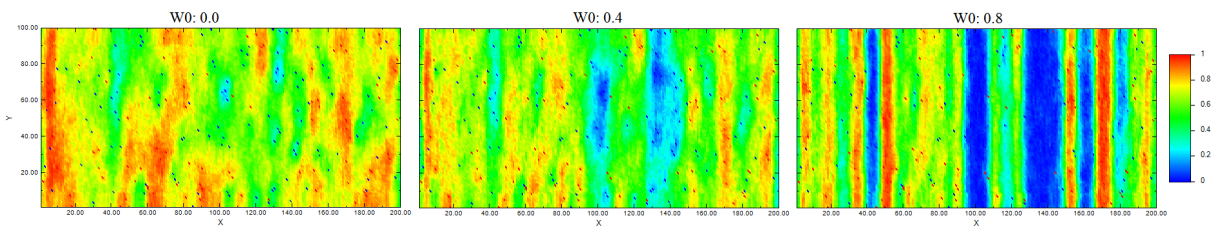


Figura 5.19: Mapas de probabilidad - 200 DC - Deriva interpretada

Pocos datos condicionantes (100 %) – Deriva impuesta

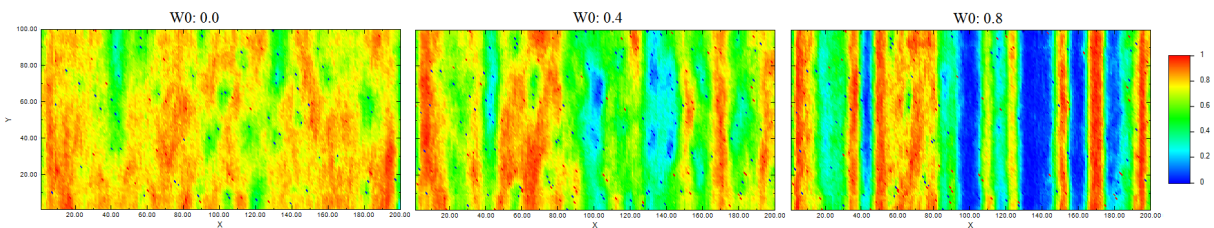


Figura 5.20: Mapas de probabilidad - 200 DC - Deriva impuesta

Se infiere de los MPs presentes en las figuras 5.17, 5.18, 5.19 y 5.20 lo siguiente:

- En primer lugar, el conjunto de 100 realizaciones por simulación entrega un grupo suficientemente vasto para caracterizar adecuadamente los sectores según su probabilidad. Además los datos condicionantes tienen una probabilidad 1 o 0, lo que está acorde a la intuición.

- Al analizar el caso $w_0: 0.0$ se observa que los MPs resultantes para el caso *Muy pocos datos condicionantes* utilizando tanto *deriva interpretada* (figura 5.17) como *deriva impuesta* (figura 5.19) tienen un alto grado de similitud. Lo mismo ocurre para el caso *pocos datos condicionantes* utilizando ambas derivas. Se corrobora así que el proceso interno de simulación se lleva a cabo de forma correcta, ya que la deriva no tiene influencia en los resultados y estos dependen sólo de los datos condicionantes.
- Continuando el análisis del punto anterior para el caso $w_0: 0.4$. Tomando en consideración sólo la *deriva interpretada* (recordar que proviene directamente de los datos condicionantes) se observa claramente cómo un aumento de los datos condicionantes lleva a caracterizar mejor la zona de estudios. En efecto, al tener información de un 1 % del sector a simular (# 200 DC) se caracteriza adecuadamente los sectores de leyes medias y bajas (bajo el umbral 0.63) no así al utilizar un 0.075 % del sector a simular (# 15 DC). Es importante recalcar la naturaleza aleatoria en la toma de muestras ya que de existir un muestreo preferencial los resultados globales se verán mermados independiente de la cantidad de datos condicionantes.
- Al observar la probabilidad asociada a las zonas de bajo y medio valor (entre 0 y 0.63) se tiene que las probabilidades asociadas están por debajo de 0.3 sólo si el entorno presenta igual condición (figuras 5.19 y 5.20 con $w_0: 0.0$ y 0.4). Para valores idénticos en un entorno de valores más altos la probabilidad asciende a 0.3 - 0.5.
- Los valores altos tienen un comportamiento vertical y angosto sumergido en una matriz de valores medios y bajos diseminados. Sin embargo, al aumentar el porcentaje de influencia de la deriva externa hasta $w_0: 0.8$ los mapas de probabilidad reflejan un comportamiento en los valores medios-bajos similar a los valores altos perdiendo su carácter diseminado. Esto lleva a asumir que sobre 80 % de influencia en la deriva externa se pierde realismo en los resultados.
- La probabilidad, sobre las 100 realizaciones, asociada a las zonas de valores altos (sobre 0.63) está muy bien caracterizada. En efecto, se observa en las 4 figuras para $w_0: 0.4$ que la probabilidad asociada está entre 0.7 – 1.0 con notoria continuidad en la vertical. Además, se observa que una influencia del 40 % lleva a mejores resultados que al utilizar 0 % o 80 %.

Análisis en el porcentaje de acierto

Habiendo previamente obtenido los mapas de probabilidad se calcula a continuación cuán asertivo es el modelo de Filtersim incorporando derivas al momento de simular el Ground Truth. Dado el umbral 0.63 utilizado en la sección anterior, se determina que un 44.7% de los 20.000 valores informados en el Ground Truth están por sobre el umbral. Se decide entonces, para cada mapa de probabilidad, fijar el valor umbral (no necesariamente 0.63) que lleve a obtener el porcentaje más cercano a 44.7% de valores sobre el umbral escogido. Luego, en los mapas de probabilidad y en Ground Truth se etiquetan con 1 y 0 los valores sobre y bajo umbral definido, respectivamente.

Se considera como acierto, luego de haber categorizado bajo un cierto umbral, cuando un punto informado como 0 en el Ground Truth es categorizado como 0 en los mapas de probabilidad. Lo mismo sucede para valores categorizados como 1 en el G.T. y como 1 en los mapas de probabilidad. Por otro lado, la sobreestimación se genera al obtener valores 1 cuando se esperaba que fuesen 0, y la subestimación en el caso inverso. Para efectos de análisis se asume como *porcentaje de acierto* a la suma de los porcentajes de acierto para valores previstos como 0 que resultan 0 y los previstos como 1 que resultan efectivamente 1.

Los principales resultados, expuestos en gráficos y tablas, se encuentran en la sección de anexos F. Para continuar con los análisis se presenta en la tabla 5.5 la cantidad de puntos según nivel de acierto para los distintos valores de w_0 . En las dos primeras columnas se denota por **Es** el valor obtenido sobre los mapas de probabilidad y por **D.S.** el valor que *debería ser* según el Ground Truth.

Tabla 5.5: Tabla resumen porcentajes de acierto - 100% datos condicionantes

		15 D.C. - Interpretado						15 D.C. - Impuesto					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	6439	6670	7127	6852	6545	6872	6511	7213	7494	8117	8323	8179
0	1	4636	4405	3948	4223	4530	4203	4564	3862	3581	2958	2752	2896
1	0	4626	4561	4297	4246	4496	4412	5166	3892	3506	3007	2753	2908
1	1	4299	4364	4628	4679	4429	4513	3759	5033	5419	5918	6172	6017
% acierto		53.7	55.2	58.8	57.7	54.9	56.9	51.4	61.2	64.6	70.2	72.5	71.0
		200 D.C. - Interpretado						200 D.C. - Impuesto					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	7924	7688	7808	7953	8011	7841	7729	8054	8448	8344	8483	8154
0	1	3151	3387	3267	3122	3064	3234	3346	3021	2627	2731	2592	2921
1	0	3425	3241	3146	3162	3022	3190	3282	2771	2814	2627	2619	2993
1	1	5500	5684	5779	5763	5903	5735	5643	6154	6111	6298	6306	5932
% acierto		67.1	66.9	67.9	68.6	69.6	67.9	66.9	71.0	72.8	73.2	73.9	70.4

Se calcula el porcentaje de acierto como la suma de los aciertos cuando el valor obtenido es 0 y debería ser 0 *más* cuando sucede lo mismo con el valor 1, sobre el total de nodos (20.000). La figura 5.21 grafica el porcentaje de acierto para los casos 200 y 15 datos condicionantes con medias inferidas e impuestas en función del nivel de influencia por parte de la deriva externa.

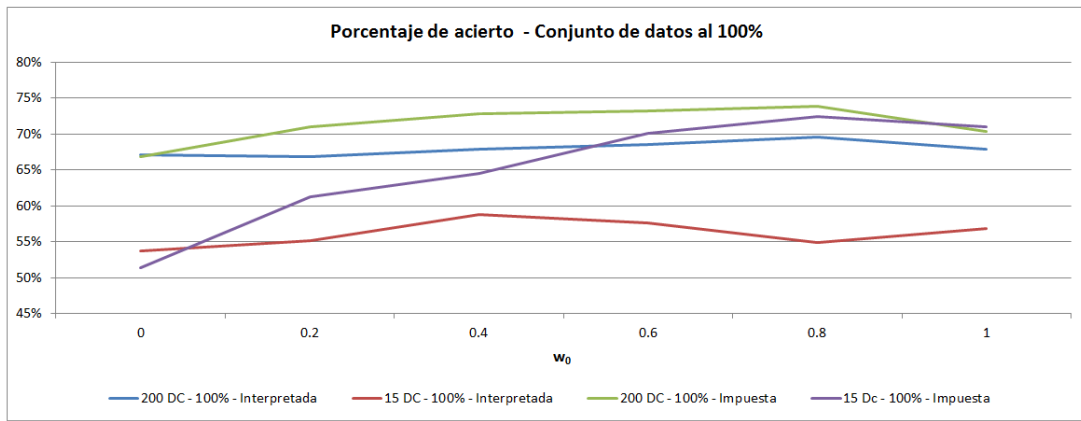


Figura 5.21: Gráfico del porcentaje de acierto utilizando *muy pocos y pocos datos condicionantes* al 100 %

De los resultados expuestos en la tabla 5.5 y la figura 5.21 se desprende:

- Sin incorporar derivas externas ($w_0 : 0.0$) se observa que el porcentaje de acierto utilizando 15 D.C. es de un 52 % mientras que al utilizar 200 D.C. es de 67 %. Además, estos valores son similares al analizar deriva interpretada e impuesta, lo cual responde a la intuición ya que no deberían tener influencia en los resultados.
- Al utilizar una deriva interpretada desde los datos condicionantes, se observa un aumento proporcional del orden de un 3 % al pasar de $w_0 : 0.0$ a $w_0 : 0.6$.
- Al utilizar una deriva impuesta desde el Ground Truth y analizar el cambio porcentual de $w_0 : 0.0$ a $w_0 : 0.6$ se observa que utilizando 15 D.C. se aumenta en un 19 % desde 51 % a 70 %. Por otro lado, utilizando 200 D.C. el cambio es menos pronunciado de 67 % a 73 % representando un incremento del 6 %.
- Al analizar solo el caso en que **es 1** y **debería ser 1** se observa que en todos los casos incluir derivas externas en la media local incrementa el nivel de acierto al simular.
- La meseta, del orden de un 72 %, observada en todos los casos al aumentar la influencia de la deriva responde al hecho que las estructuras obtenidas al simular tienen un carácter vertical-lineal distando del carácter vertical-sinusoidal del Ground Truth.
- Al analizar los casos de sobre y subestimación de los valores se tiene que la cantidad de nodos sobreestimados y subestimados son similares para todos los niveles de influencia de deriva externa.
- Sobre el punto anterior, el caso 15 D.C.- Deriva Interpretada tiene como media 4.380 nodos entre sub y sobreestimados. Al pasar a 15 D.C.- Deriva Impuesta este valor cae a 3.480. Al pasar luego al caso 200 D.C.- Deriva Interpretada este valor cae a 3.200, y finalmente el caso 200 D.C.- Deriva Impuesta este valor cae a 2.860.
- Un aumento por sobre $w_0 : 0.6$ o bien no mejora los resultados o los empeora, salvo en el caso 15 D.C. - deriva interpretada. Esto, junto con el análisis visual, lleva a establecer que aumentar el porcentaje de influencia no mejora los resultados y distorsiona las estructuras verticales.

Finalmente, al incorporar derivas externas el algoritmo incrementa el porcentaje de acierto en los resultados en comparación a Filtersim convencional. El nivel óptimo de influencia por parte de la deriva externa debe ser fijado bajo otros parámetros que den compatibilidad a este indicador de certeza en la predicción (por ejemplo, un análisis visual y variográfico). Otra aplicación interesante sería definir, para todos los mapas de probabilidad, el umbral 0.63 como corte en la categorización y comparar luego los porcentajes de acierto.

Error Cuadrático Medio global sobre la deriva obtenida – Sin umbral

Dado que la deriva se presenta a lo largo de la vertical, se calcula el promedio de los valores verticales en cada posición horizontal sobre una realización. De este modo se tiene una noción de la deriva obtenida en esta realización. Se calcula un Error Cuadrático Medio (E.C.M.) a priori al comparar una realización contra el Ground Truth. La comparación consiste en determinar el Error Cuadrático Medio a priori al promediar las diferencias cuadráticas de los promedios verticales en cada posición horizontal, tal como se ve en la ecuación (5.2).

$$E.C.M.^{(a\ priori)} = \left(\sum_{i=1}^{200} (\bar{x}_i - \bar{g}_i)^2 \right) / 200 \quad (5.2)$$

Con \bar{x}_i y \bar{g}_i el promedio en la vertical en la posición i para la realización y el Ground Truth respectivamente.

Se repite el proceso anterior para cada una de las 100 realizaciones obteniendo 100 Errores Cuadráticos Medios a priori que son luego promediados para obtener un Error Cuadrático Medio Global (ecuación (5.3)). El proceso completo se aplica sobre cada valor de w_0 midiendo con ello el efecto que tiene la ponderación de la deriva externa en los resultados.

$$E.C.M.^{Global} = \left(\sum_{j=1}^{100} E.C.M.^{(a\ priori)} \right) / 100 \quad (5.3)$$

El promedio de los Errores Cuadráticos Medios a priori (figura 5.22) da cuenta del valor esperado. Sin embargo no da cuenta de la dispersión entre los valores obtenidos en cada realización. Por ello se decide obtener una medida de dispersión considerando así la varianza presente entre los 100 E.C.M. a priori (figura 5.23).

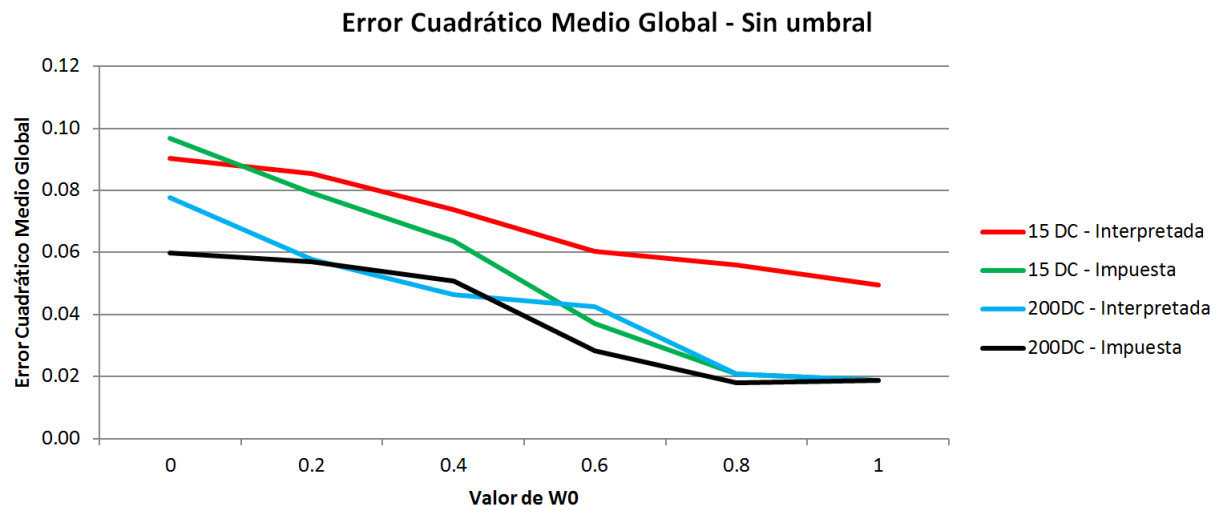


Figura 5.22: Error Cuadrático Medio Global - Sin umbral

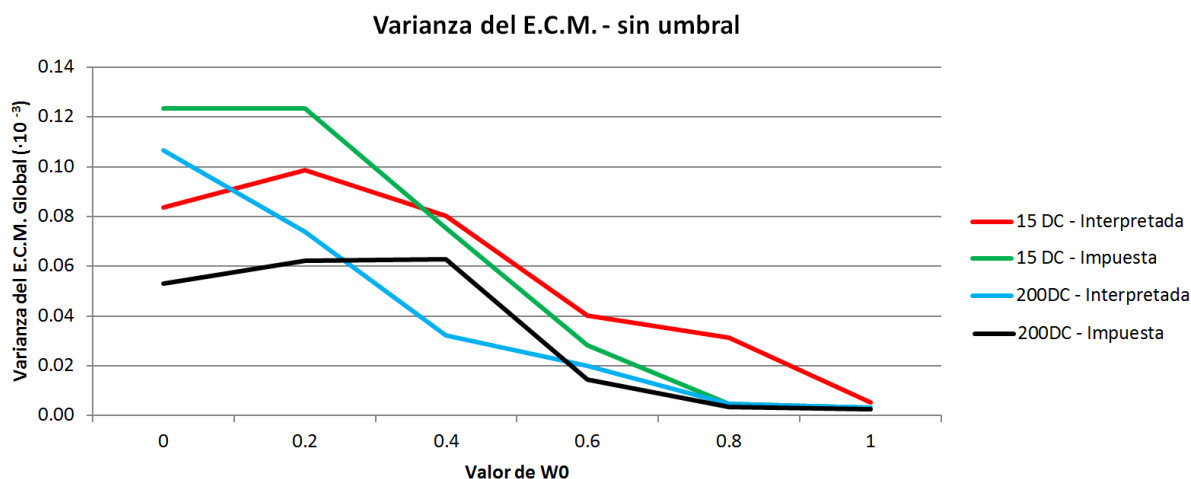


Figura 5.23: Varianza del Error Cuadrático Medio – Sin umbral

Aplicando un análisis similar al *ceteris paribus* (un parámetro es analizado mientras el resto permanece constante) salvo que en este caso se analizan uno o más parámetros con libertad, se obtienen las siguientes interpretaciones:

- Para un w_0 igual a 0.0 (sin incorporación de derivas externas) se tiene que el mayor E.C.M. global lo entrega la combinación *muy pocos datos condicionantes – deriva interpretada* por sobre el resto. Lo mismo sucede con la varianza. Lo anterior respeta los resultados esperados intuitivamente ya que a menor grado de información mayor imprecisión en los resultados y mayor variabilidad entre ellos.
- Al considerar w_0 igual a 0.0 los casos *15 DC – interpretada* y *15 DC – impuesta* tienen los mismos parámetros de entrada. Lo mismo sucede con los casos *200 DC – Interpretada* y *200 DC – impuesta*. No obstante se observa una diferencia entre los E.C.M. globales obtenidos. En el primer caso la diferencia relativa es de 0.01 mientras que en segundo caso es de 0.02. Esta diferencia relativa debería ser cercana a cero a medida que el número de realizaciones aumenta de 100 a 1.000 o 10.000. Por lo tanto se puede atribuir la diferencia relativa inicial a la estocasticidad presente en la simulación.
- Si se analiza la influencia de las derivas externas (aumento de w_0) a medida que se asigna mayor ponderación a éstas los E.C.M. globales tienden a disminuir así como también las varianzas asociadas. Esto sucede en todas las combinaciones datos condicionantes – deriva externa. Este comportamiento no es estrictamente decreciente. En efecto se observan pequeños aumentos en el E.C.M. global al aumentar el w_0 los cuales se deben principalmente al proceso aleatorio de la simulación ya que se esperan que este valor decrezca.
- Dado que la deriva impuesta proviene del Ground Truth debiese observarse un mayor aporte de información en éstas que la entregada por la deriva inferida. Esta intuición se corresponde con los resultados. En efecto, al analizar el caso w_0 igual a 1.0 (sólo consideración de la deriva externa y no de los datos condicionantes) se tiene que los menores E.C.M. son obtenidos por las combinaciones *200 DC – Impuesta* y *15 DC – impuesta*. En particular, analizando sólo el caso muy pocos datos condicionantes (15 DC) el E.C.M. global al utilizar la deriva interpretada es un 66.6 % mayor al obtenido bajo la deriva impuesta.
- La validación visual entregó como observación que al utilizar un w_0 en el intervalo [0.4-0.8] las realizaciones respetan el comportamiento sinusoidal de la variable regionalizada. Al estudiar este intervalo se aprecia que en todo el rango la disminución del E.C.M. global es siempre decreciente. Esto se visualiza en las 4 combinaciones *datos condicionantes – deriva externa*. El mismo fenómeno sucede al analizar la varianza del E.C.M. respecto a ambos

intervalos. De lo anterior se infiere que asociar un 60 % de influencia a la deriva externa ($w_0=0.6$) es sugerido, y que al aumentar este valor a 0.8 las mejoras están presentes pero no son sustanciales y que al disminuirlo, $w_0=0.4$, se aprecian deterioros en el E.C.M. global y en la representación visual.

- La deriva impuesta al provenir del Ground Truth es más rica en información que la deriva interpretada que proviene de los datos condicionantes. Esto ya que la primera cuenta con 20.000 datos informados mientras que la segunda a lo más con 200. Entre los datos condicionantes el conjunto de 200 D.C. es más rico en información que el grupo de 15 D.C. Por lo tanto al combinarlos se esperaría que la combinación *200 D.C. – Deriva impuesta* sea la mejor informada y la combinación *15 D.C. – Deriva interpretada* la menos informada. Dado que a mayor información mejor debiesen verse los resultados, se espera una jerarquía en la calidad de los resultados. En efecto, al analizar el caso $w_0=0.6$ se observa que el E.C.M. global está completamente jerarquizado (ver figura 5.22). Esto último verifica el comportamiento esperado de la simulación. Es importante destacar que el mismo fenómeno no sucede al analizar la varianza de las cuatro combinaciones.

Error Cuadrático Medio global sobre la deriva obtenida - Con umbral 0.63

Otra métrica interesante de obtener es la diferencia cuadrática promedio que existe sobre el umbral 0.63. Para ello se determinan las coordenadas X en las cuales el valor es igual o mayor a 0.63 y sobre este conjunto de coordenadas se calcula el E.C.M. a priori (ecuación (5.4)). Sea N_{um} la cardinalidad del conjunto de coordenadas cuya media en Y es mayor o igual a 0.63. Sean \bar{x}_i e \bar{g}_i el promedio en la vertical del elemento i , en el conjunto de coordenadas antes explicado, para la realización y el Ground Truth respectivamente.

$$E.C.M.^{(a \text{ priori} - \text{Umbral})} = \left(\sum_{i=1}^{N_{um}} (\bar{x}_i - \bar{g}_i)^2 \right) / N_{um} \quad (5.4)$$

El proceso anterior se lleva a cabo para cada una de las 100 realizaciones obteniendo 100 Errores Cuadráticos Medios a priori - Umbral que son luego promediados para obtener un Error Cuadrático Medio Global - Umbral (ecuación (5.5)). El proceso completo se aplica sobre cada valor de w_0 .

$$E.C.M.^{Global-Umbral} = \left(\sum_{j=1}^{100} E.C.M.^{(a \text{ priori} - \text{Umbral})} \right) / 100 \quad (5.5)$$

Cabe notar que las coordenadas que estén por debajo del umbral 0.63 **no** son tomadas en consideración y que el conjunto de elementos N_{um} es constante para todos los escenarios de estudio ya que se define sobre la deriva del Ground Truth. Los resultados del E.C.M. medio global se grafican en la figura 5.24 y las varianzas respectivas entre las 100 realizaciones se grafican en la figura 5.25.

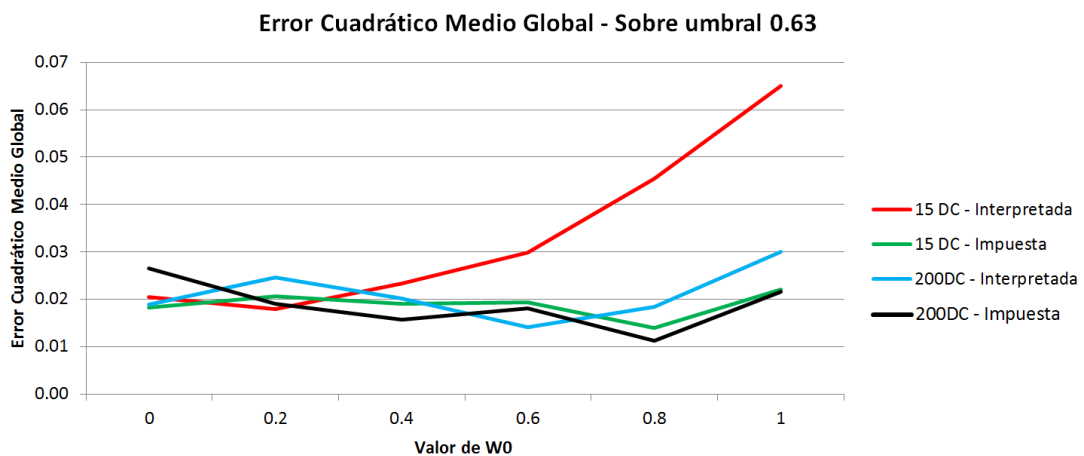


Figura 5.24: Error Cuadrático Medio Global - Sobre umbral 0.63

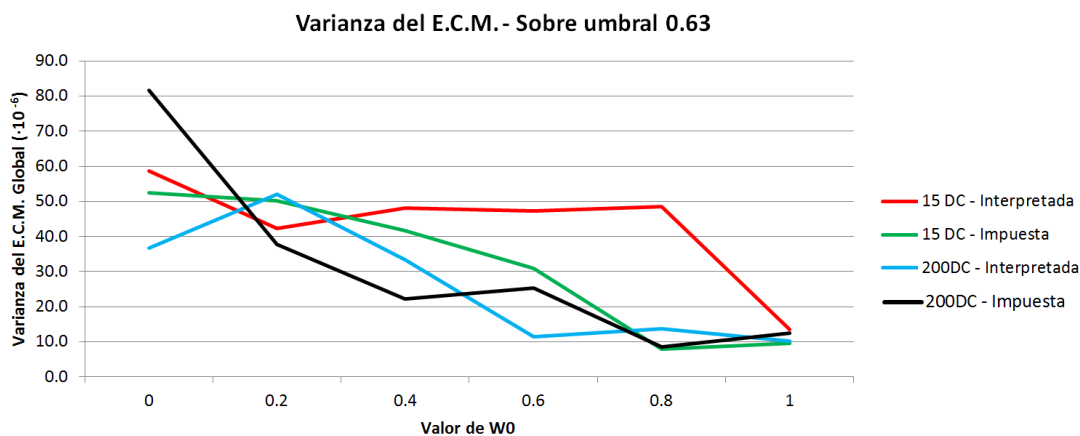


Figura 5.25: Varianza del Error Cuadrático Medio – Sobre umbral 0.63

De los resultados anteriores se interpreta:

- Analizando los E.C.M no se aprecia mejoras en la incorporación de derivas en el intervalo $w_0 :]0.0 - 0.4]$ en ninguno de los escenarios.
- No se aprecia jerarquía entre los escenarios en el intervalo $]0.0 - 0.4]$ en los E.C.M. Globales.
- Analizando las varianzas: para la deriva impuesta un aumento en la influencia lleva a disminuir la dispersión entre los E.C.M. a priori. El caso particular *200 DC - Deriva impuesta* presenta el mismo E.C.M. Global con $w_0 : 0.2$ que con $w_0 : 0.6$ pero una varianza claramente más baja con 0.6 que con 0.2.
- El caso particular del escenario *15 DC - Deriva interpretada* incrementa el E.C.M. a partir de $w_0 : 0.2$ lo cual se explica por la escasa información desde la cual se infiere la deriva. Con tan sólo 15 datos condicionantes un gran número de coordenadas X tienen valores bajo 0.63 como deriva en la media interpretada que distan en gran medida del valor real de la deriva en esa coordenada.
- Es interesante notar que para el caso $w_0 : 0.6$ los E.C.M. para los escenarios *15 DC - Deriva impuesta*, *200 DC - Deriva interpretada* y *200 DC - Deriva impuesta* son relativamente iguales.

Dado todo lo anterior, se concluye que obtener el Error Cuadrático Medio sobre un cierto umbral **no** ayuda a la elección del mejor escenario ni a la mejor proporción de influencia de la deriva. Sin embargo, puede ser de utilidad para detectar cuan ricas en información son las derivas obtenidas en los sectores de interés para cierto porcentaje de influencia (como es el caso de 15 DC - interpretado sobre $w_0 : 0.2$).

Comparación con Kriging

El Kriging es una de las técnicas de estimación más utilizadas en geoestadística. Dada su naturaleza de estimador, difiere de las simulaciones al generar un mapa de valores esperados y no un conjunto de posibles escenarios. Existen diversas modificaciones a la técnica, sin embargo, el Kriging Simple (KS) y el Kriging Ordinario (KO) son las más utilizadas (ver sección *Estado del arte > Estimación mediante Kriging*) por lo que son estas las empleadas para comparar sus resultados contra la versión modificada de Filtersim.

El KO toma en consideración la correlación espacial de los datos condicionantes a través del variograma y al estimar pondera los datos condicionantes dentro de una vecindad bajo su disposición espacial. Además considera que la media sobre el dominio a estimar es un parámetro desconocido. Por su parte el KS solo se diferencia del KO por incorporar como variable conocida la media global. Dos son los conjuntos de datos condicionantes: *muy pocos datos (15)* y *pocos datos (200)* sobre los cuales se aplican ambas técnicas considerando para el KS una media global igual a 0.49 que es la media presente en el Ground Truth (figura 5.2.c).

Los parámetros utilizados para determinar el variograma experimental se presentan en la tabla 5.6.

Tabla 5.6: Parámetros para el cálculo del variograma experimental (15 DC y 200DC)

Parámetros	N0°E	N90°E
Azimut	0.0	90.0
Tolerancia en azimut	45.0	45.0
Ancho de banda horizontal	10.0	10.0
Dip	0.0	0.0
Tolerancia en Dip	5.0	5.0
Ancho de banda vertical	10.0	10.0
Número de pasos	100.0	100.0
Distancia entre pasos	1.0	1.0
Tolerancia en el paso	1.0	1.0

En la figura 5.26 se visualiza el variograma experimental (líneas discontinuas) y modelado (líneas continuas) para ambos conjuntos de datos condicionantes. En color (**negro**) la dirección N0E y en (**rojo**) la dirección N90E.

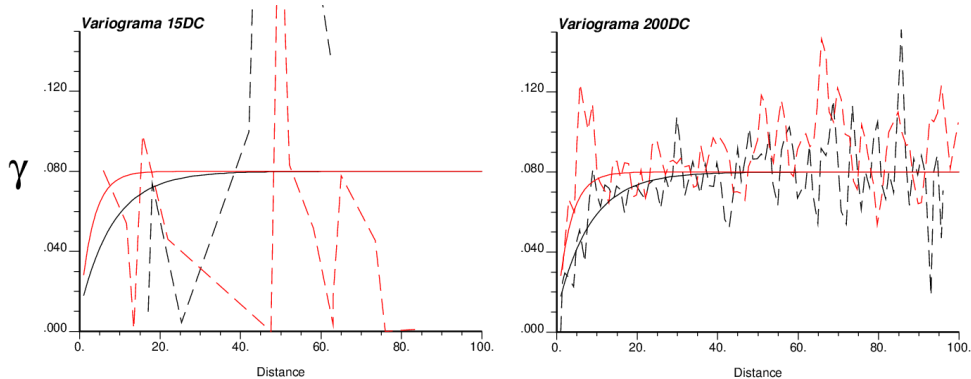


Figura 5.26: Variogramas experimentales y modelados para *muy pocos* y *pocos datos condicionantes*

Dada la baja cantidad de datos para la inferencia del variograma en el conjunto *muy pocos datos condicionantes*, se decide utilizar el mismo modelo variográfico utilizado en *pocos datos condicionantes*. El modelo variográfico se presenta en la ecuación (5.6).

$$\gamma = 0.01 + 0.07 \cdot \exp(25, 10) \quad (5.6)$$

La vecindad de búsqueda utiliza como radios 70 y 70 unidades en las direcciones X e Y respectivamente. Se establece un mínimo y máximo número de datos condicionantes de 1 y 8 respectivamente sin cantidad mínima por octante. Las figuras 5.27 y 5.28 presentan los resultados al utilizar Kriging ordinario y simple, respectivamente, considerando 15 y 200 datos condicionantes.

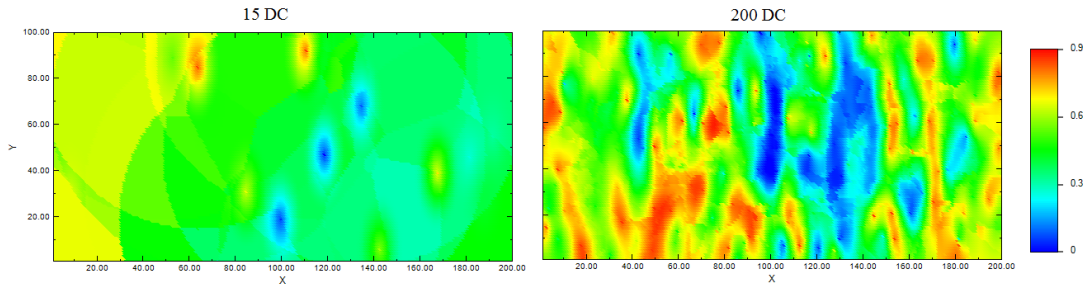


Figura 5.27: Kriging ordinario. (Izq) 15 datos condicionantes – (der) 200 datos condicionantes

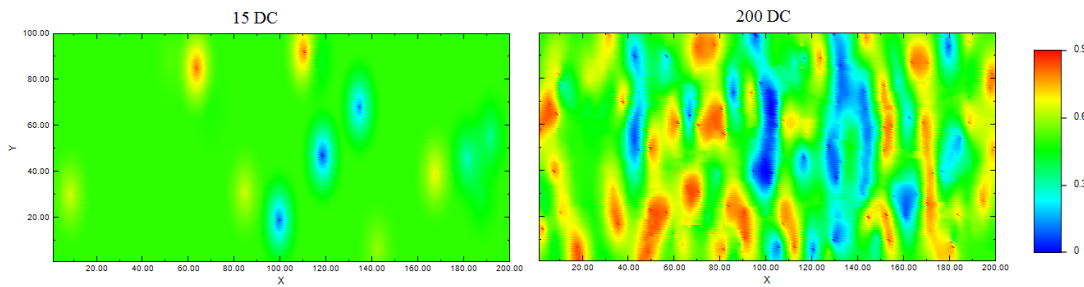


Figura 5.28: Kriging simple. (Izq) 15 datos condicionantes – (der) 200 datos condicionantes

De los resultados anteriores se desprenden las siguientes observaciones:

- Al visualizar la estimación mediante KO (figura 5.27) y KS (figura 5.28) queda claro que ninguno logra reproducir adecuadamente el carácter continuo de los valores altos en la dirección vertical. Podría decirse que el KO entrega resultados mejores que el KS ya que sectores de alto valor tienen un carácter más continuo, sin embargo esto se debe que al no considerar una media conocida la influencia de los datos condicionantes tiene un mayor alcance.
- La diferencia de valores entre pares de puntos cercanos es bastante pequeña, es decir, la variable se desestructura de manera suave lo cual no representa la desestructuración real (abrupta, ver figura 5.1).
- El principal problema del Kriging, para este caso de estudio, es su propiedad intrínseca de *suavizamiento*. El interés principal es reproducir el comportamiento de la variable regionalizada en sus valores altos, sin embargo la propiedad de suavizamiento disminuye considerablemente la proporción de estos en la estimación, subestimándolos. Un proceso similar sucede con los valores bajos, en este caso sobrestimándolos (ver gráficos de distribución de la figura 5.29). Por lo tanto se pierde el sentido final de la estimación.
- Sumado a lo anterior, la forma en la distribución de los valores estimados tiende a ser Gaussiana. No obstante, la distribución de los valores tanto en la T.I. (inferencia del variograma) como de la G.T. (sector a estimar, origen de los datos condicionantes) difieren bastante de una distribución tipo Gaussiana.

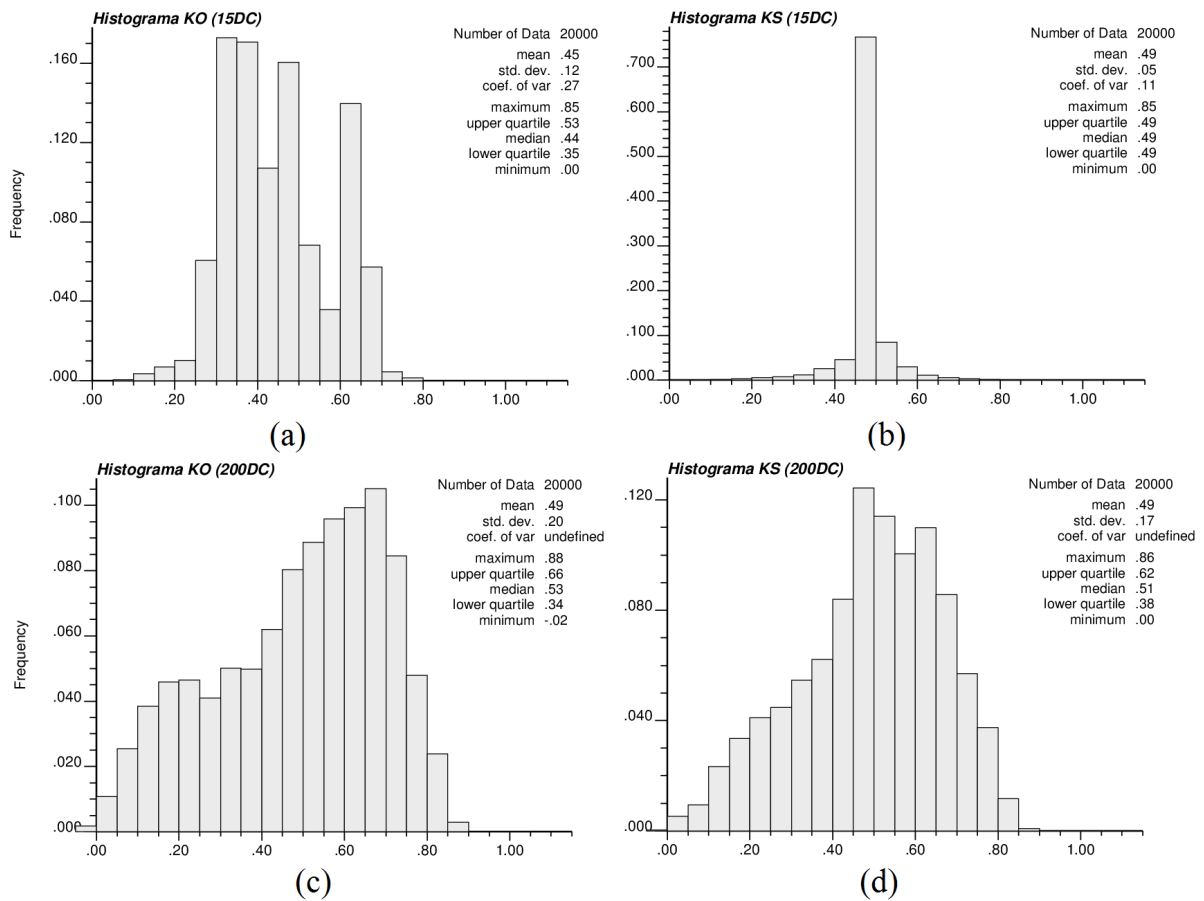


Figura 5.29: Histogramas resultados Kriging Simple y Ordinario para 15 y 200 datos condicionantes

- El Kriging adolece de la incapacidad de reproducir las características de la variable regionalizada si no se cuenta con información para el condicionamiento.
- Contar con un 0.075 % del sector en estudio con información condicionante sólo genera ruido en los resultados reflejados como un halo de valores que tienden hacia la media al alejarse del punto condicionante. El aumento en el área informada de un 0.075 % a un 1 % produce mejoras considerables a la hora de utilizar Kriging. En particular, se observan elipses con eje mayor en la dirección vertical acorde con lo esperado por parte de las estructuras. El KO logra unir las elipses generando así mayor continuidad de los sectores de alto valor en contraste con el KS que genera halos gradacionales aislados en una matriz de valores medios (cabe recordar que la vecindad de búsqueda es fija en toda la comparación).

Comparación con algoritmo de Bandas Rotantes

A diferencia del Kriging que funciona como estimador minimizando la varianza del error el algoritmo de Bandas Rotantes (BR) es un algoritmo de simulación que busca representar la variabilidad de la función aleatoria. Para la inferencia de la continuidad espacial se utiliza el variograma de los datos condicionantes cuyos parámetros se describen en la tabla 5.6 y se visualizan en la figura 5.26 (se utiliza el mismo variograma modelado). Se generan 100 realizaciones (con semilla igual a 30061971) sobre la grilla de simulación a una discretización de bloques de 1x1. La vecindad de búsqueda es de 70x70 condicionando con un mínimo de 1 dato y un máximo de 8 (sin uso de octantes). Se utiliza Kriging Ordinario en el sub-proceso de estimación (la simulación por parte de BR se realiza en un conjunto de direcciones unidimensionales (rectas) a lo largo de todo el espacio (en este caso 2D)). Para la extrapolación de las colas se utiliza el valor mínimo y máximo según la imagen de entrenamiento (0.0 y 0.897).

En la figura 5.30 se presenta una de las cien realizaciones utilizando BR condicionado por *muy pocos datos condicionantes* y *pocos datos condicionantes*. La figura 5.31 presenta el promedio de las cien realizaciones.

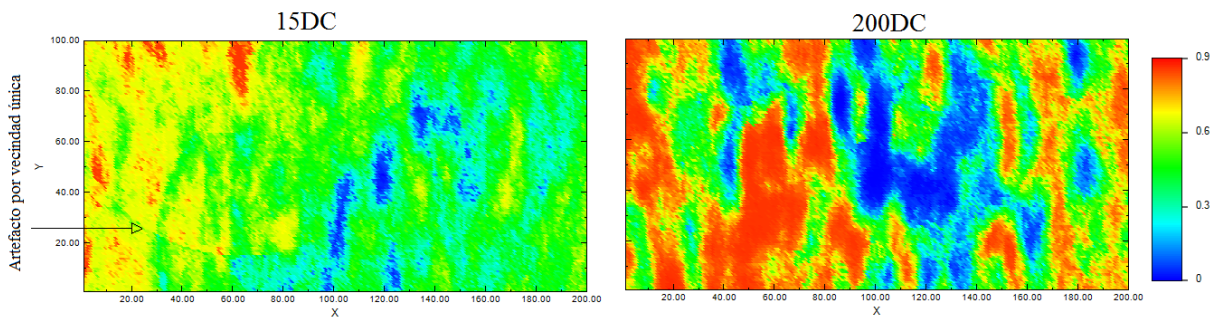


Figura 5.30: Bandas Rotantes. Una realización. (Izq) 15 D.C. – (der) 200 D.C.

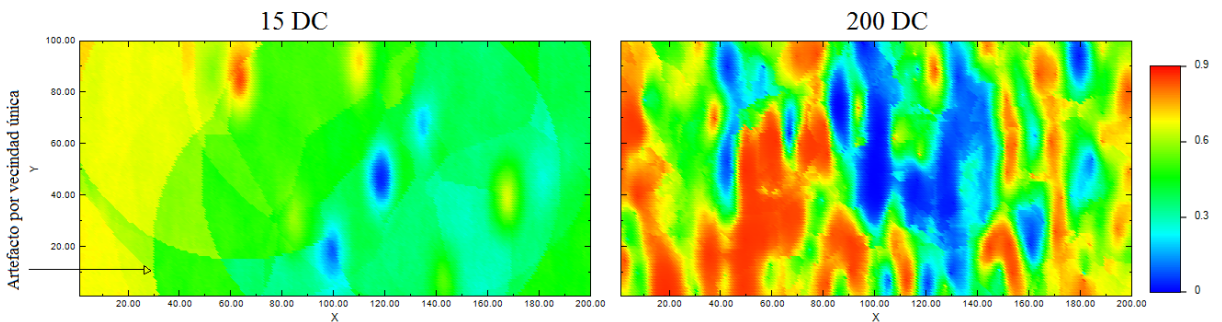


Figura 5.31: Bandas Rotantes. Promedio de 100 realizaciones. (Izq) 15 D.C. – (der) 200 D.C.

En la figura 5.32 se muestra el histograma de una realización y el promedio de entre las cien, tanto para los 15 DC como para los 200 DC.

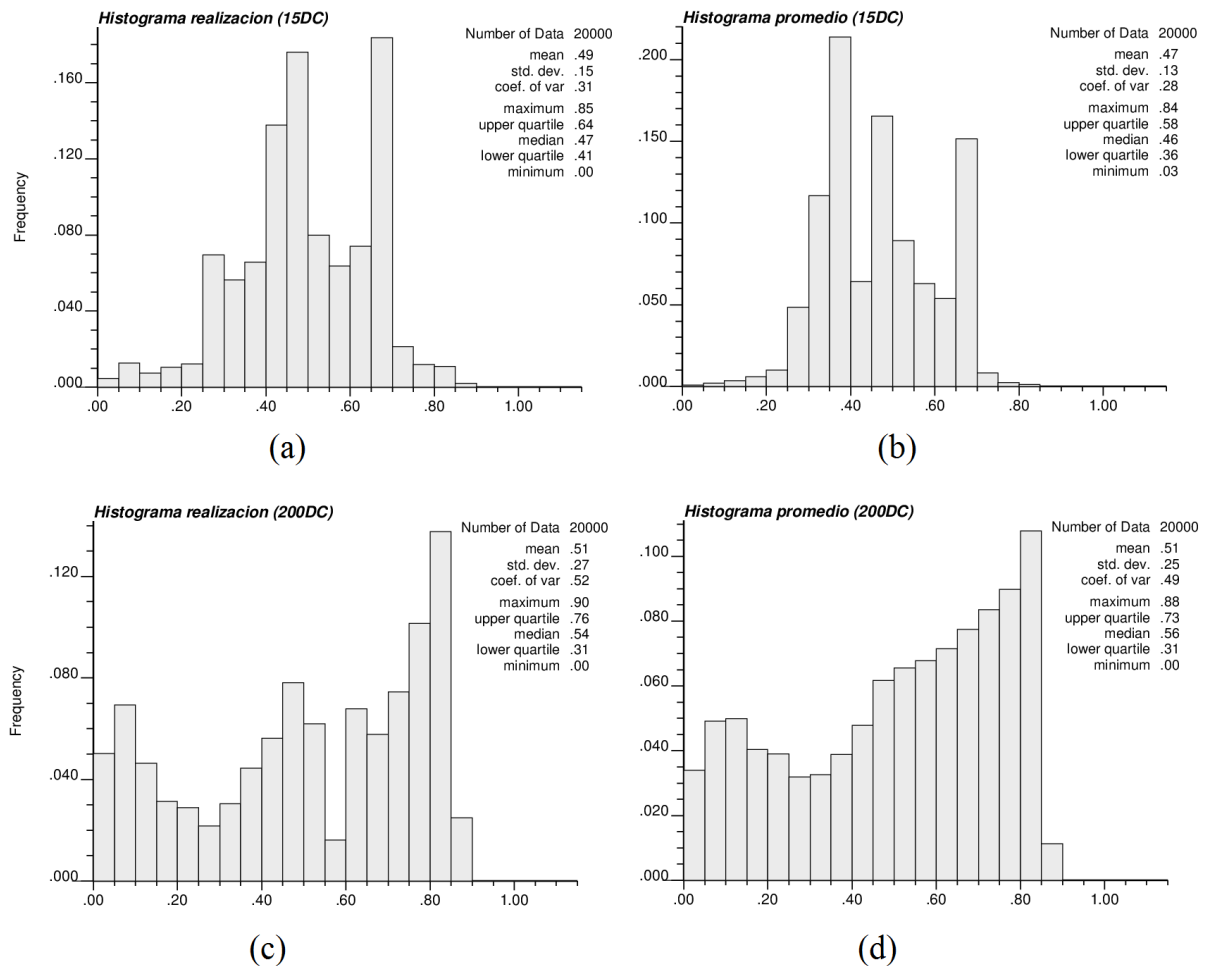


Figura 5.32: Bandas Rotantes. Histogramas: (a) una realización con 15DC, (b) promedio de 100 realizaciones con 15DC, (c) una realización con 200DC, y (d) promedio de 100 realizaciones con 200DC

De los resultados anteriores se desprenden las siguientes observaciones:

- Las realizaciones, utilizando 15 datos condicionantes, generan estructuras en forma de elipse con eje mayor en orientación vertical. Los contactos entre sectores de valores altos-medios-bajos son borrosos y presentan ruido. Se observa jerarquía entre los valores, es decir, no hay presencia de valores altos y bajos juntos. Además, la distribución estadística de los valores simulados sigue un comportamiento Gaussiano (figura 5.32.a) alejándose por completo de la real distribución del Ground Truth.
- El promedio de las 100 realizaciones, al considerar 15 datos condicionantes, presenta halos de influencia muy marcados que corresponden al radio de influencia considerado en la vecindad de búsqueda de 70×70 .
- Al considerar 200 datos condicionantes, se presentan estructuras verticales más claras y menos elipsoidales. Estas estructuras se observan tanto en valores altos como bajos mientras que los valores medios tienen un carácter más diseminado. Se continúan observando contactos rugosos con presencia de ruido entre valores cercanos al límite (bajos-medios o medios-altos). El histograma de las realizaciones (figura 5.32.c) se asemeja más al histograma del Ground Truth.
- El promedio de las 100 realizaciones, al considerar 200 datos condicionantes, aumenta la proporción de valores sobre 0.5 (figura 5.32.d) presentando estructuras mejor delimitadas con contornos más suaves y orientación vertical bien definida.

Finalmente, considerar un número muy bajo de datos condicionantes (15 DC) lleva a realizaciones que distan mucho de la realidad. Por otro lado, al aumentar el número a 200 datos condicionantes se mejora considerablemente la reproducción de estructuras tanto en las realizaciones como en el promedio. Aún así las estructuras no son comparables a las presentes en el Ground Truth.

Dado que al utilizar 200 DC el promedio de las realizaciones establece claramente las zonas de valores altos y valores medios-bajos, este puede ser utilizado para definir las derivas en la media para simulación con Filtersim modificado.

5.2. Segundo Caso de estudio - Yacimiento Cerro Colorado

5.2.1. Introducción y objetivos

En los proyectos mineros la estimación de recursos está presente a lo largo de toda la vida de la mina. En sus primeras etapas cuantifica los recursos/reservas que luego son levantadas como los principales activos de la mina. Durante la explotación agrega valor cuantificando los recursos/reservas en sectores potenciales para realizar expansiones, entre otras aplicaciones, estableciendo el riesgo asociado a estos recursos/reservas. Actualmente son los métodos de estimación (Vecino más cercano, Inverso de la distancia, Kriging y sus variaciones, entre otros) quienes rigen la evaluación del yacimiento. Sin embargo, técnicas de simulación (Simulación secuencial Gaussiana, Multipuntos, Bandas rotantes, entre otros) han ido incorporándose durante el último tiempo.

El presente caso de estudio tiene lugar en la yacimiento Cerro Colorado (propiedad de BHP Billiton) aplicando en este los algoritmos de Simulación Filtersim modificado (incorporando derivas en la media local) y Secuencial Gaussiana (a modo de comparación). La aplicación se enmarca en la simulación de leyes de Cobre Total en bancos a explotación futura en el corto-mediano plazo (bancos inferiores a un banco en específico). A partir de los resultados, diversas decisiones pueden ser evaluadas entre ellas: replanificación de corto plazo, cuantificación de la incertidumbre en las leyes previamente estimadas con las leyes actualmente simuladas, análisis de variación en la ley media esperada del banco completo, entre otras.

5.2.2. Descripción Yacimiento Cerro Colorado

La Compañía Minera Cerro Colorado (CMCC) es propiedad y operada 100 % por la empresa BHP Billiton - División Metales Base. Se sitúa en la región de Tarapacá a 125 kilómetros de la ciudad de Iquique, Chile a una altura de 2600 msnm (ver figura 5.33) y posee propiedades de explotación sobre un área de 16700 hectareas.

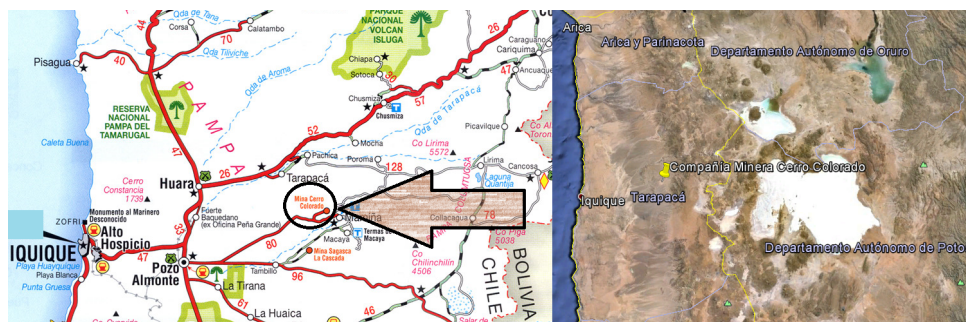


Figura 5.33: Ubicación del yacimiento Cerro Colorado

Cerro Colorado es un depósito tipo pórfido cuprífero explotado a cielo abierto. Su principal producción es cobre en forma de cátodos con una producción anual del orden de 130000 toneladas anuales. El actual rajo posee 2 y 1.5 kms de largo en las direcciones Este-Oeste y Norte-Sur respectivamente. Se presentan dos zonas principales de mineralización: una zona mineralizada de 50 a 250 metros de espesor y una capa de material lixiviado de 50 a 150 metros. En el sector Este del depósito se observan múltiples capas de óxidos y sulfuros cuyos contactos e interacciones dan origen a cuerpos de complejas geometrías. Por el lado Oeste se presentan óxidos sobre una unidad de sulfuros menos compleja en su disposición espacial ¹.

Se data su génesis en aproximadamente 52 millones de años. Procesos supérgenos generaron perfiles complejos de meteorización a una profundidad promedio de 450 metros (como se detalló

¹Fuente: chile.infomine.com

previamente) generando reservas estimadas en 228 millones de toneladas de cobre con una ley promedio de 1.0 % [4]. Al analizar el perfil de la mineralización supérgena se caracterizan cuatro zonas principales (ver figura 5.34): lixiviada superior; supérgenos superiores; lixiviada inferior; y supérgena inferior. Las zonas lixiviadas tienen alta presencia de hematita. En la zona de supérgenos superiores predominan la brocantita y atacamita con una importante inclusión de vetas de crisocola. En la zona de supérgenos inferiores se observa una matriz arcillosa principalmente caolinita (arcillas blancas) y filosilicatos de aluminio rica en Calcocita.



Figura 5.34: Caracterización de las principales zonas litológicas de Cerro Colorado

Se observa en las figuras 5.34 y 5.35 que en la pared Este del rajo aflora un flujo ignimbítico que cubre la mayor parte de la capa lixiviada, y que a su vez está cubierta por grava. Una segunda unidad, más compacta, de gravas rojas se acumula en la zona inferior. Por la pared Oeste la capa de lixiviados es principalmente hematítica en una matriz originalmente rica en brechas de sulfuros. Por su parte, hay presencia de jarosita en las rocas andesíticas levemente alterada y crisocola en la zona oxidada.

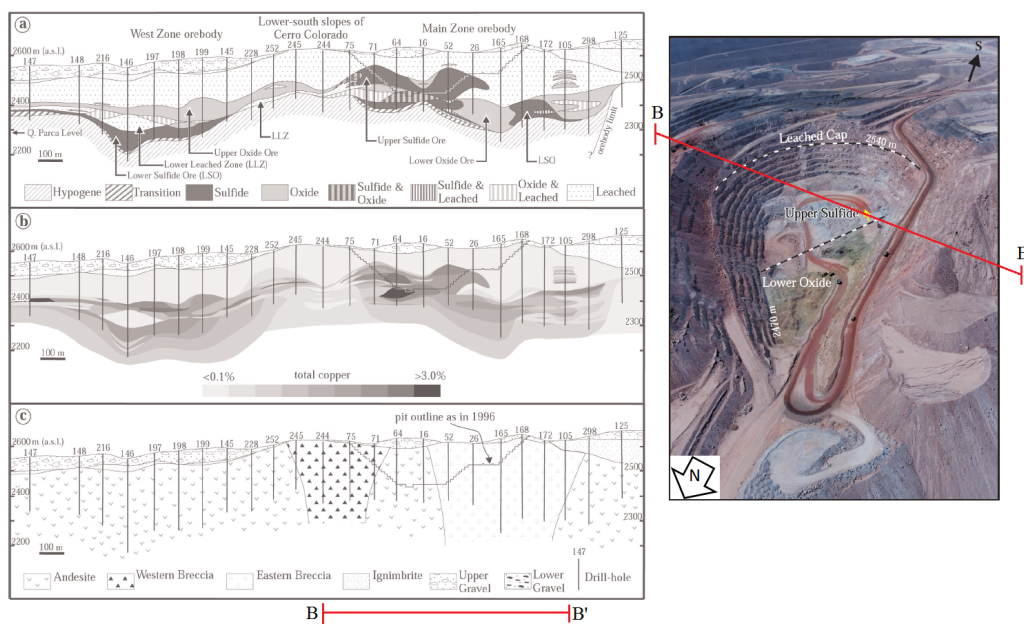


Figura 5.35: Perfil transversal alteraciones, ley de cobre y litología - Cerro Colorado

5.2.3. Metodología

La presente metodología será aplicada a lo largo de todo el caso de estudio. A modo de resumen: se comienza analizando las bases de datos; se define la variable (propiedad) a simular y donde se llevará a cabo la simulación; se define desde que zona inferir el comportamiento de la variable regionalizada; se validan las simulaciones; y finalmente se analizan los resultados e infieren métricas relevantes.

1. Estudio exploratorio de datos

Se cuenta con dos bases de datos al año 2010: pozos de tronadura y sondajes de exploración. Ambas son analizadas estadísticamente en pos de familiarizarse y detectar características/anomalías en el yacimiento (alto de banco, distribución de leyes, unidades geológicas, entre otras).

- a) Se presentan las vistas en planta, transversales, longitudinales e isométricas de los pozos y de los sondajes.
- b) Se presentan, analizan y comentan los histogramas y tablas resúmenes con las estadísticas básicas de las bases de datos.
- c) Se visualizan las derivas en las tres direcciones principales (este, norte, elevación).
- d) Se buscan anomalías utilizando gráficos de dispersión entre el Cobre Total y el Cobre Soluble.

2. Simulación

En base a las estadísticas de las bases de datos se determina la unidad geológica (UG) a considerar en la simulación. Dentro de esta UG se analizan los bancos (pozos de tronadura) para determinar un banco (cota Z_1) adecuado para extraer la imagen de entrenamiento. Se procede luego a simular los 4 bancos inferiores ($Z_1 - \text{Altura de banco}$, $Z_1 - 2 \cdot \text{Altura de banco}$, $Z_1 - 3 \cdot \text{Altura de banco}$ y $Z_1 - 4 \cdot \text{Altura de banco}$).

- a) Se determina la UG más adecuada para la simulación basado en criterios de continuidad espacial, suficiencia de datos, representatividad de la distribución de leyes, y en general de la calidad de la información.
- b) Habiendo determinado la UG se analizan los pozos de tronadura (banco a banco) y se determina el más adecuado para la extracción de la imagen de entrenamiento.
- c) Creación de una grilla para la imagen de entrenamiento ubicándola sobre el banco seleccionado. Dado que los pozos no están ubicados de forma regular en el espacio, estos deben ser migrados a la grilla regular previamente creada. Para ello se utiliza el método Vecino más cercano centrado en el nodo y cuyo radio de búsqueda es igual a la mitad + 1 del largo/ancho de cada bloque.

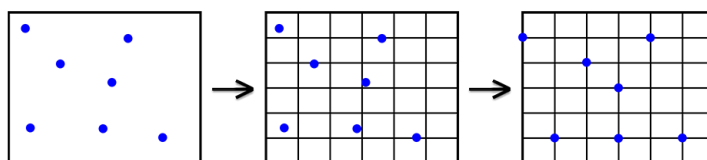


Figura 5.36: Migración de datos a una grilla regular

- d) Al término de la etapa anterior un conjunto de nodos quedará sin información. Estos son estimados con la técnica Inverso de la distancia (potencia 2, $c = 0$, mínimo y máximo de datos 1 y 4 respectivamente).
- e) En el banco inmediatamente inferior se genera una segunda grilla (grilla a simular). Esta se debe ubicar en una zona con presencia de información tanto de sondajes de exploración como de pozos de tronadura. Además se debe procurar que ésta esté lo más centrada posible respecto al centro de la imagen de entrenamiento.

- f) Se migran los datos de sondajes a la grilla a simular.
- g) Se extraen, a partir de los datos de sondajes migrados:
 - 1) Deriva en coordenada X.
 - 2) Variograma omnidireccional (por cada banco).
- h) Se simulan las leyes de Cobre Total en el banco generando los siguientes casos:
 - 1) Caso base: simulación utilizando SGSim.
 - 2) Caso F1: simulación utilizando Filtersim modificado sin incorporación de derivas.
 - 3) Caso F2: simulación utilizando Filtersim modificado incorporando derivas con un 40 % de influencia.
- i) Se repite el proceso para los bancos siguientes.

3. Validación

Al generar modelos matemáticos que representen comportamientos de fenómenos naturales estos deben ser validados desde distintas aristas. Se debe verificar no solo la reproducción de los valores observados en el sistema sino además la estructura interna implícita que lleva a la generación del fenómeno.

a) Revisión de sesgo

La imagen de entrenamiento extraída del banco (Z_1) presenta una distribución de leyes de Cobre total con una media ($m_{T.I.}$) y una varianza ($\sigma_{T.I.}^2$) específica. Por su parte, los sondajes asociados al banco inferior ($Z_2 = Z_1 - \text{altura de banco}$), los pozos de tronadura y especialmente el modelo simulado presentan medias y varianzas definidas como: $m_{sondajes}$, m_{pozos} , $m_{simulacion}$ y $\sigma_{sondajes}^2$, σ_{pozos}^2 , $\sigma_{simulacion}^2$ respectivamente. Dado que tres serán los casos simulados (Caso base, Caso F1 y Caso F2) se tiene así 3 medias y 3 varianzas para el caso $m_{simulacion}$ y $\sigma_{simulacion}^2$ respectivamente.

Una vez obtenidas las medias y varianzas se calculan los errores relativos tal como se presentan en la ecuación (5.7).

$$\text{Error relativo} = \epsilon_m(\%) = \frac{m_i - m_j}{m_j} \cdot 100\% \text{ y } \epsilon_\sigma(\%) = \frac{\sigma_i^2 - \sigma_j^2}{\sigma_j^2} \cdot 100\% \quad (5.7)$$

b) Verificar continuidad

Para medir la continuidad espacial de las simulaciones se utiliza el variograma. Para ello se utilizan los mismos parámetros variográficos de la Imagen de entrenamiento en los pozos de tronadura, en los sondajes, en la deriva impuesta y esencialmente en las realizaciones obtenidas al simular (tal como se observa en el bosquejo de la figura 5.37).

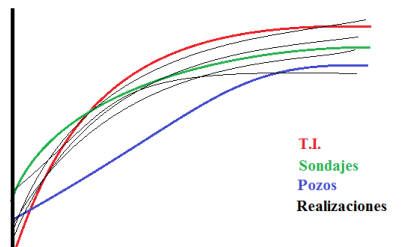


Figura 5.37: Ilustración análisis variográfico para validación de continuidad espacial

4. Análisis y métricas

Habiendo validado las simulaciones se procede a la obtención de métricas para luego ser analizadas.

a) Error Cuadrático Medio

El Error Cuadrático Medio (ECM) es el encargado de medir la dispersión de las realizaciones respecto a la realidad. En efecto, se calcula el ECM entre los valores simulados (caso base, caso F1 y caso F2) y los reales (pozos de tronadura).

Sean:

- # L: número de realizaciones.
- # N: número de nodos informados en pozos de tronadura.
- (u_n) : la posición de un nodo con información de pozos de tronadura ($u_n : 1 \dots N$).
- $Z_{real}(u_n)$: ley de Cobre total en la posición u_n extraído de los pozos de tronadura (valor real).
- $Z_{sim}(u_n)$: ley de Cobre total simulado en la posición u_n . Este valor varía dependiendo del caso: caso base, caso F1 y caso F2.

$$ECM = \frac{1}{L \times N} \sum_{l=1}^L \sum_{n=1}^N (Z_{real}(u_n) - Z_{sim}(u_n))^2 \quad (5.8)$$

b) Derivas

Cada banco tiene una deriva en la media que proviene de la información entregada por los pozos de tronadura. Esta deriva será aceptada como la realidad. Por cada realización obtenida al simular (en cada uno de los 3 casos) se obtendrá una deriva, así es posible obtener un conjunto de derivas por cada simulación entregando la posibilidad de compararla respecto a la realidad. En el caso particular del Caso F2, en el cual se impone una deriva externa, se agrega esta para su comparación. Todos los resultados son presentados gráficamente a lo largo de la dirección de análisis.

c) Mapas de probabilidad

Por cada simulación se generan 100 realizaciones, lo cual lleva a tener un vector de 100 posibles valores para cada bloque simulado. Definiendo un valor umbral se calcula el porcentaje de valores al interior del vector que igualan o superan el valor umbral. Este valor es subjetivo, sin embargo, y dado que la variable en estudio es la Ley de cobre total se opta por utilizar como umbral la Ley de Corte crítica. Esta Ley de Corte responde a condiciones técnico-económicas y se obtiene cuando el costo de producir una tonelada de cobre iguala a la ganancia (beneficio iguala a cero).

Habiendo obtenido los mapas de probabilidad banco a banco es posible determinar criterios de incertidumbre en superar la Ley de Corte. Con lo anterior cada bloque queda categorizado con valores 1 si el porcentaje de veces que la ley simulada está por sobre la ley de Corte está por sobre un porcentaje-umbral y 0 en caso que la cantidad de veces sea menor al umbral anterior.

d) Curvas Tonelaje-Ley

Las curvas de Tonelaje-Ley son una representación gráfica de la cantidad de material que superan una ley de corte y la ley media que en ellos se encuentra. Es posible inferir para cada realización su curva de Tonelaje - Ley, la cual es contrastada con la inferida desde los pozos de tronadura (realidad).

5.2.4. Estudio exploratorio de datos

Se cuenta con dos bases de datos: sondajes de exploración (o directamente *sondajes* en su forma abreviada) y pozos de tronadura (*bancos* o *pozos* en su forma abreviada) al año 2010. Se estudia primero la base de sondajes y luego la base de pozos.

Sondajes

Se cuenta con una base de datos de 21268 muestras de sondajes. Las propiedades medidas, o clasificadas, para cada muestra son: largo de la muestra, cobre soluble, cobre total, unidad geológica del cobre soluble, unidad geológica del cobre total y la razón de solubilidad. En búsqueda de datos anómalos (*outliers*) se presentan en la figura 5.38 los histogramas para cada uno de ellos.

Se observa que todas las muestras cuentan con información del largo. El largo de compósito es de 10 metros, sin embargo hay un grupo de muestras cuyo largo varía entre 0 y 8 las cuales serán enmascarados (restringidos). Analizando el Cobre Soluble (CuS) se observa que todas tienen información de CuS y están dentro de un rango normal (encerrado en rojo están los valores sobre 1% los cuales **no** necesariamente se consideran anómalos). Lo mismo sucede al analizar el Cobre total (CuT), es decir, están todas las muestras informadas dentro de un rango normal y encerrado en rojo están los valores sobre 1.5% los cuales no necesariamente son aberrantes.

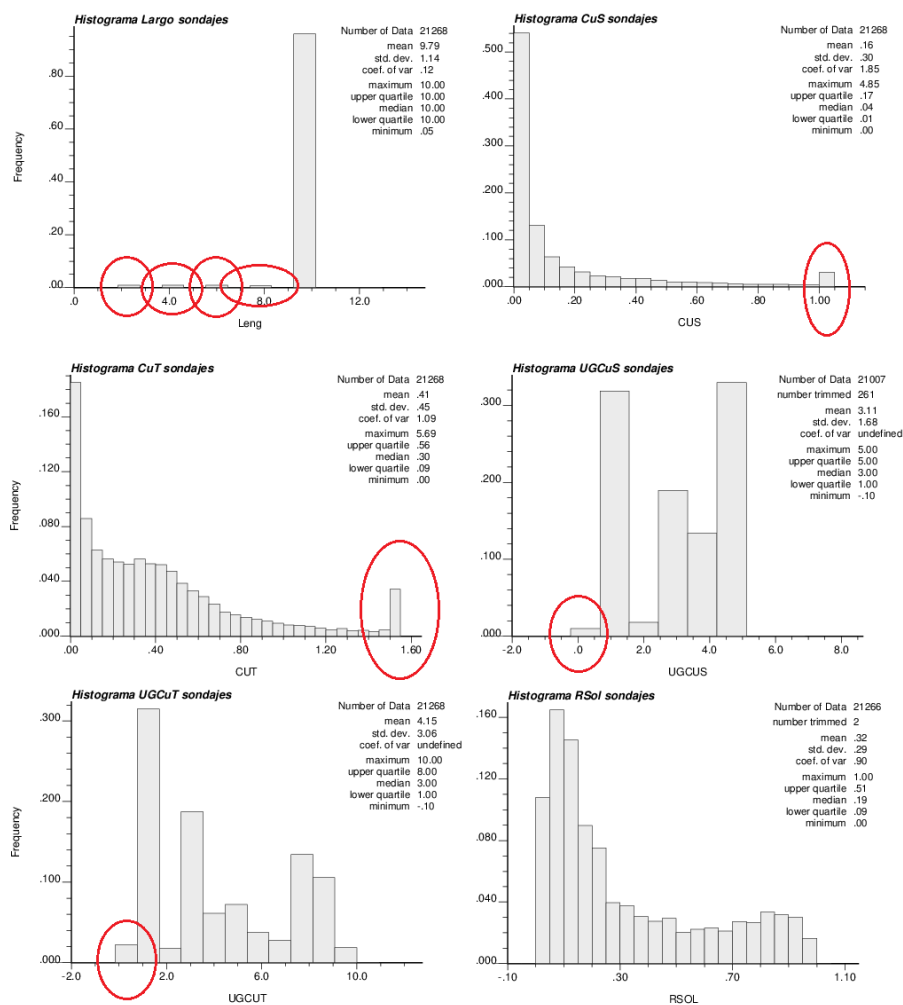


Figura 5.38: Histogramas y estadísticas básicas para cada propiedad en las muestra de sondajes

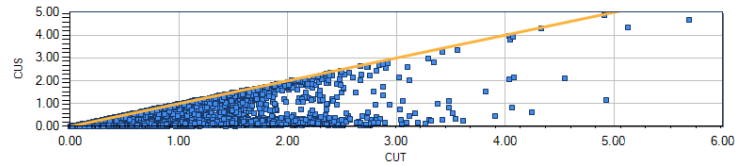


Figura 5.39: Gráfico de dispersión CuT vs CuS en los sondeos de exploración

Se observa en la figura 5.39 que todos los valores para el Cobre soluble son menores o iguales al valor del Cobre total sobre la misma muestra lo que está acorde al comportamiento mineralógico esperado. Al realizar un estudio según unidades geológicas (UG) se tiene: al agrupar los datos de CuS según la procedencia de UG cinco grupos quedan conformados y 261 muestras quedan codificadas como -0.1, es decir, sin pertenecer a una UG en particular las cuales son luego enmascaradas. Al analizar la base de datos según CuT se observan 10 UGs principales y el total de las muestras está informada. En las UGCuT los óxidos son representados por las UG 2 y 3 mientras que los sulfuros por las UG 4 y 5. A criterio del autor serán analizadas estas 4 UGs para la aplicación de las simulaciones. Enmascarando los datos aberrantes se presenta en la tabla 5.7 el resumen de las estadísticas básicas analizadas para cada Unidad geológica del Cobre total en óxidos y sulfuros.

Tabla 5.7: Sondeos - Estadísticas básicas

UGCuT	2	3	4	5
Cantidad	375	3935	1288	1520
Promedio	0.538	0.737	0.852	0.798
Mínimo	0.076	0.009	0.014	0.017
Máximo	2.516	5.692	4.548	4.926
Desviación estandar	0.334	0.503	0.573	0.607
Varianza	0.111	0.253	0.328	0.368

En la figura 5.40 se presentan las vistas en planta de las 4 unidades geológicas (UGCuT) para dar una noción de la continuidad espacial de los datos informados (se elige presentar el Cobre total).

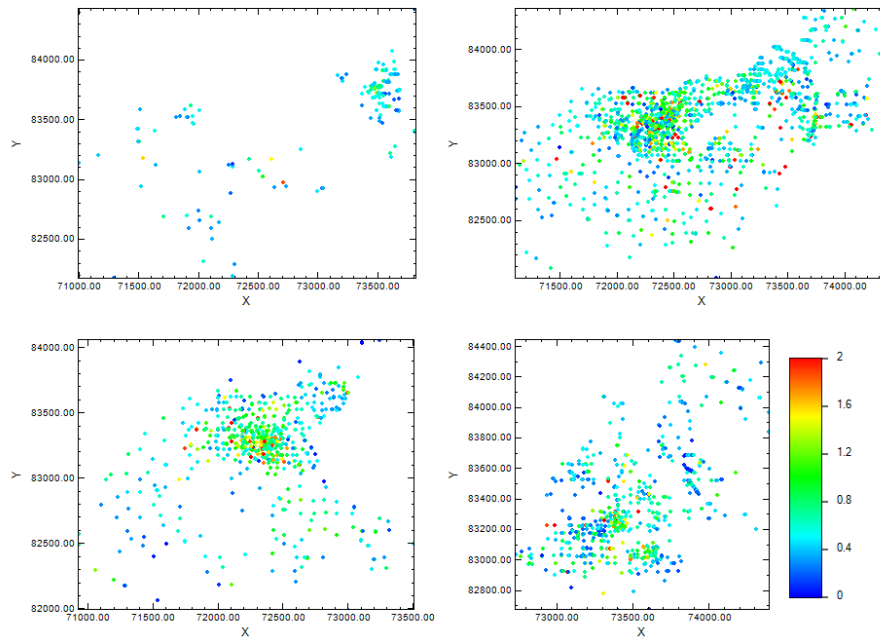


Figura 5.40: Sondajes compositados a 10 m - Vista en planta. (Izq. superior) UGCuT 2, (Der. superior) UGCuT 3, (Izq. inferior) UGCuT 4, y (Der. inferior) UGCuT 5

Las vistas en planta, de perfil e isométricos se encuentran en la sección de Anexos G.

Pozos de tronadura

Un análisis similar al de los sondajes se aplica sobre los pozos de tronadura. La base de datos cuenta con 183313 puntos informados los cuales corresponden a pozos de tronadura extraídos banco a banco. Las propiedades medidas son: largo del pozo (leng), calidad de las arcillas (malas, regulares y buenas) (Arc), Cobre soluble (CuS), Cobre total (CuT), densidad seca (Dseca), Unidad geológica del Cobre soluble (UGCUS), Unidad geológica del Cobre total (UGCUT) y la razón de solubilidad (Rsol). El histograma y las estadísticas básicas asociadas a cada propiedad se presentan en la figura 5.42.

Se observa en la figura 5.42: todos los pozos cuentan con información del largo siendo este de 10 metros, sin embargo un pequeño grupo presenta largos de 20 metros y otros casi nulos. Estos son enmascarados por posible error en la toma de información. Respecto a las arcillas, 607 pozos no cuentan con información al respecto. Un gran porcentaje de muestras se codifica con 0 ya que se desconoce su clasificación respecto a su UG. El resto se divide similar entre las arcillas malas (1,2,3 y 9), regulares (4 y 5) y buenas (6, 7 y 8). Respecto al Cobre total y Cobre soluble todos los pozos tienen información y los valores se presentan en un rango normal. Respecto a la densidad seca hay 38 datos con densidad igual a 0 (encerrado en rojo) lo cual no es coherente por lo que se decide enmascararlos. Analizando la razón de solubilidad todos los pozos presentan información y están dentro del rango esperado ($[0.0 - 1.0]$). Finalmente, analizando las Unidades geológicas: en el Cobre soluble hay 181 datos no informados y el resto pertenece a una de las 5 unidades (1,2,3,4 o 5). Por su parte el Cobre Total presenta 20 datos no informados (que coincide en no informar la UG del CuS respectivo) mientras el resto está presente en una de las 10 UGs (0,1,2,3,4,5,6,7,8 o 9) sin embargo entre las UG 1,3,5,7 y 9 se concentra el 86 % de los pozos (representando óxidos, sulfuros y zonas arcillosas).

Analizando la gráfica de dispersión de la figura 5.41 se observa que un solo valor presenta un $CuS > CuT$ lo cual es mineralógicamente imposible por lo que este dato es enmascarado.

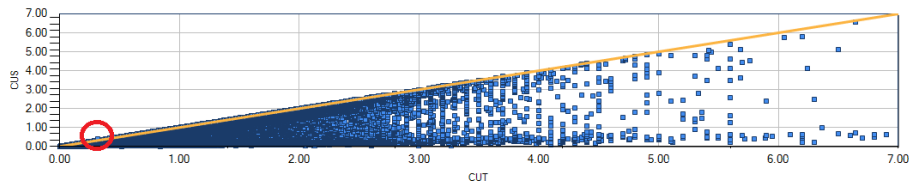


Figura 5.41: Gráfico de dispersión CuT vs CuS en los pozos de tronadura

Bajo el mismo criterio utilizado en los sondeos se decide analizar las Unidades geológicas asociadas a los óxidos (2 y 3) y sulfuros (4 y 5). Se resume en la tabla 5.8 las estadísticas básicas para UGCuT 2-3-4-5.

Tabla 5.8: Pozos de tronadura - Estadísticas básicas

UGCuT	2	3	4	5
Cantidad	10,878	68,797	8,791	28,679
Promedio	0.624	0.748	0.927	0.898
Mínimo	0.020	0.001	0.010	0.010
Máximo	4.150	13.700	6.640	8.080
Desviación estandar	0.375	0.562	0.627	0.590
Varianza	0.140	0.316	0.393	0.348

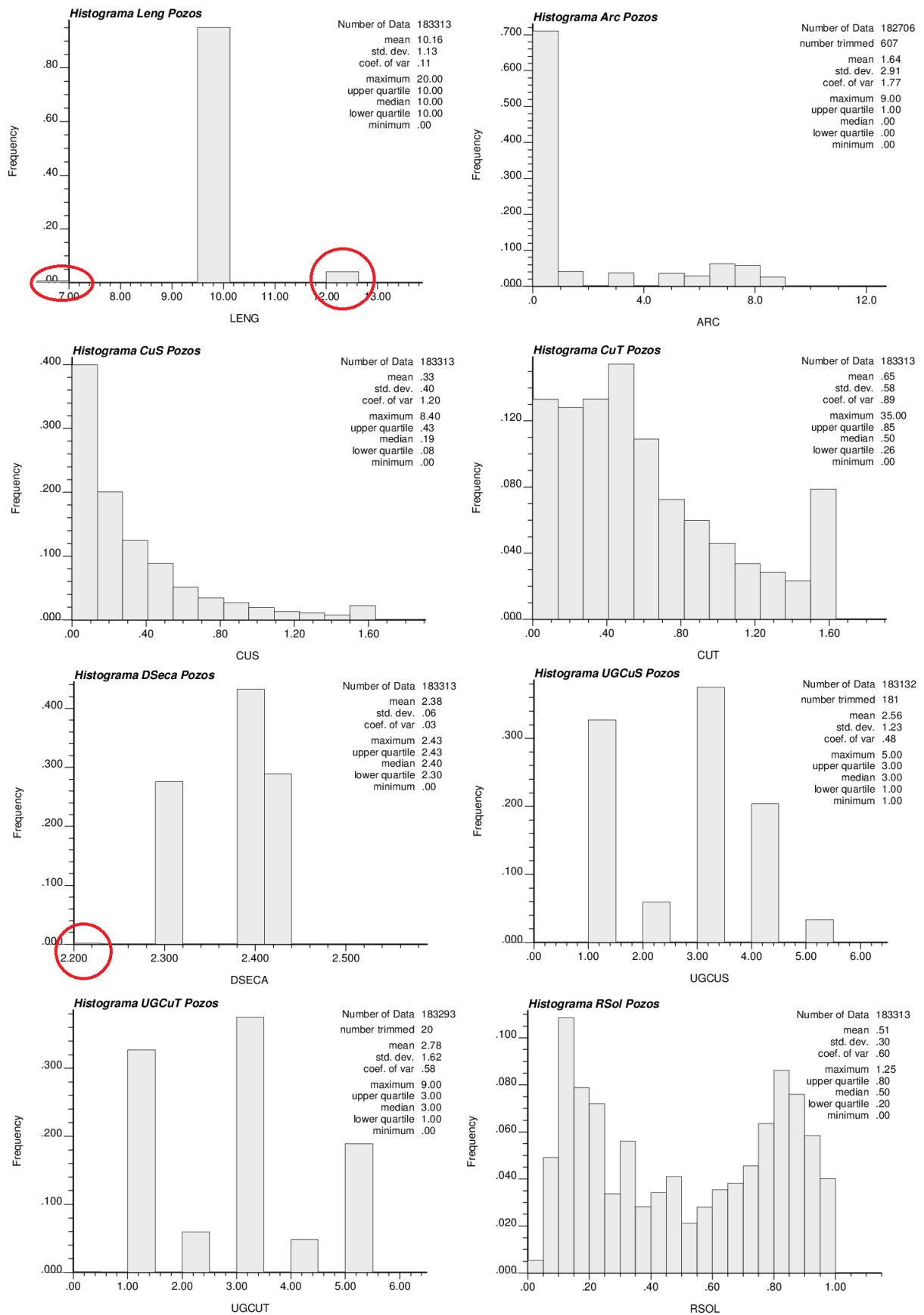


Figura 5.42: Histogramas y estadísticas básicas para cada propiedad en los pozos de tronadura

Del mismo modo que en los sondajes, se presenta en la figura 5.43 las vistas en planta de las UGCuT 2-3-4-5 para un próximo análisis de la continuidad espacial (un criterio parcial de análisis visual **no** reemplaza métricas de análisis de continuidad espacial como el *variograma*).

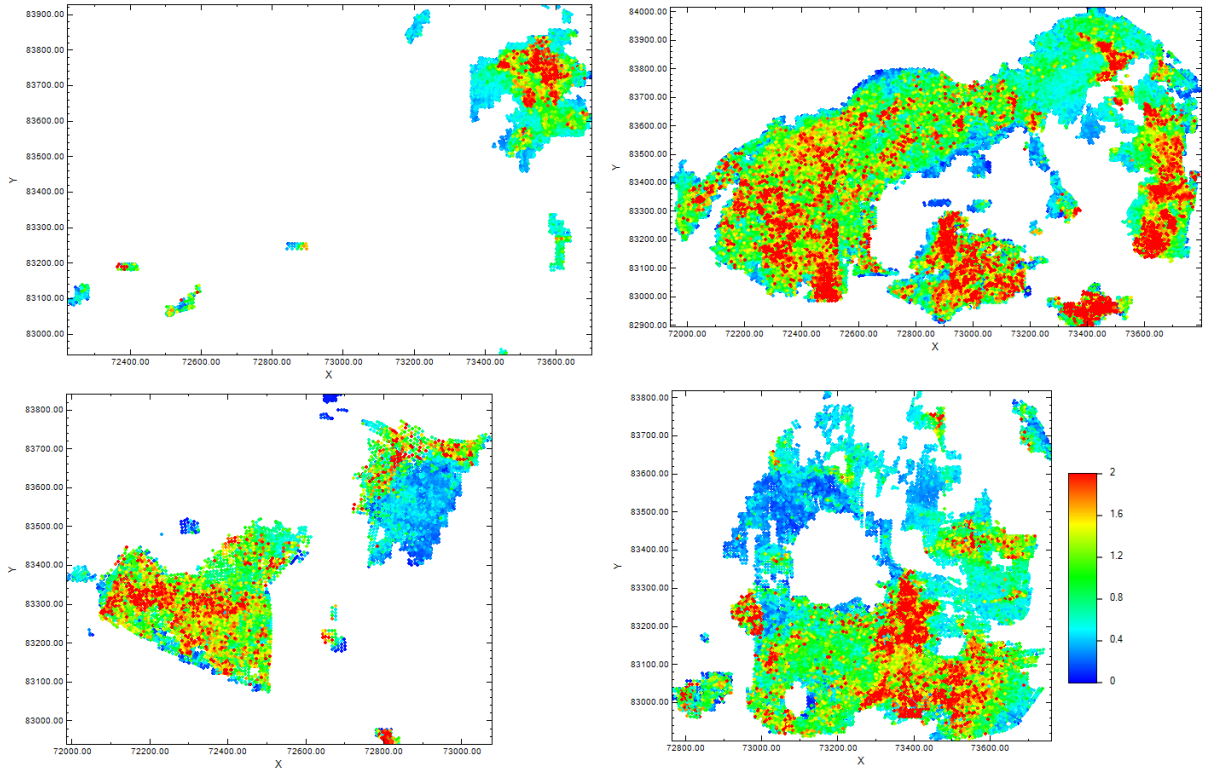


Figura 5.43: Pozos de tronadura - Vista en planta. (Izq. superior) UGCuT 2, (Der. superior) UGCuT 3, (Izq. inferior) UGCuT 4, y (Der. inferior) UGCuT 5

Las vistas en planta, de perfil e isométricos se encuentran en la sección de Anexos G.

5.2.5. Simulación

Imagen de entrenamiento, datos condicionantes y derivas externas

Se decide utilizar como imagen de entrenamiento la información entregada por los pozos de tronadura y como datos condicionantes la información de los sondajes de exploración. En la figura 5.43 se visualizaron en planta las 4 unidades geológicas (asociadas a óxidos y sulfuros) donde se aprecia que las unidades 2 y 4 presentan menos cantidad de datos que las unidades 3 y 5, respaldado por las estadísticas de la tabla 5.8. Dado lo anterior se decide no utilizar estas unidades para la obtención de la imagen de entrenamiento (criterio subjetivo). Queda aún escoger entre las unidades 3 (óxidos) y 5 (sulfuros). A criterio del autor se utiliza la unidad 3 por poseer aproximadamente 2.4 veces más datos informados que la unidad 5 dejando abierta la opción futura de aplicación sobre las unidades restantes.

Imagen de entrenamiento

Habiendo definido la unidad 3 como unidad de estimación (simulación) se procede a establecer la cota desde la cual se infiere la imagen de entrenamiento. Esta T.I. será la utilizada en la simulación de leyes de Cobre Total (CuT) en los bancos inferiores. La cota del banco escogido debe poseer una cantidad aceptable de puntos informados que den consistencia a la inferencia de patrones (confección de la T.I.). Emery y Lantuéjoul [11] sugieren utilizar una T.I. lo más grande posible (de preferencia infinita) para capturar por completo la distribución de la función aleatoria. Además, la distribución espacial que tienen los datos en el banco debe procurar ser continua con la finalidad de no escoger un sector con pequeños sectores interiores no informados.

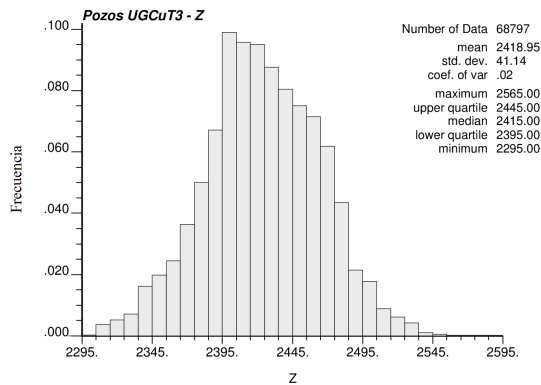


Figura 5.44: Histograma de valores en Z - Pozos de tronadura UGCuT 3

En el histograma de la figura 5.44 se aprecia que el banco con mayor cantidad de información es el **2395**. La figura 5.45 presenta la vista en planta del Banco 2395 donde se aprecia una clara continuidad espacial en 2 sectores preferenciales. Dado que la densidad de información es mayor en el primer sector que en el segundo, se elige el Sector 1 para extraer la T.I.

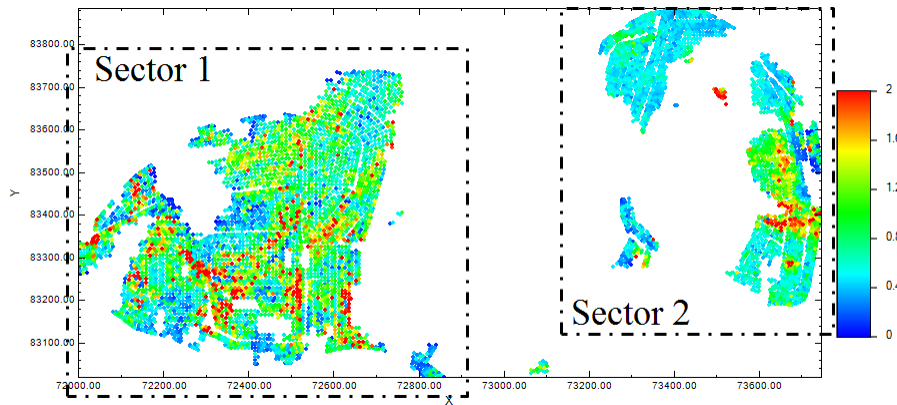


Figura 5.45: Vista en planta Banco 2395 - Pozos de tronadura UGCuT 3

La grilla que describe la imagen de entrenamiento es de 10x10x10 (basado en la altura de los bancos). En la figura 5.46 se aprecia la grilla sobre el sector 1 y en la tabla 5.9 la descripción de la grilla (imagen de entrenamiento).

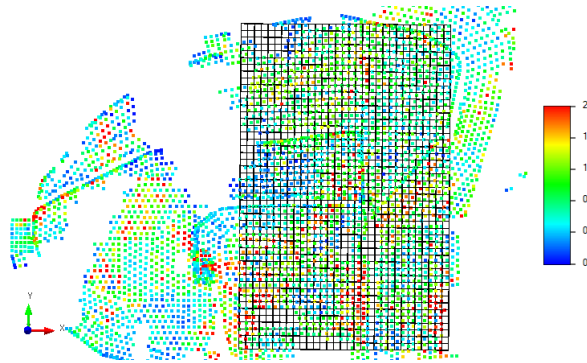


Figura 5.46: Vista en planta Banco 2395 - Sector 1 - Grilla T.I.

Tabla 5.9: Grilla asociada a la imagen de entrenamiento

Orientación	Número de nodos	Comienzo	Tamaño
X	30	72340	10
Y	47	83150	10
Z	1	2390	10

Habiendo definido la grilla de la T.I. y la zona del Sector 1 se procede a completar la grilla. Primero se migra a cada nodo el valor informado más cercano que no sobrepase los 5 metros de distancia (5 metros equivalentes a la mitad del tamaño de bloque). Con lo anterior, 612 son los nodos finalmente informados de un total de 1410. Dado que la T.I. debe estar completamente informada se vuelve necesario completar la T.I. estimando los valores no informados (57% de los nodos en la grilla). Se decide utilizar el *Inverso de la distancia* para la estimación de estos valores. Los parámetros asociados son: potencia 2, mínimo y máximo de 1 dato, radio de búsqueda de 150 metros (15 nodos). Finalmente la imagen de entrenamiento es obtenida tal como se presenta en la figura 5.47.

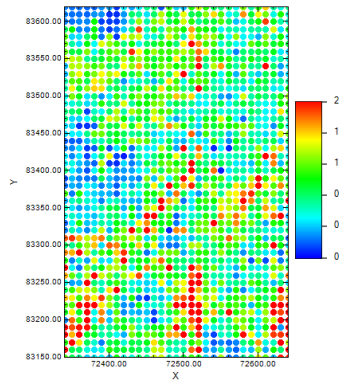


Figura 5.47: Imagen de entrenamiento obtenida del Sector 1 - Banco 2395

A modo de validar la representatividad de la T.I. se genera el gráfico Cuantil-Cuantil entre el Cobre Total del Banco 2395 y el Cobre Total en la imagen de entrenamiento. Se observa en la figura 5.48 como el gráfico describe una línea aproximadamente recta interpretando así que ambas distribuciones tienen alto grado de similitud (cerca a la diagonal).

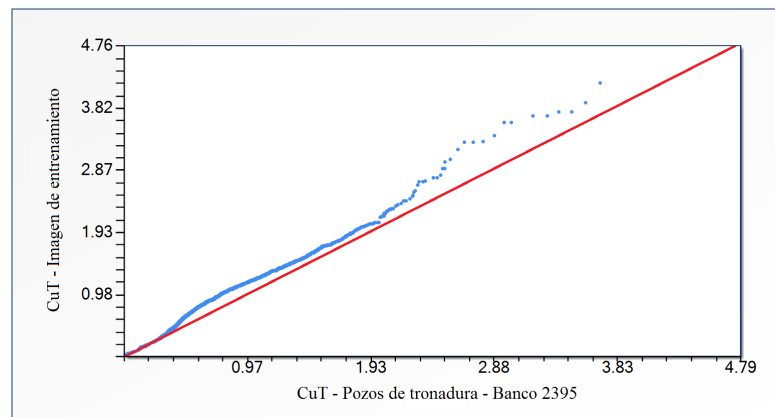


Figura 5.48: Gráfico Cuantil-Cuantil entre el Banco 2395 - UGCuT y la imagen de entrenamiento

Datos condicionantes

Dado que la imagen de entrenamiento fue obtenida a partir del banco 2395 y que la altura de banco es de 10 metros, se simulará la ley de Cobre total para los bancos 2385, 2375, 2365 y 2355. Son los sondeos los considerados como datos condicionantes. Para lograr una adecuada interpretación de estos, se debe establecer el sector a simular y con ello la información de los sondeos a utilizar.

Comenzando con el análisis del banco 2385 se grafica la vista en planta (figura 5.49) tanto para sondeos como pozos en todo el sector. Dado que los resultados de la simulación se contrastarán con la información medida en los pozos, se debe buscar un sector lo más altamente informado tanto con sondeos como con pozos. Se observa un grupo de sondeos informados alejados de la zona central y dos zonas predominantes con una alta densidad de información.

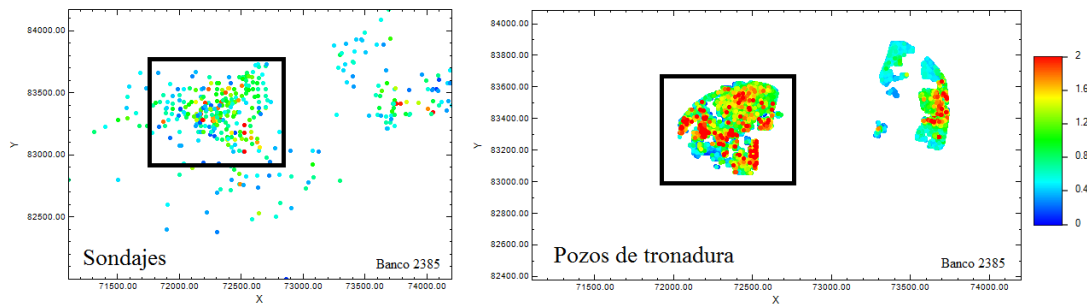


Figura 5.49: Vista en planta Banco 2385. (Izq) Sondajes - (Der) Pozos de tronadura

En la figura 5.49 se observa que una mayor cantidad de información puede ser extraída del sector delimitado en el cuadrado negro. Este a su vez coincide con estar relativamente por debajo del sector de extracción de la T.I. lo cual es coherente al suponer un comportamiento similar de mineralización. Al realizar un acercamiento al cuadrado negro se tienen las vistas en planta presentes en la figura 5.50 donde (a) sondaje y (b) pozos de tronadura. Utilizando una grilla de 10x10x10 se migran los datos a una grilla regular tal como se observa en (c) y (d). Para la migración se utiliza el estimador por Vecino más cercano con radios (6,6,10) considerando solo los sondajes en la banda $2380 \leq z < 2390$.

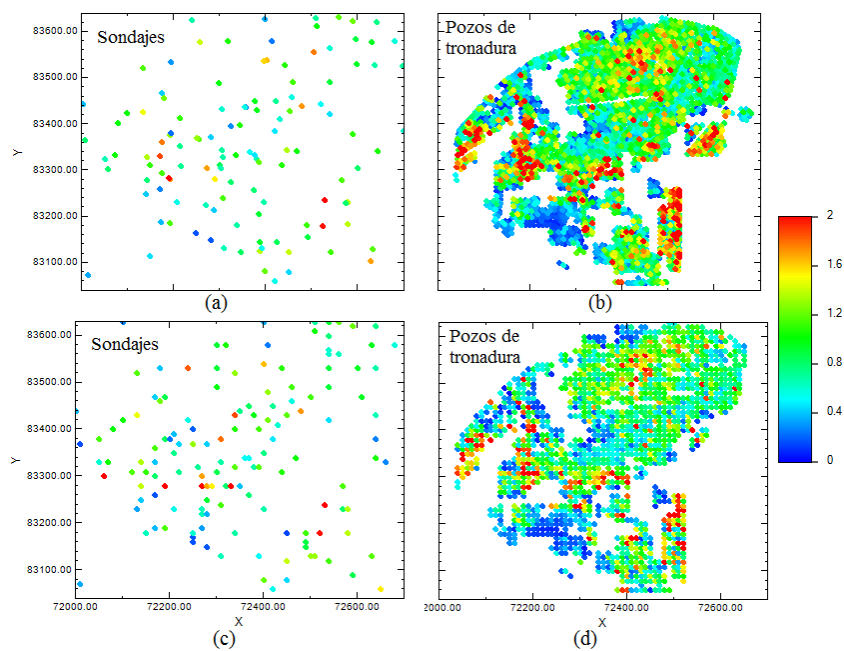


Figura 5.50: Datos condicionantes y pozos de tronadura antes y despues de migrar a grilla de 10x10x10 - Banco 2385

Se procede de igual forma para el banco 2375 con los datos condicionantes de sondajes y datos reales de los pozos de tronadura (figura 5.51).

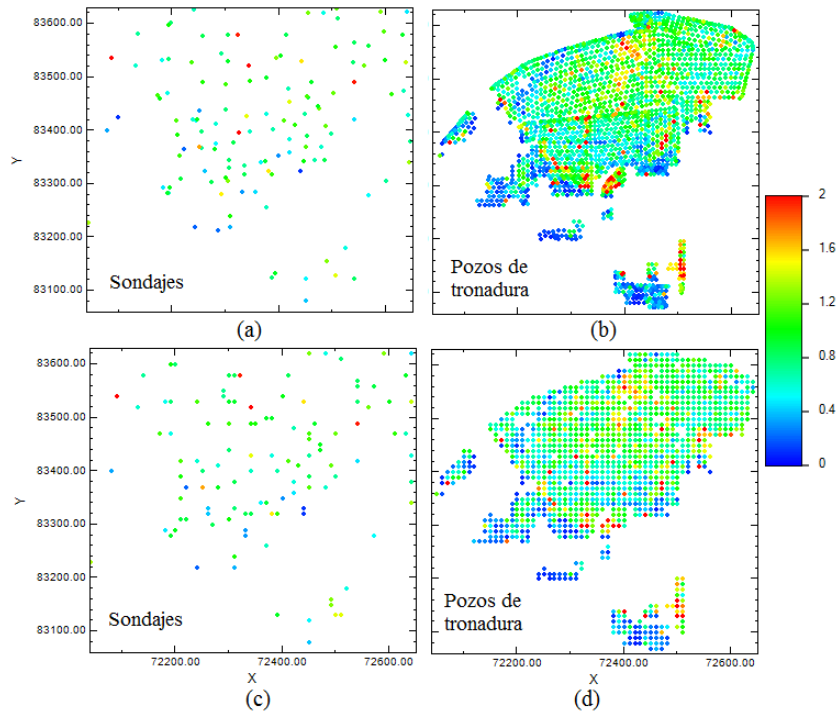


Figura 5.51: Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2375

De la misma manera se procede con los datos condicionantes de sondajes y datos reales de los pozos de tronadura del banco 2365 (figura 5.52).

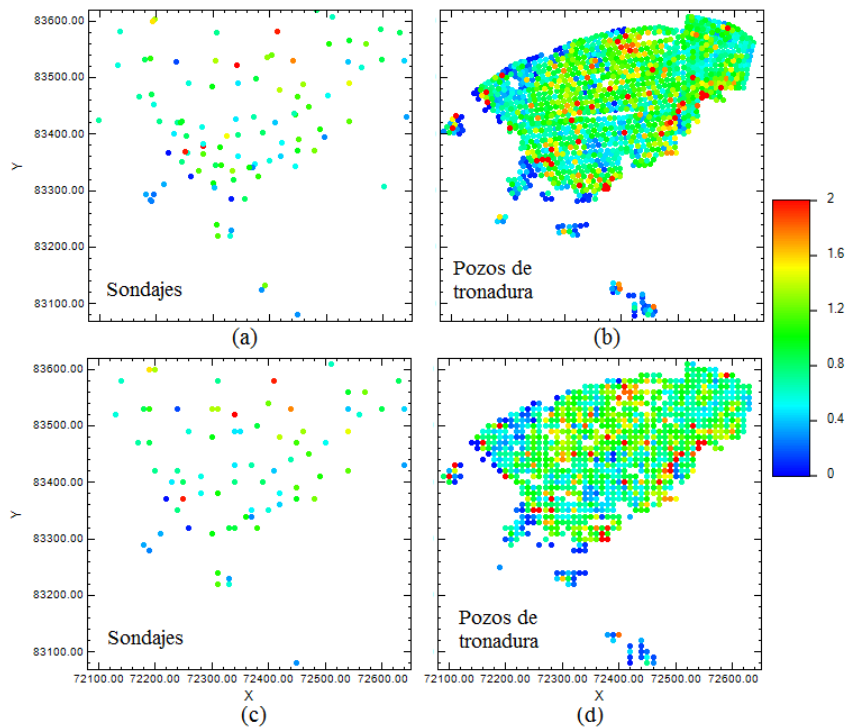


Figura 5.52: Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2365

Finalmente se repite el proceso para el banco 2355 con los datos de sondajes y datos reales de los pozos (figura 5.53).

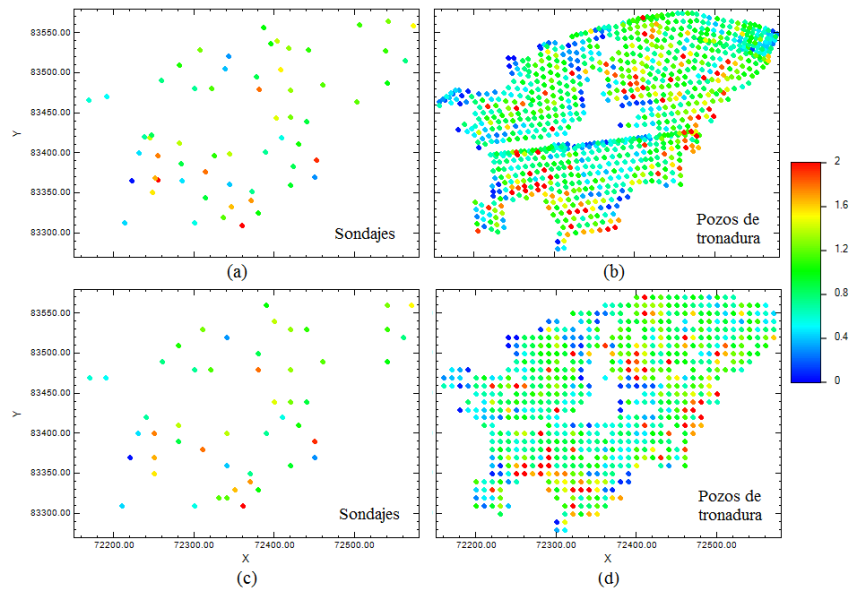


Figura 5.53: Datos condicionantes y pozos de tronadura ante y despues de migrar a grilla de 10x10x10 - Banco 2355

A modo de resumen se presenta en la tabla 5.10 las estadísticas básicas de los datos condicionantes asociados a cada banco a simular.

Tabla 5.10: Estadísticas básicas - Datos condicionantes - Bancos 2385, 2375, 2365 y 2355

Métricas	Banco 2385	Banco 2375	Banco 2365	Banco 2355
Cantidad	143	115	76	50
Mínimo	0.135	0.097	0.025	0.031
Promedio	0.897	0.885	0.828	1.065
Máximo	2.378	2.488	2.302	2.348

En la tabla 5.11 se resumen las características asociadas a las grillas de simulación para cada banco. En particular se detallan las coordenadas y cantidad de bloques en cada dirección

Tabla 5.11: Parámetros grillas de simulación

	Coordenada	Min	Max	# Bloques
Banco 2385	X	72000	72690	69
	Y	83040	83630	59
	Z	2380		1
Banco 2375	X	72040	72650	61
	Y	83060	83630	57
	Z	2370		1
Banco 2365	X	72080	72650	57
	Y	83070	83620	55
	Z	2360		1
Banco 2355	X	72150	72580	43
	Y	83270	83580	31
	Z	2350		1

Derivas externas

Sobre cada grilla a simular se debe definir la deriva en la media local que se impondrá en el algoritmo Filtersim modificado. Para ello se analiza primero qué dirección (X o Y) presenta una deriva en la media con mayor notoriedad. Dado que la imagen de entrenamiento fue extraída en el Banco 2395 a partir de los pozos, se analiza la deriva en la media para el mismo banco. Para ello si se desea determinar la deriva en X se elige una posición en específico y se promedian todos los pozos de tronadura con igual coordenada X. Lo mismo para la deriva en Y. Cabe notar que son los datos migrados a la grilla regular los utilizados ya que de otra manera la deriva no sería representativa por posible sobre muestreo. En la figura 5.54 se observan las derivas tanto en X como Y.

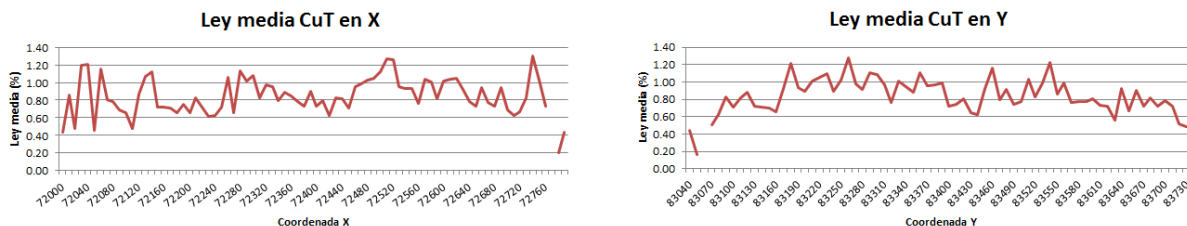


Figura 5.54: Deriva en la media en el Banco 2395 para las coordenadas X e Y

Se observa que ambas direcciones tienen variaciones en la media. Al realizar un análisis en la coordenada X se aprecia una variación en la varianza de la misma y una tendencia al alza a medida que aumenta la coordenada X. Por su parte, la deriva en la coordenada Y presenta un aumento hacia la mitad de las coordenadas y luego una leve disminución. A criterio del autor se decide estudiar sólo las derivas en la coordenada X dejando como propuesto un análisis en Y o ambos incluidos.

Para los 4 bancos en estudio se utilizaron los datos condicionantes (descritos en la sección anterior) para la inferencia de la deriva en la media. Para ello se calcula el promedio de los valores, ya migrados a una celda regular, de los sondeos que poseen las mismas coordenadas en X. Dada la poca densidad de datos condicionantes existen coordenadas en X que no presentan valor promedio ya que hay ausencia de datos. Para suplir esta información se decide tomar el valor promedio del valor entregado por las 2 coordenadas más cercanas (una a la izquierda y otra a la derecha). Si una coordenada no informada está próxima a uno de los bordes, en la grilla a simular, es probable que sólo una coordenada cercana esté informada (la más próxima a la derecha si está cerca del borde izquierdo o la más próxima a la izquierda si está cerca del borde derecho) y en este caso toma directamente este valor (lo *copia*). Cabe notar que este último criterio para definir las celdas no informadas es un criterio subjetivo y puede modificar los resultados si se utiliza otra metodología.

Habiendo definido el criterio a utilizar para definir las derivas en la media, se presenta en la figura 5.55 las derivas finales a utilizar en cada banco.

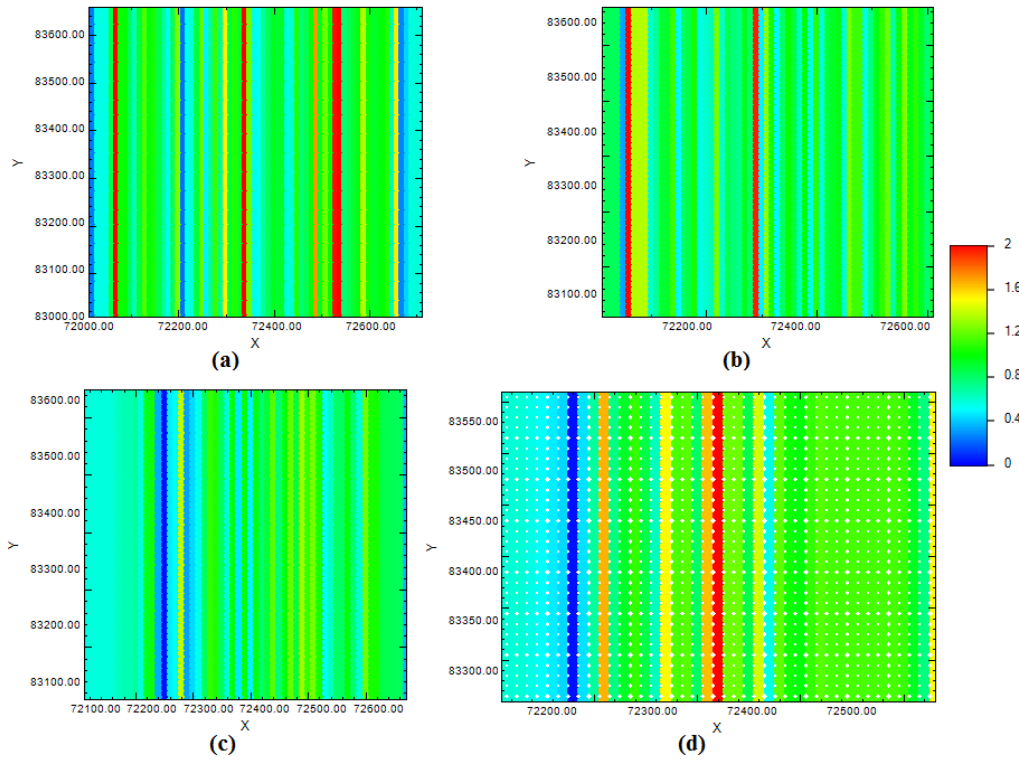


Figura 5.55: Derivas externas inferidas desde los sondajes. Bancos: (a) 2385, (b) 2375, (c) 2365 y (d) 2355.

Variogramas tras anamorfosis Gaussianas

Para la aplicación de la Simulación Secuencial Gaussiana (SGSim) se requiere que la variable en estudio presente una distribución normal. Dado que la distribución de la ley de Cobre total no sigue un carácter Gaussiano, se determina la función de transformación (anamorfosis) al transformar el CuT del banco 2395 a una variable Gaussiana (ver figura 5.56).

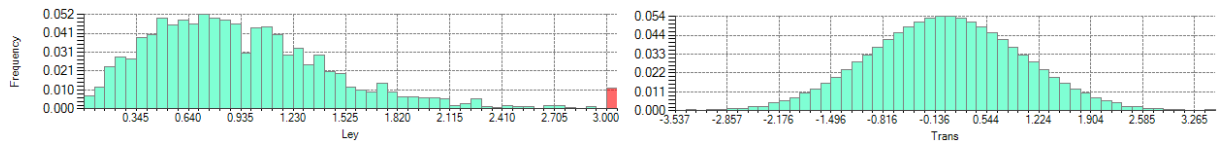


Figura 5.56: Histograma de las leyes de cobre en la imagen de entrenamiento (izquierda) y de las leyes luego de aplicar la anamorfosis (derecha)

Una vez transformado los datos (y desagrupados si fuese necesario) se infiere el variograma experimental para luego ser modelado (banco a banco). Los parámetros de entrada para la determinación de estos se detallan en la tabla 5.12 mientras que la cantidad de datos utilizado es equivalente a la cantidad de muestras de sondajes migrados en la grilla (ver tabla 5.10).

Luego de obtener los variogramas experimentales estos deben ser modelados a través de estructuras conocidas que cumplen con las condiciones para resolver el sistema de Kriging. Los variogramas modelados se presentan en las ecuaciones (5.9).

$$\begin{aligned}
 \gamma_{2385} &= 0.2 + 0.8exp(80, 80) \\
 \gamma_{2375} &= 0.15 + 0.75esf(40, 40) + 0.2exp(120, 120) \\
 \gamma_{2365} &= 0.1 + 0.75exp(30, 30) + 0.15esf(90, 90) \\
 \gamma_{2355} &= 0.1 + 0.9exp(30, 30)
 \end{aligned}
 \tag{5.9}$$

Tabla 5.12: Parámetros para cálculo del variograma experimental

	Banco 2385	Banco 2375	Banco 2365	Banco 2355
Azimuth	0	0	0	0
Tolerancia Azimuth	90	90	90	90
Dip	0	0	0	0
Tolerancia Dip	90	90	90	90
Ancho de banda Horizontal	150	150	150	150
Ancho de banda Vertical	150	150	150	150
Número de pasos	33	30	28	21
Largo del paso	10	10	10	10
Tolerancia al paso	10	10	10	10

La figura 5.57 presenta los variogramas tanto experimentales como modelados para los 4 bancos en estudio.

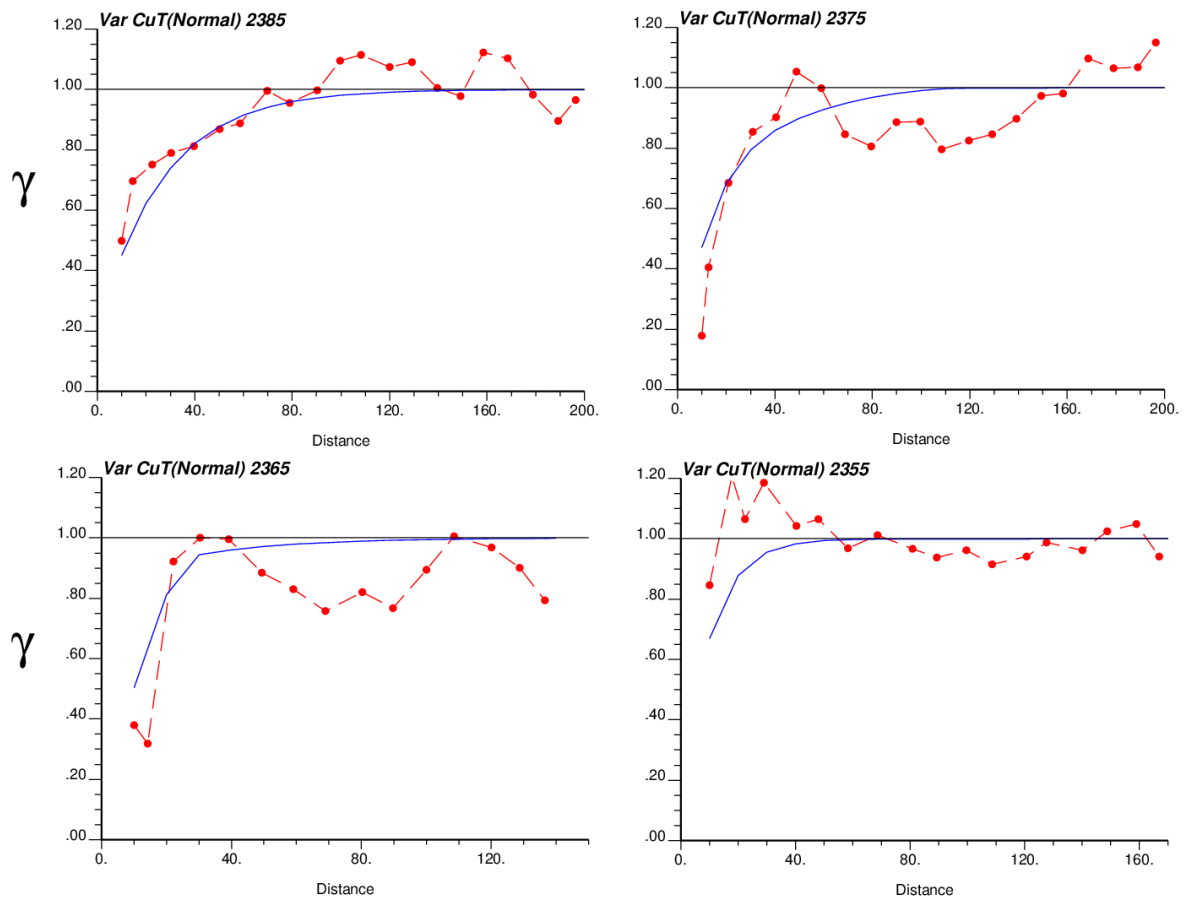


Figura 5.57: Variogramas experimentales en los bancos: 2385, 2375, 2365 y 2355

Descripción de los escenarios (subcasos) de estudio

A continuación se presentan los parámetros de entrada para la aplicación de los algoritmos de simulación en los 3 escenarios a estudiar.

Caso base

El caso base corresponde a la aplicación del algoritmo de Simulación Secuencial Gaussiana. Los parámetros de entrada son idénticos para los 4 bancos cambiando solamente el variograma y los datos condicionantes. Cabe recordar que la simulación utiliza como inferencia de la continuidad espacial los variogramas sobre los datos ya transformados a una distribución Gaussiana. Los parámetros comunes se presentan en la tabla 5.13.

Tabla 5.13: Parámetros Simulación Secuencial Gaussiana

Realizaciones	100	Nivel de <i>debugging</i>	1
Extrapolación de cola: min	0.0	Semilla	69069
Extrapolación de cola: max	2.5	Mínima cantidad de datos condicionantes	1
<i>Lower tail option</i> y parámetro	1 y 0.0	Máxima cantidad de datos condicionantes	8
<i>Upper tail option</i> y parámetro	1 y 1.5	Cantidad de datos previamente simulados a usar	12.0
Multigrilla	2	Uso de octantes	No
Máximos radios de búsqueda	120, 120 y 120	Tipo de Kriging	Simple
Tamaño matriz de covarianzas	51, 51 y 11	-	-

Caso F1

El caso F1 es la aplicación de Filtersim sin la incorporación de derivas. Los parámetros de entrada son idénticos para los 4 bancos en estudio dado que la diferencia es la grilla de simulación, los datos condicionantes y las derivas en la media pero no en los parámetros de entrada. Se generan 100 realizaciones para tener un conjunto suficientemente informado al momento de calcular los mapas de probabilidad. Se decide utilizar una grilla interna del mismo tamaño que la grilla de búsqueda ambas cuadradas de lado 70 (sólo 7 celdas por lado). Se decide tomar este tamaño para que la cantidad de patrones desde la imagen de entrenamiento sea suficientemente representativo. Dos serán los niveles de multigrilla. El ponderador de la deriva externa es nulo. La cantidad de *clusters* en el proceso de agrupamiento es de 50 (lo cual entrega un promedio a priori de 21 patrones por prototipo en el primer nivel de multigrilla a la hora de inferir los patrones desde la imagen de entrenamiento). Los ponderadores asociados a los datos condicionantes, los datos simulados en el nivel actual de grilla y los datos simulados en el nivel superior de grilla son 0.6, 0.25 y 0.15 respectivamente. En la tabla 5.14 se resumen estos parámetros.

Tabla 5.14: Parámetros de entrada Filtersim sin incluir derivas

Repeticiones	100	Mínima cantidad datos condicionantes	2
Grilla de búsqueda: largo en X	70	Ponderador Deriva (w_0)	0.0
Grilla de búsqueda: largo en Y	70	Número de clusters	50
Grilla interna: largo en X	70	Ponderador datos condicionantes (w_1)	0.6
Grilla interna: largo en Y	70	Ponderador datos simulados (w_2)	0.25
Multigrilla	2	Ponderador sim. Grilla superior (w_3)	0.15

Caso F2

El caso F2 es la aplicación de Filtersim incorporando derivas en la media. Tal como en el caso F1 los parámetros de entrada son idénticos para los 4 bancos en estudio. Se generan 100 realizaciones. La grilla interna es del mismo tamaño que la grilla de búsqueda, ambas cuadradas de lado 70 (7 celdas por lado). Dos son los niveles de multigrilla. La deriva externa se pondera un 40 %, es decir, 0.4. La cantidad de *clusters* en el proceso de agrupamiento es de 50. Los ponderadores asociados a los datos condicionantes, los datos simulados en el nivel actual de grilla y los datos simulados en el nivel superior de grilla son 0.6, 0.25 y 0.15 respectivamente. En la tabla 5.15 se resumen estos parámetros.

Tabla 5.15: Parámetros de entrada Filtersim incluyendo derivas

Repeticiones	100	Mínima cantidad datos condicionantes	2
Grilla de búsqueda: largo en X	70	Ponderador Deriva (w_0)	0.4
Grilla de búsqueda: largo en Y	70	Número de clusters	50
Grilla interna: largo en X	70	Ponderador datos condicionantes (w_1)	0.6
Grilla interna: largo en Y	70	Ponderador datos simulados (w_2)	0.25
Multigrilla	2	Ponderador sim. Grilla superior (w_3)	0.15

5.2.6. Resultados y validación

A continuación se presentan visualmente una de las 100 realizaciones junto con el promedio de las mismas para los 3 escenarios en estudio y los 4 bancos simulados.

Caso base

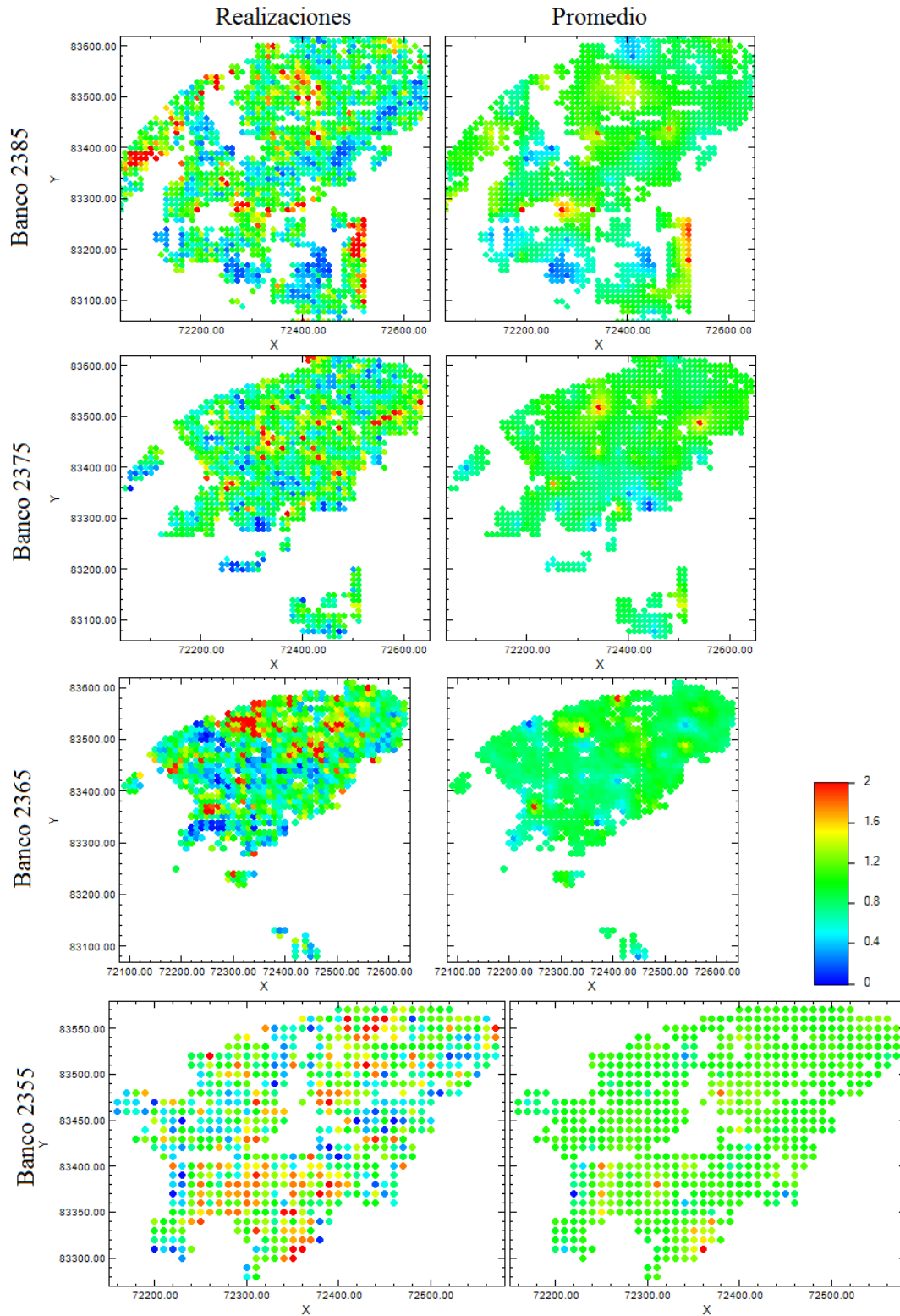


Figura 5.58: Caso base (SGSim) - Una de cien realizaciones y los promedios para los 4 bancos en estudio

Caso F1

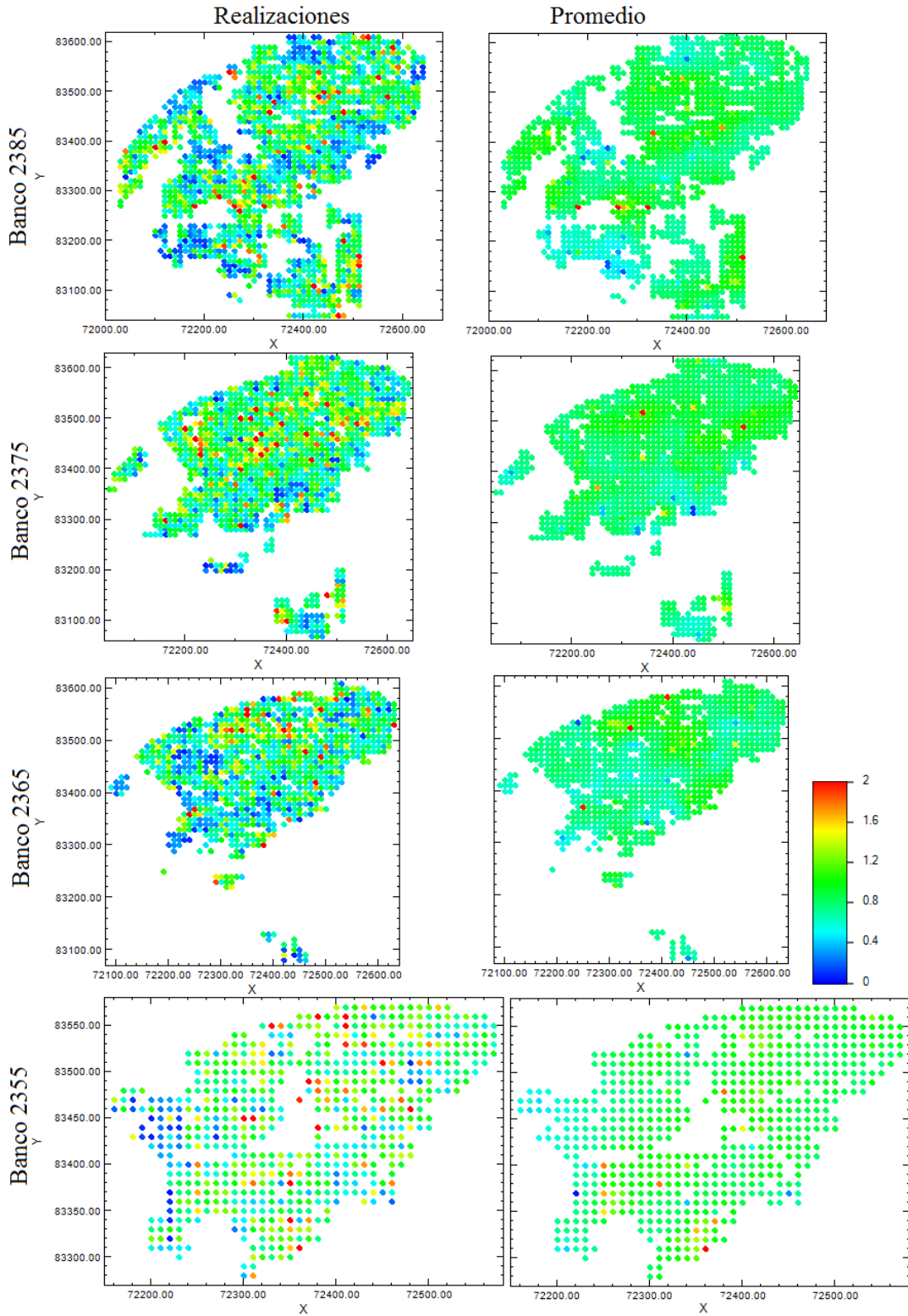


Figura 5.59: Caso F1 - Una de cien realizaciones y los promedios para los 4 bancos en estudio

Caso F2

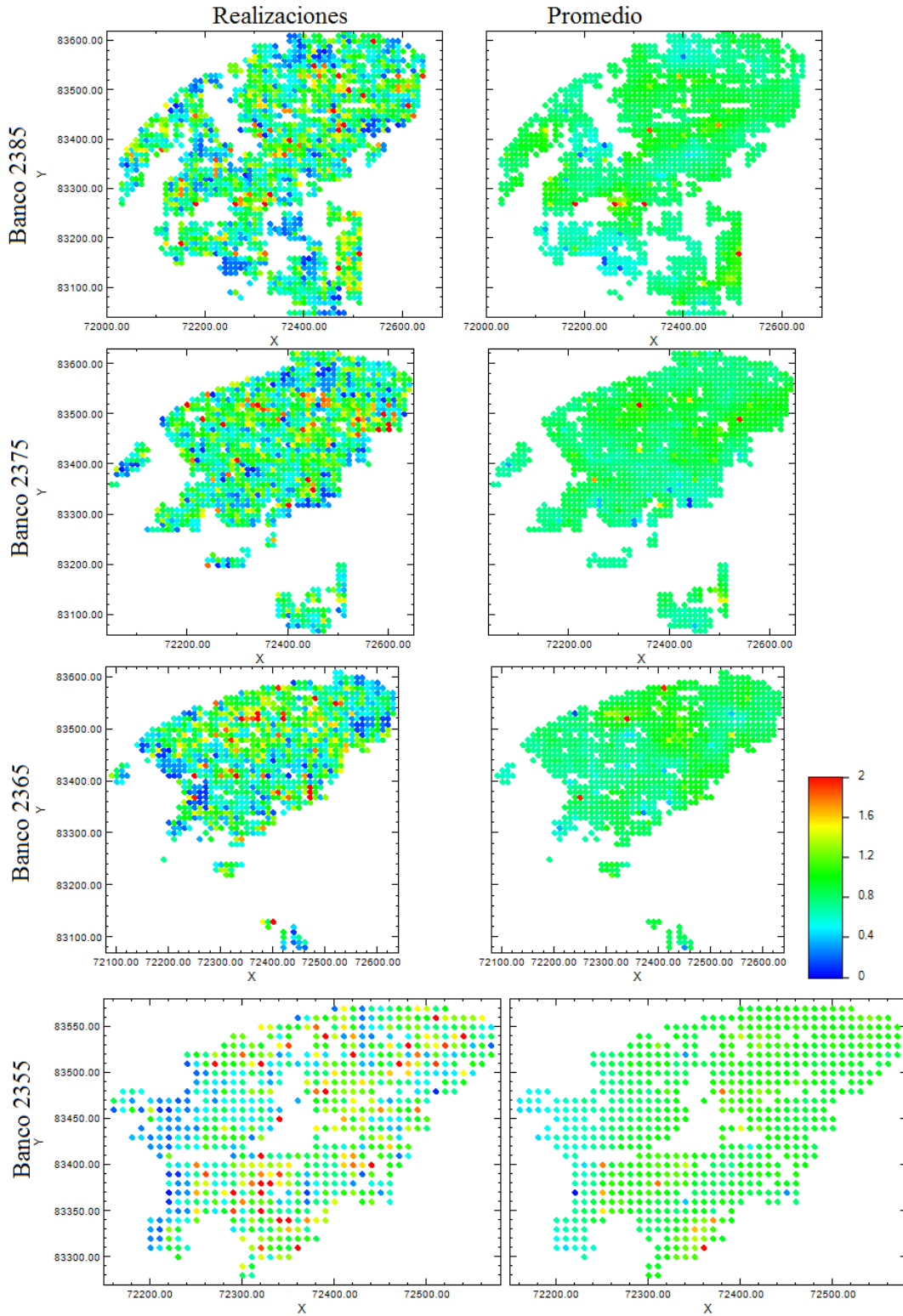


Figura 5.60: Caso F2 - Una de cien realizaciones y los promedios para los 4 bancos en estudio

Verificación de sesgo

Se entiende por *sesgo* a la diferencia entre el valor esperado del estimador y el valor real de un parámetro en estudio. Un estimador se dice insesgado cuando su sesgo es nulo. En el presente caso de estudio se calcula el sesgo que hay respecto a la **media** y a la **varianza** para cada realización en cada simulación. Los valores considerados son sólo para los que, en la misma ubicación, existe valor en los pozos de tronadura. La tabla 5.16 presenta la media y la varianza de los datos de pozos y sondajes para cada banco ya que contra ellos se calcula el sesgo de las realizaciones. En el anexo H se presentan los gráficos detallados para cada cálculo del sesgo.

Tabla 5.16: Media y varianza en las muestras de pozos y sondajes banco a banco

	Sondajes	Sondajes	Pozos	Pozos		Imagen de
	Media	Varianza	Media	Varianza		entrenamiento
Banco 2385	0.897	0.231	0.864	0.313	Mínimo	0.050
Banco 2375	0.885	0.161	0.817	0.220	Media	0.952
Banco 2365	0.828	0.197	0.887	0.238	Máximo	4.900
Banco 2355	1.065	0.229	0.926	0.272	Varianza	0.332

A modo de resumen se calcula el sesgo promedio (promedio de los 100 valores determinados por cada simulación) graficandolos en la figura 5.61 y resumidos en la tabla 5.17.

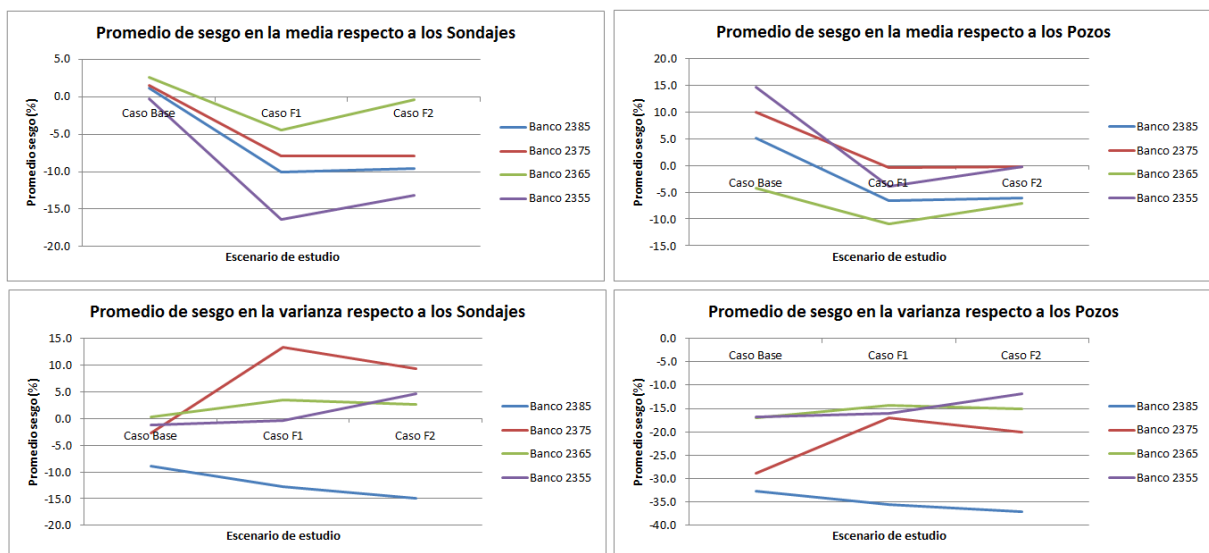


Figura 5.61: Promedio en los sesgos, banco a banco, para la media y la varianza respecto a los sondajes y pozos de tronadura

Si se comparan los resultados obtenidos al calcular el sesgo en la media respecto a los datos de sondajes se observa claramente que la Simulación Secuencial Gaussiana es más insesgada que Filtersim para los 4 bancos. Esto se debe a que la función aleatoria generadora en SGSim es inferida directamente desde los sondajes (previa anamorfosis) mientras que en los casos F1 y F2 la inferencia proviene de la T.I. del banco superior 2395. Lo mismo sucede al analizar el promedio del sesgo en la varianza, para los 4 bancos SGSim mostró el menor sesgo promedio. Por otro lado, la incorporación de derivas (caso F2) en Filtersim mejoró considerablemente el sesgo al compararlo con el caso sin deriva (caso F1) en los bancos 2365 y 2355. Lo anterior no ocurre en los primeros dos bancos inferiores (2385 y 2375). Esto lleva a interpretar que el algoritmo Filtersim modificado no entrega beneficios sustanciales, desde el punto de vista de reducir el sesgo en la media, para los primeros dos bancos inferiores mientras que sí mejora los resultados para los 2 bancos subsiguientes a ellos.

Tabla 5.17: Tabla resumen: promedio en los sesgos para la media y la varianza respecto a los sondeos y pozos

	Banco 2385			Banco 2375		
	C. Base	C. F1	C. F2	C. Base	C. F1	C. F2
Media - Sondeos	1.2	-10.0	-9.6	1.5	-7.9	-7.9
Media - Pozos	5.1	-6.5	-6.1	9.9	-0.3	-0.3
Varianza - Sondeos	-8.8	-12.8	-15.0	-2.7	13.3	9.4
Varianza - Pozos	-32.6	-35.5	-37.1	-28.9	-17.1	-20.0

	Banco 2365			Banco 2355		
	C. Base	C. F1	C. F2	C. Base	C. F1	C. F2
Media - Sondeos	2.6	-4.5	-0.3	-0.2	-16.4	-13.2
Media - Pozos	-4.3	-10.9	-7.0	14.7	-3.8	-0.2
Varianza - Sondeos	0.3	3.4	2.6	-1.3	-0.4	4.7
Varianza - Pozos	-16.9	-14.4	-15.1	-16.9	-16.1	-11.9

Realizando la misma comparación ahora respecto a los pozos de tronadura, si se analiza el banco 2355 (último banco) el promedio del sesgo en la media entregado por SGSim es de 14.7% mientras que Filtersim convencional y con incorporación de deriva es de -3.8% y -0.2% respectivamente. Esto señala lo sesgado que está el aplicar SGSim en este banco y las ventajas que tiene la aplicación de Filtersim y más aún la incorporación de derivas. El sesgo en la varianza también disminuye al pasar de SGSim al caso F2. Al analizar el banco 2365 el caso F1 empeora los resultados al compararlo con SGSim pero se mejora al incorporar derivas. Por otro lado, en los primeros dos bancos inferiores (2385 y 2375) se observa una mejora al utilizar Filtersim por sobre SGSim, sin embargo la incorporación de derivas no mejora los resultados (mismo fenómeno observado al comparar con los sondeos).

Es importante recalcar que los resultados anteriores dependen fuertemente de la *consistencia* en las estadísticas de los datos condicionantes y la imagen de entrenamiento. Esto implica que los resultados, y en particular su consistencia, mejorarán al utilizar una media local que tenga *buen* sentido físico/geológico. Todo lo anterior guarda relación con la susceptibilidad que tienen los resultados a la imagen de entrenamiento y los datos condicionantes.

Verificación de continuidad espacial

Uno de los aspectos a verificar de las realizaciones que se obtienen al simular es la reproducción de la distribución de la función aleatoria generadora. Sin embargo, la distribución espacial que tienen los valores dentro del dominio (estructura de los valores en el espacio) es igual de importante. Para cuantificar la continuidad espacial, y tal como en el caso de estudio en la Roca Sedimentaria, se utiliza el *variograma experimental*. La variografía es aplicada sobre la Imagen de Entrenamiento ((pozos de tronadura en el banco 2395) **rojo**), sobre los sondajes en el banco simulado (**negro**), sobre los pozos de tronadura en el mismo banco (**verde**) y sobre 22 realizaciones (**celeste**) aleatoriamente escogidas. Para el cálculo de los variogramas experimentales se utilizan los parámetros, banco a banco, entregados en la tabla 5.18.

Tabla 5.18: Parámetros para el cálculo del variograma experimental

	Banco 2385	Banco 2375	Banco 2365	Banco 2355
Azimuth	0	0	0	0
Tolerancia Azimuth	90	90	90	90
Dip	0	0	0	0
Tolerancia Dip	90	90	90	90
Ancho de banda Horizontal	150	150	150	150
Ancho de banda Vertical	150	150	150	150
Número de pasos	15	15	15	15
Largo del paso	10	10	10	10
Tolerancia al paso	10	10	10	10

Caso base

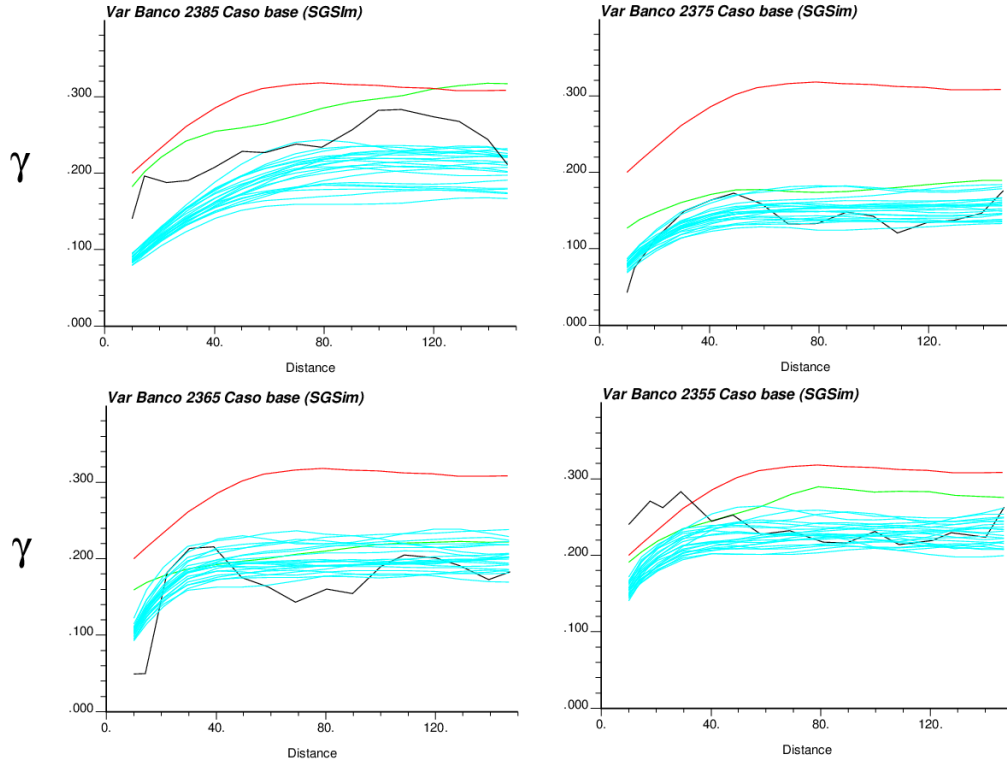


Figura 5.62: Estudio variográfico - Caso base (SGSim)

Caso F1

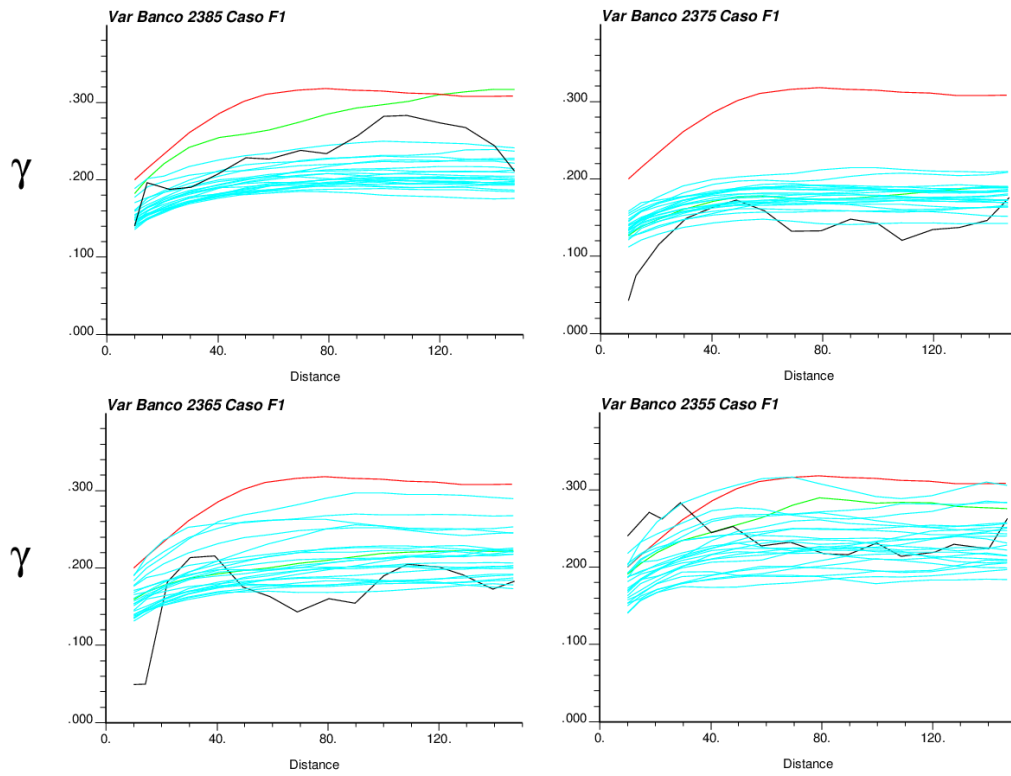


Figura 5.63: Estudio variográfico - Caso F1

Caso F2

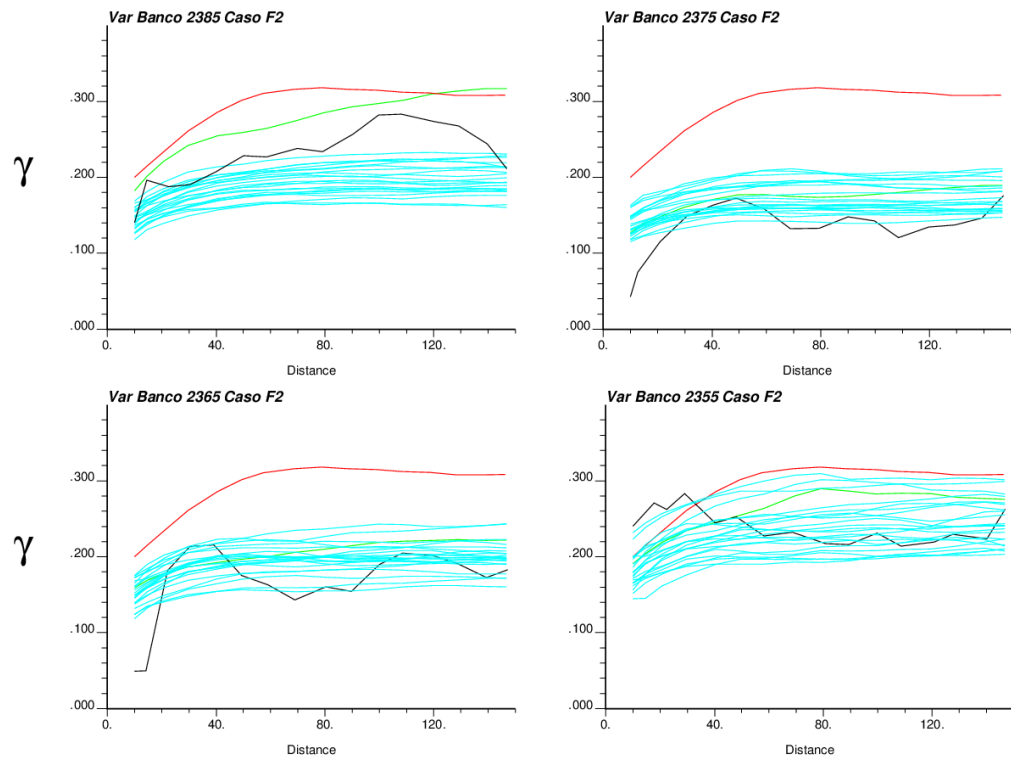


Figura 5.64: Estudio variográfico - Caso F2

Lo primero que se deduce es que, a pesar de que las simulaciones entregadas por Filtersim se basan en la información entregada por la imagen de entrenamiento, los resultados muestran que las realizaciones no siguen el mismo comportamiento variográfico que la mostrada por la T.I. sino que están fuertemente controladas por los datos condicionantes (incluso más que SGSim). Por otro lado, en el banco 2385 los resultados mostrados por SGSim no se asemejan ni al variograma de los sondeos ni a los pozos de tronadura, mientras que el caso F1 se aumenta el efecto pepita de los variogramas y se asemeja más a los sondeos que a los pozos. Si al caso F1 se le incorporan derivas se obtiene lo mostrado por el caso F2 en donde no se aprecia una mejora ni en la estructura del variograma ni en la dispersión entre las distintas realizaciones. Para los bancos 2375, 2365 y 2355 SGSim reproduce la conducta variográfica de los sondeos en cada una de las realizaciones mientras que Filtersim reproduce el comportamiento variográfico de los pozos de tronadura por sobre los sondeos (valores condicionantes). El incorporar derivas mejoró los resultados al disminuir la dispersión (amplitud) que existía entre las realizaciones. Esto último se observa en el banco 2365, sin embargo los bancos 2355 y 2375 no presentan grandes cambios.

Finalmente, las simulaciones por Filtersim entregan mayor dispersión entre los posibles escenarios que las entregados por SGSim. Además, con Filtersim/Filtersim modificado se logran estructuras variográficamente similares a las presentadas por los pozos de tronadura (valores fidedignos del banco simulado). Sin embargo, la incorporación de derivas o bien mejora la calidad de los resultados o bien no tiene mayor influencia. Cabe notar que en ningún caso se arruinan los resultados al incorporar derivas. Por todo lo anterior, la decisión sobre qué banco aplicar la técnica depende de cuán similares son respecto a la imagen de entrenamiento, a la cantidad de datos disponibles para el condicionamiento, y un conjunto de otras variables geológicas que a priori no son simples de obtener, sabiendo que en la realidad sólo se cuenta con información de sondeo en los bancos simulados.

5.2.7. Análisis de resultados

Error cuadrático medio

El Error Cuadrático Medio (ECM) es el encargado de medir la dispersión de las realizaciones respecto a la realidad (pozos de tronadura). Para ello se toman en consideración sólo los bloques simulados que tienen información de pozos de tronadura. Para cada escenario de simulación se obtiene un único valor ECM mediante la ecuación (5.10).

Sean:

- # L: número de realizaciones (100).
- # N: número de nodos informados en pozos de tronadura (relativo a cada banco en estudio).
- (u_n) : la posición de un nodo con información de pozos de tronadura ($u_n : 1 \dots N$).
- $Z_{real}(u_n)$: ley de Cobre total en la posición u_n extraído de los pozos de tronadura (valor real).
- $Z_{sim}(u_n)$: ley de Cobre total simulado en la posición u_n . Este valor varía dependiendo del caso: caso base, caso F1 y caso F2.

$$ECM_{global} = \frac{1}{L \times N} \sum_{l=1}^L \sum_{n=1}^N (Z_{real}(u_n) - Z_{sim}(u_n))^2 \quad (5.10)$$

En la tabla 5.19 se presentan los resultados para cada caso de estudio en cada banco simulado. Los mismos son graficados en la figura 5.65.

Tabla 5.19: Errores Cuadráticos Medio por banco para los distintos casos

	Caso base	Caso F1	Caso F2
Banco 2385	0.451	0.474	0.476
Banco 2375	0.339	0.376	0.370
Banco 2365	0.399	0.424	0.417
Banco 2355	0.495	0.475	0.481

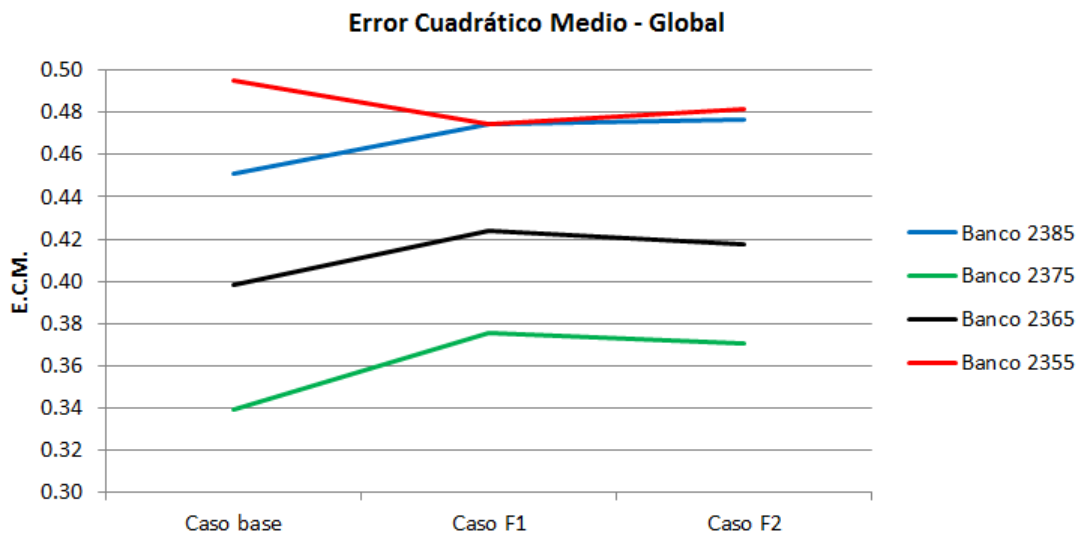


Figura 5.65: Gráfico E.C.M. banco a banco para cada escenario de estudio

En los primeros tres bancos (2385, 2375 y 2365) se observa un comportamiento similar, presentado más claramente en la figura 5.66. El ECM entregado por la Simulación Secuencial Gaussiana es reiteradamente menor en comparación a los entregados por Filtersim. En el último banco, banco 2355, el ECM de SGSim es mayor que los entregados por Filtersim. Por otro lado, la incorporación de derivas externas con un 40 % de influencia en las realizaciones solo mejora levemente los resultados al disminuir el ECM en los bancos 2375 y 2365, mientras que en el banco 2385 prácticamente no altera los resultados y en el banco 2355 empeora levemente los resultados.

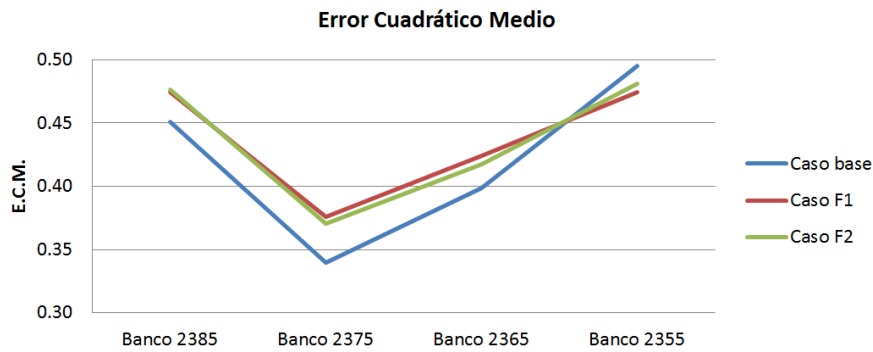


Figura 5.66: Error Cuadrático Medio banco a banco

Lo anterior guarda relación con los sesgos analizados previamente, donde el banco 2375 tenía el menor sesgo ahora presenta menor ECM respecto a los otros 3. En específico el caso F1 entrega mayor sesgo en la media en concordancia con entregar el mayor ECM. Sin embargo, la mejora al incorporar deriva se aprecia de mejor manera analizando el sesgo en la media que con la métrica del ECM. El caso particular del último banco también es similar donde SGSim entrega uno de los sesgos más altos en la media (respecto a los pozos de tronadura) mientras que su ECM es también el más alto de todos. En este caso particular la utilización de Filtersim mejora la precisión del estimador en comparación a SGSim.

Derivas

Para medir la reproducción de la deriva en las simulaciones se grafican las deriva en la media a lo largo de la coordenada X para los 4 bancos de estudio y los 3 casos simulados. Todos los resultados se encuentran detallados en la sección de Anexos i. Cada gráfico contiene la deriva en X de 22 realizaciones (color negro) y de los pozos de tronadura (color rojo) como métrica de comparación con la realidad.

Analizando el banco 2385, se tiene que los resultados mediante SGSim siguen la tendencia de la deriva, pero ésta se encuentra desplazada siendo relativamente sobrestimada. Por su parte, Filtersim convencional entrega una deriva debidamente ajustada a la encontrada en los pozos de tronadura. Al incorporar derivas externas en un 40 % no se aprecian mejoras significativas al obtener esta métrica. Cabe destacar que su incorporación no genera artefactos que lleven a empeorar los resultados.

Analizando el banco 2375, se observa que los tres casos (SGSim, Filtersim convencional y Filtersim con deriva) generan derivas similares y que se ajustan al modelo de deriva extraído de los pozos de tronadura. Sin embargo, Filtersim logra caracterizar mejor la deriva en la zona Este del banco donde la cantidad de bloques informados mediante pozos es más escasa (unos 13 bloques por cada coordenada X) y por ende genera mayor variabilidad. La incorporación de deriva ayuda a disminuir la dispersión (amplitud) observada en la respuesta de las 22 realizaciones para este mismo sector. No obstante, la zona opuesta - el lado Oeste - presenta una deriva más errática debido a la mayor escasez de bloques informados (entre 1 y 4) para cada coordenada.

El banco 2365 presenta un fenómeno similar al descrito en el banco 2375. La escasez de bloques informados en la zona Oeste lleva a un aumento en la dispersión de las respuestas al extraer las derivas, independiente del caso analizado. La incorporación de derivas no ayuda a una mejora consistente de las derivas ya que aplicando Filtersim convencional ya se logran buenos ajustes a la realidad (pozos).

Analizando el banco 2355 (ver figura 5.67), se observa que SGSim entrega resultados cuya deriva sigue la tendencia presentada por los pozos de tronadura pero con una desviación clara siendo sobrestimada en gran parte del banco. Los resultados al aplicar Filtersim convencional se ajustan de mejor forma a la deriva de los pozos, y al incorporar derivas se ajusta aún mejor disminuyendo la dispersión de las derivas provenientes de las realizaciones. El sector Oeste es el más despoblado del banco y su deriva queda mejor caracterizada al utilizar el Caso F2 que en cualquiera de los otros dos casos.

Lo anterior se resume en que la influencia de la deriva no siempre resulta en una mejora clara. Sin embargo, cuando tiene una influencia clara ayuda a ajustar mejor la deriva al modelo real y no empeora los resultados. Por otro lado, en todos los escenarios se caracterizó mejor la deriva por parte de Filtersim o Filtersim con deriva que al utilizar Simulación Secuencial Gaussiana.

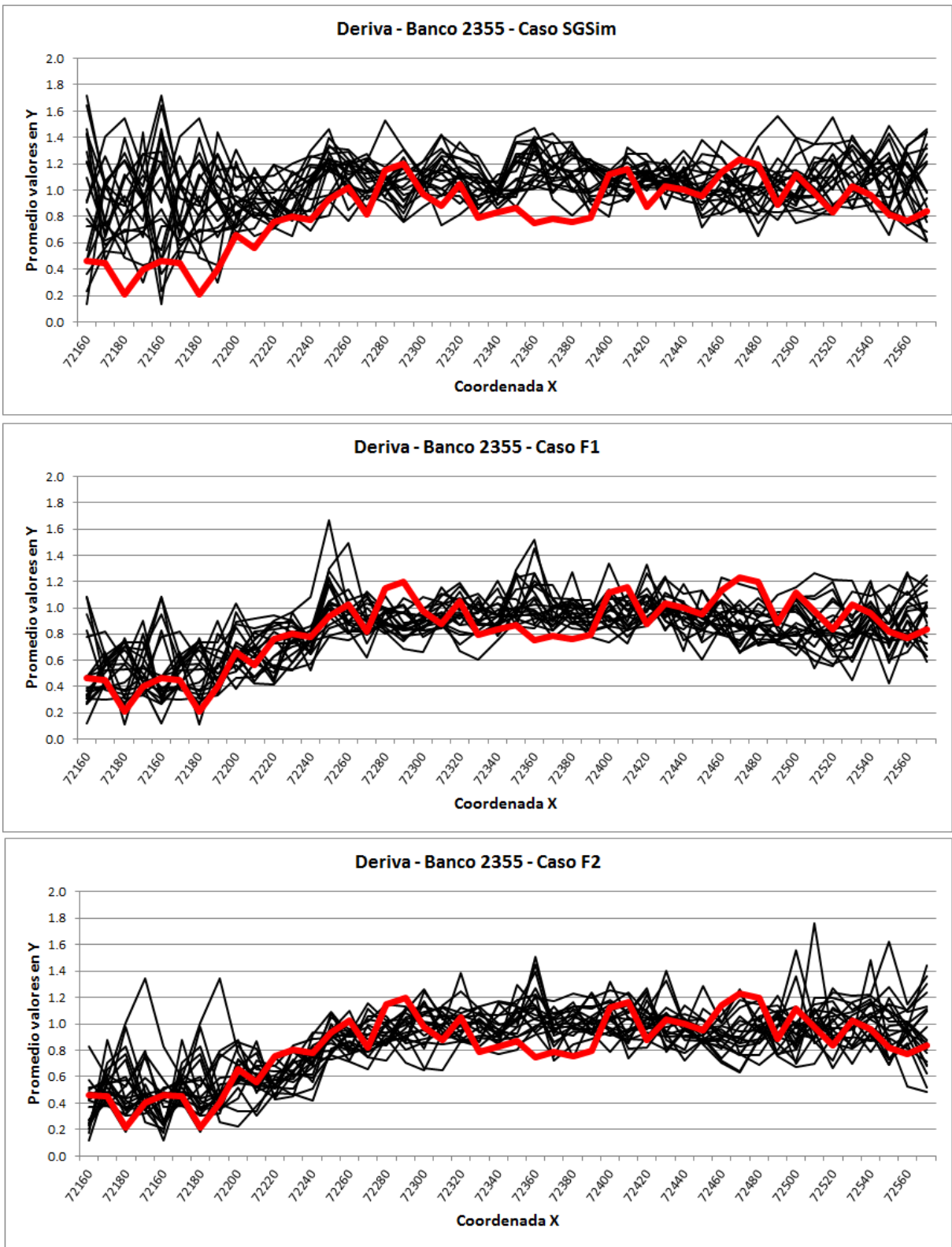


Figura 5.67: Derivas en la media - Coordenada X - Banco 2355

Mapas de probabilidad sobre ley de corte

Dado el conjunto de 100 realizaciones obtenidas en cada simulación y asumiendo que cada una de ellas representa un posible estado de la realidad es posible cuantificar la cantidad de veces que un bloque tiene asociada una ley de Cobre Total mayor a una Ley de corte crítica. Lo anterior se visualiza en los denominados mapas de probabilidad (MP). Por otro lado, se entiende por Ley de corte crítica al porcentaje de mineral que al ser minado, trasladado, procesado, refinado y vendido entrega una ganancia igual a los costos acarreados. Se infiere de lo anterior que la Ley de corte debe ser definida previo a determinar los mapas de probabilidad.

La ley de corte se determina directamente de la ecuación de beneficio (5.11) al igualarla a cero. ($B = 0$).

$$\text{Beneficio} = (P_{Cu} - C_{venta}) \cdot Rec \cdot Ley \text{ de mineral} \cdot Ton - (C_{Mina} - C_{Planta}) \cdot Ton \quad (5.11)$$

Al igualar a cero, dividir por el *Tonelaje(Ton)* y despejar la ley del mineral se obtiene finalmente la expresión (5.12).

$$\text{Ley de corte} = \frac{C_{Mina} - C_{Planta}}{Rec \cdot (P_{Cu} - C_{venta})} \cdot 100 \quad (5.12)$$

Dado que Compañía Minera Cerro Colorado explota su yacimiento mediante extracción a cielo abierto, se decide suponer los parámetros técnico-económicos expuestos en la tabla 5.20 para la determinación de la Ley de corte crítica. Cabe notar que se deben aplicar criterios de conversión entre unidades y el paso de porcentaje a valores numéricos para una adecuada interpretación de los resultados.

Tabla 5.20: Parámetros técnico-económicos para el cálculo de la ley de corte

Ítem	Valor	Unidad
Precio Cu	2.5	USD/lb
Costo Mina	5	USD/t
Costo F&R	6	USD/t
Costo Venta	0.7	USD/lb
Recuperación	87	%
Conversión lb/t	2204	lb/t

Como resultado de lo anterior se obtiene una Ley de corte, para Cobre total, de **0.32 %**.

Dado lo anterior se procede a determinar los MP banco a banco para cada caso de simulación. La totalidad de los mapas de probabilidad se encuentran en el Anexo J.

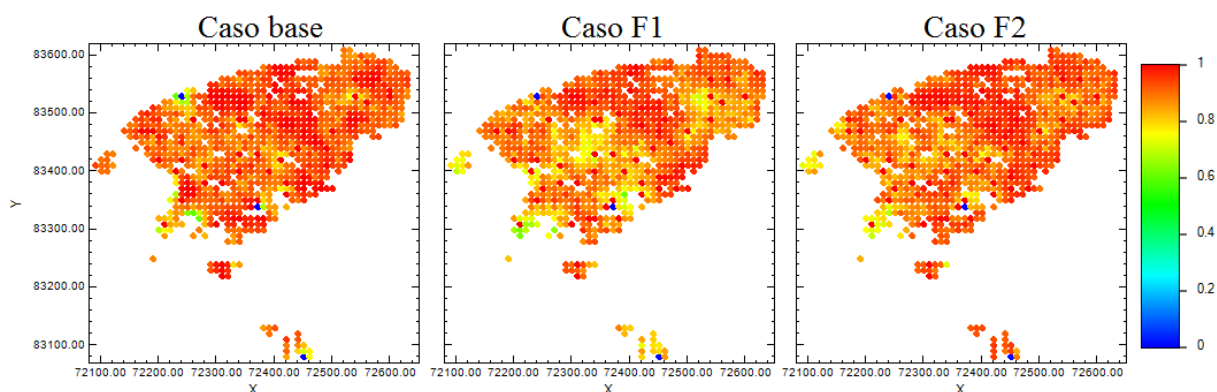


Figura 5.68: Mapas de probabilidad en banco 2365

Se presenta el caso del banco 2365 (figura 5.68) a modo de ejemplo. Al utilizar SGSim todo el banco tiene una ley por sobre 0.32 % con una probabilidad de al menos 80 % de ocurrencia. Por su parte Filtersim disminuye la probabilidad de ocurrencia de al menos 70 % en todo el banco. Al incorporar derivas, caso F2, se eleva nuevamente el porcentaje de ocurrencia a al menos 80 %. Lo anterior se traduce en que Filtersim convencional genera un rango de Leyes de cobre con mayor porcentaje de leyes bajo 0.32 % al interior del banco. Para tener una idea si se está subestimando la Ley de cobre en el caso F1 o bien sobrestimando con SGSim y caso F2, se propone categorizar según niveles de certidumbre y comparar con lo informado por los pozos de tronadura.

Resulta interesante observar que a distintos niveles de incertidumbre se determinan distintas zonas de ocurrencia. En otras palabras, al contar con los MP es posible fijar un nivel de incertidumbre y categorizar cada bloque como *está por sobre* o *está por debajo* de un porcentaje de ocurrencia tolerable. A modo de análisis se decide presentar los sectores al considerar un 85, un 90 y un 95 por ciento de ocurrencia de estar por sobre la Ley de corte crítica. Los resultados banco a banco y para cada caso se presentan en la sección de Anexos J.

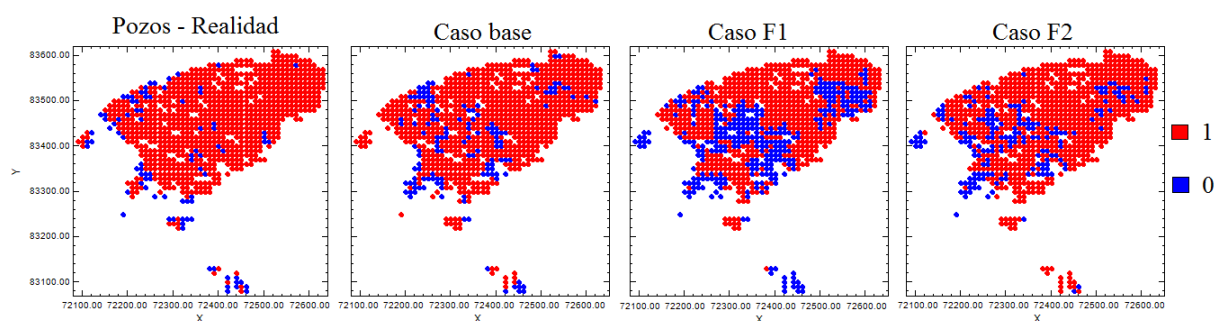


Figura 5.69: Categorización sobre banco 2365 - Umbral de 85 % de certeza sobre Ley de corte

La figura 5.69 presenta la categorización sobre el mapa de probabilidad, para el banco 2365, con un umbral de certeza de al menos 85 % de estar por sobre la Ley de corte. Con valor 1 se indexan todos los bloques que presentan en más de un 85 % de las ocasiones valores iguales o superiores a 0.32 % y con 0 el complementario. El MP categorizado más cercano a los pozos de tronadura lo entrega la Simulación Secuencial Gaussiana. Por su parte el algoritmo Filtersim (caso F1), en su forma original, entrega dos zonas principales cuya certidumbre es menor al umbral del 85 %. Esto llevaría a enviar los bloques indexados en 0 a botadero/planta lixiviación/stock cuando en realidad el beneficio sería mayor si se enviase a planta como parte de la producción. La integración de las derivas (caso F2) mejora de buena manera la categorización de los sectores. En efecto, es posible visualizar el aumento del área en rojo al interior del caso F2 en comparación al caso F1 llevando a sectores similares a los valores reales (pozos de tronadura).

La categorización (sectorizar) es sensible al valor umbral definido como nivel aceptable de certidumbre, tal como se observa en el anexo J. En los bancos 2385 y 2375 la categorización más cercana a la realidad se obtiene con un valor umbral de 85 % de ocurrencia para los 3 casos de estudio. El banco 2365 fue estudiado anteriormente. Por último en el banco 2355 los resultados más similares se generan con un umbral de ocurrencia del 90 % para SGSim, sin embargo Filtersim convencional e incorporando derivas generan artefactos en la zona Oeste, subestimando las Leyes de cobre, lo cual disminuye el volumen de material económicamente extraíble.

Finalmente, se observan mejoras al incorporar derivas en los bancos 2385, 2375 y 2365 sin embargo en el banco 2355 hay presencia de artefactos en las realizaciones que llevan a subestimar las leyes en el sector Oeste. Se infiere que la calidad en los resultados depende de un conjunto de factores como: el algoritmo que está siendo aplicado, factores geológicos (relación con la distribución de leyes en la T.I.), cantidad y calidad de datos condicionantes, entre otros.

Curvas Tonelaje-Ley

Las Curvas de Tonelaje-Ley son una representación gráfica que cuantifica la cantidad de mineral y la ley media asociada presente en un determinado dominio sobre una determinada Ley de corte. En el eje de las abscisas se gradúa la ley de corte del mineral de interés. En el lado izquierdo del eje de las ordenadas se grafica la cantidad de material acumulado y a la derecha la Ley media del material antes mencionado sobre una determinada Ley de corte. Las curvas Tonelaje-Ley son calculadas para todos los bancos y bajo los 3 escenarios de estudio. Los resultados son agrupados por banco y presentados en la sección de Anexos K.

Las curvas Tonelaje-Ley sobre el banco 2385 declaran recursos totales por 1.6 [Mton] a una ley de 1.0 % de Cobre Total. Por parte de SGSim se observa una sobrestimación de los recursos al definir una Ley de corte en el intervalo [0.0, 1.2] % y una subestimación sobre 1.2 % además de subestimar la Ley media sobre una ley de corte 0.5 %. Por otro lado, Filtersim convencional sobrestima, en menor grado, las curvas Tonelaje en el intervalo [0.0, 0.7] % y subestima, en menor grado también, para Leyes de corte mayores a 0.7 %. La curva de Ley media está siempre ajustada a la realidad salvo pequeñas desviaciones. La incorporación de derivas no mejora ni empeora la cuantificación de los recursos ya que las curvas observadas tienen un alto grado de similitud al caso de Filtersim convencional.

Al analizar el banco 2375 (figura 5.70), se declaran 1.23 [Mton] a una ley media del 1.0 % de Cobre total. Los resultados expuestos por SGSim difieren claramente de la realidad entregando una sobreestimación de 200.000 toneladas a lo largo de todo el intervalo [0.0, 1.2] % ajustándose solo en la última parte de la curva. Sin embargo, la ley media coincide (como consecuencia de lo anterior) con la de los pozos de tronadura. Por su parte, la simulación de Filtersim convencional ajusta de buena manera a las curvas de los pozos de tronadura (realidad) tanto para el Tonelaje como la Ley media. La incorporación de derivas en este banco ayuda a minimizar aun más la dispersión mostrada entre las distintas curvas por parte de las realizaciones. En este caso si tuvo una repercusión positiva la incorporación de derivas ya que define rangos más pequeños del recurso mineral esperado en la extracción y la ley media para una Ley de corte fija.

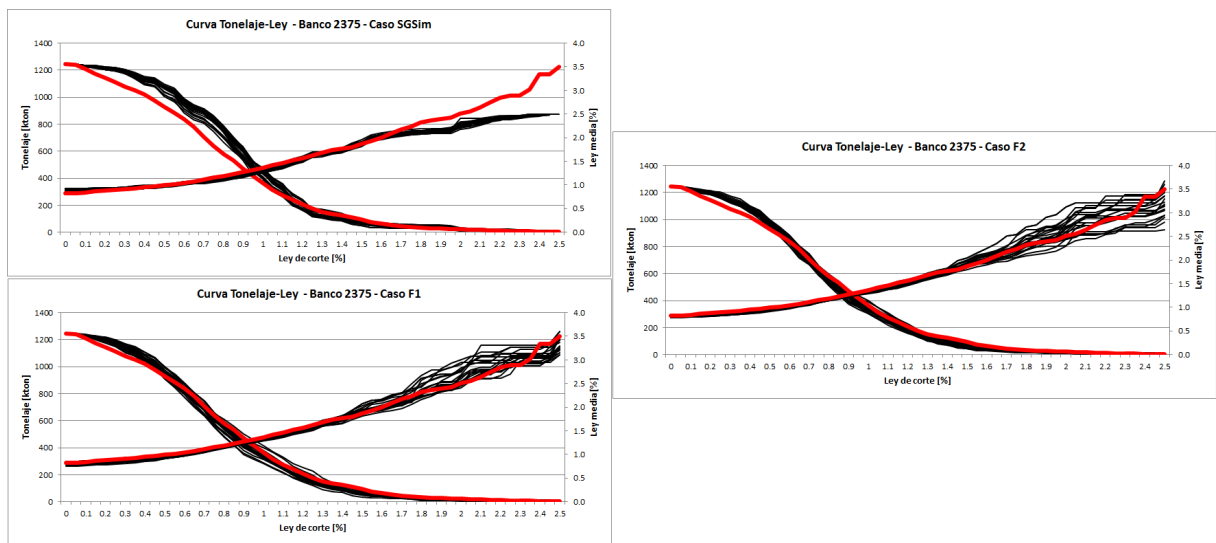


Figura 5.70: Curvas Tonelaje-Ley banco 2375. Utilización del caso base, caso F1 y caso F2

En el caso del banco 2365, se declaran 0.93 [Mton] a una ley media del 0.9 % de Cobre total. Los resultados al utilizar SGSim describen de forma adecuada el comportamiento real de las curvas Tonelaje-Ley en el intervalo [0.0, 1.6] %, sobre 1.6 % la Ley media se encuentra muy subestimada. Por parte de Filtersim convencional los resultados no son auspiciosos ya que subestima en todo momento los recursos reales disminuyendolos en unas 100.000 toneladas. Al incorporar un 40 % de influencia a la deriva externa se mejoran considerablemente los resultados. Sin embargo, aun así se subestiman los recursos en unas 50.000 toneladas. Resulta interesante proponer un aumento en la influencia de la deriva y cuantificar si se logran mejores ajustes a las curvas esperadas. Para los 3 casos (SGSim, F1 y F2) las Leyes medias fueron adecuadamente interpretadas.

En el caso final del banco 2355, se declaran recursos por 0.615 [Mton] a una ley aproximada del 1.0 % de Cobre total. Los resultados obtenidos mediante SGSim sobrestiman los recursos en unas 100.000 toneladas, sobrestimando la ley media en el intervalo [0.0, 0.9] % y subestimandola para Leyes de corte sobre 0.9 %. Por su parte, la simulación mediante Filtersim convencional caracteriza de muy buena manera las curvas Tonelaje-Ley estando ajustadas a la realidad en gran parte del intervalo de Leyes de corte. La incorporación de derivas mantiene el buen ajuste mejorando más las respuestas al disminuir la dispersión existente entre las realizaciones.

Se concluye de lo anterior que la influencia de la deriva o bien no repercute en la reportabilidad de recursos o bien la mejora ajustando las curvas a la realidad y disminuyendo además la dispersión por parte de las realizaciones.

Capítulo 6

DISCUSIÓN Y CONCLUSIONES

En el presente capítulo se resume, discute y concluye la investigación sobre el estado actual de la simulación multipuntos, los resultados obtenidos en los casos de estudio, y recomendaciones para trabajos futuros.

6.1. Estado del arte e incorporación de derivas externas

Se estudiaron los actuales algoritmos multipuntos y modificaciones a los mismos que se emplean actualmente en la simulación de variables categóricas y continuas (Síntesis textural, Snesim (2002), Simpat (2005), Filtersim (2006), DisPat (2010), Direct Sampling (2010), e Impala (2011)). Se elige Filtersim por sobre el resto de los métodos para la incorporación de derivas en la media local ya que puede ser aplicado tanto a variables categóricas como continuas, su proceso interno está descrito a cabalidad, presenta más alternativas de modificación para la incorporación que sus pares, y por la velocidad de cómputo y la baja demanda de memoria, entre otras cualidades. Para ello se describe la estructura interna de funcionamiento desde la inferencia de patrones, el proceso de agrupamiento, el cálculo de la similitud entre la realidad y el patrón más semejante, y la reproducción de este sobre el dominio a simular. Se propone a ello 3 opciones de integración de deriva, resultando escogida la que modifica la función distancia agregando un nuevo ponderador, w_0 , a la diferencia entre la media externa y la media mediante filtro en el punto (o bloque) a simular. Para llevar a cabo la incorporación se implementa el algoritmo Filtersim en Python.

6.2. Resultados

Caso de estudio - Roca sedimentaria

El desarrollo del caso de estudio - Roca sedimentaria se realiza para medir el impacto en los resultados que tiene la incorporación de derivas externas bajo distintos niveles de influencia. Además, se cuantifica la repercusión que tiene el grado de información por parte de las derivas y la cantidad de datos condicionantes sobre un dominio determinado. Se entiende por grado de información al nivel de detalle coordenada a coordenada que entrega la deriva.

La no incorporación de derivas lleva a resultados que distan de la realidad esperada sujeta a datos condicionantes. Esto se debe en gran medida a la presencia mayoritaria de ciertas estructuras al momento de inferir los comportamientos desde la imagen de entrenamiento, lo que influye luego en la reproducción de la realidad ya que existiría una elección preferencial por las estructuras predominantes. Al fijar un 20 % de influencia de las derivas en la elección del patrón se mejora considerablemente la reproducción, entregando estructuras más continuas y mejor delimitadas. Se aprecia rápidamente la importancia del grado de información por parte de las derivas. Utilizar derivas con alto grado de información (impuesta) lleva a caracterizar de mejor manera las zonas de estructuras que se busca reproducir. Sin embargo, se observa mucha disimilitud entre las realizaciones. Al aumentar la influencia de la deriva entre un 40 % y un 60 % las realizaciones

entregan zonas aún mejor demarcadas manteniendo la sinuosidad de las estructuras y mejorando la continuidad en la vertical. Dado que las derivas presentan un carácter continuo en la vertical, aumentar su porcentaje de influencia acarrea consigo la pérdida del carácter sinusoidal por parte de las estructuras llevándolas a un comportamiento más lineal alejándolas del comportamiento real esperado. En efecto, al medir el error cuadrático medio entre la deriva externa y la derivada de los resultados se tiene que el menor valor (ECM: 0.02) se obtiene con una influencia entre 80 % y 100 % en contraste con ECM sin incorporar derivas (ECM: 0.06).

Al tener informado un 1 % del dominio se logra una reproducción estructural adecuada y ajustada a la realidad en contraste con la pobre reproducción que se obtiene con un 0.075 % informado. Lo anterior se deduce con una nula influencia de la deriva, respaldando con esto la importancia de los datos condicionantes en el proceso de simulación. Por otro lado, unas de las derivas se obtienen al ser inferidas de los datos condicionantes y otras directamente del dominio que se desea simular (impuesta). Esta última tiene un carácter irreal ya que normalmente se desconoce la zona a simular. Los resultados observados con un 40 % de influencia por parte de la deriva inferida con 0.075 % del dominio informado distan bastante de la realidad esperada y de los resultados presentados por las derivas impuestas. Por su parte, la deriva inferida con un 1 % del dominio informado logra una buena reproducción de la realidad, muy semejante a la obtenida con la deriva impuesta. Lo anterior lleva a concluir que no es necesario conocer a cabalidad el dominio para obtener una deriva suficientemente informada. Teniendo un 1 % de datos condicionantes, distribuidos aleatoriamente sobre el dominio y sin muestreo preferencial, es posible inferir una deriva suficientemente ad-hoc para el proceso de simulación. La validación variográfica respalda la reproducción de los variogramas y se observa que el variograma de las realizaciones tiende al variograma de la imagen de entrenamiento cuando no se consideran derivas externas, y a medida que el porcentaje de influencia aumenta de 0 % a 100 % la variografía pasa de tender al variograma de la imagen de entrenamiento a tender al variograma del Ground Truth (sector a simular – origen de la deriva impuesta y datos condicionantes).

Analizando el nivel de asertividad (porcentaje de acierto) al incorporar derivas externas se observa que en todas las combinaciones *datos condicionantes - deriva externa* se observa un incremento en el porcentaje de acierto del orden de un 10 % al pasar de una influencia de 0 % ($w_0 : 0.0$) a 60 % ($w_0 : 0.6$). Un aumento sobre el 60 % o bien mantiene este nivel o lo empeora.

Al comparar los resultados obtenidos mediante métodos de estimación (Kriging Simple y Ordinario) y simulación (Bandas Rotantes) con los resultados obtenidos por Filtersim incorporando derivas los primeros no logran reproducir la realidad esperada mientras Filtersim modificado se aproxima de buena manera. Esto garantiza la utilización de simulaciones multipuntos por sobre estimaciones/simulaciones convencionales y más aún, la aplicación de Filtersim incorporando derivas por sobre Filtersim convencional. Inconvenientes en la reproducción de estructuras son observados a niveles más pequeños donde contactos entre zonas de valores altos, medios y bajos son confusos y presentan ruido. Esto se produce por el tamaño pequeño de la grilla de búsqueda en comparación al tamaño de las estructuras que se desea simular. Se utiliza una grilla múltiple de nivel 2 para palear este problema, sin embargo no se logran estructuras con valores iguales en su interior sino estructuras bien delimitadas pero con valores medios y altos en su interior. Por lo tanto, si el objetivo de implementar el método modificado es caracterizar bien a pequeña escala el algoritmo adolece de no ser capaz de obtener estas reproducciones.

Finalmente, se concluye que el óptimo porcentaje de influencia de la deriva externa está entre un 40 % y un 60 % logrando mejorar los resultados en comparación a Filtersim convencional manteniendo las propiedades de la estructura que se pierdan al aumentar la influencia a un 80 %. La inferencia de las derivas desde los datos condicionantes es óptima si el porcentaje de datos informantes es del orden del 1 % del dominio a simular.

Caso de estudio - Cerro Colorado

En el caso de estudio – Cerro Colorado se utilizó el resultado del caso de estudio – Roca sedimentaria en cuanto al porcentaje óptimo de influencia de la deriva (40 %) para la simulación de leyes en bancos inferiores basado en la información de un banco superior. Visualmente las realizaciones obtenidas con Filtersim incorporando deriva en contraste con las obtenidas mediante Filtersim convencional dependen del banco analizado. Sin embargo, las mejoras o se observan sólo en ciertos sectores (sector medio en el banco 2365, Sur en el 2385 y Oeste en el 2355) o bien no son apreciables (banco 2375).

El estudio de la continuidad espacial, variografía, da cuenta que las realizaciones entregadas por Filtersim siguen un comportamiento similar al presentado por los pozos de tronadura. No así los resultados por SGSim que siguen un comportamiento variográfico similar al presentado por los sondeos de exploración. La incorporación de derivas o bien llevó a una disminución en la dispersión en los variogramas de las realizaciones o bien no tuvo mayor influencia.

Los sesgos en la media, respecto a los pozos de tronadura o sondeos de exploración, entregados por parte de Filtersim incorporando derivas es menor que en su versión original para los bancos ubicados a 30 y 40 metros del banco desde el cual se infirió la imagen de entrenamiento. Sin embargo para los bancos ubicados a 10 y 20 metros no hubo variación en el sesgo esperado. Por lo anterior la disminución de sesgo no es algo intrínseco, aunque no se presentó un aumento, lo cual es positivo.

Respecto a las derivas que se infieren de las realizaciones al simular, al utilizar Filtersim convencional se tiene una adecuada representación y las realizaciones se establecen en torno a la deriva de los pozos de tronadura para todos los bancos simulados. Al incorporar derivas externas la variación (amplitud – dispersión) que existe entre las derivas de las diferentes realizaciones se disminuye considerablemente en el banco 2355, levemente en los bancos 2365 y 2375 y no se aprecian cambios en el banco 2385. Por su parte, la simulación por parte de SGSim en el banco 2385, si bien sigue la tendencia de la deriva, ésta se encuentra desplazada por sobre la deriva de los pozos en promedio por 0.3 % Ley de cobre total. En los bancos 2375 y 2365 SGSim sigue adecuadamente las tendencias en derivas de los pozos de tronadura salvo en los sectores Este y Oeste por falta de datos condicionantes. En el banco 2355 SGSim sobrestima la deriva por sectores, en el lado Oeste en 0.9 % Ley de cobre total, en el centro por 0.2 % y en el lado Este por 0.4 %. Dado lo anterior, la reproducción de las derivas es mejor utilizando Filtersim incluyendo derivas externas al 40 %, luego por Filtersim convencional y atrás queda SGSim con una sobrestimación marcada de la deriva en algunos bancos y ciertos sectores.

El error cuadrático medio obtenido en los bancos 2375 y 2365 es menor (ECM: 0.37 y 0.417) al utilizar Filtersim incorporando derivas que Filtersim convencional (ECM: 0.376 y 0.424). Sin embargo siempre están por sobre los valores obtenidos por la Simulación Secuencial Gaussiana (ECM: 0.339 y 0.339). En el banco 2385 la incorporación de derivas no afecta ni positiva ni negativamente los ECM obtenidos. En el caso particular del banco 2355 la Simulación Secuencial Gaussiana (ECM: 0.495) está por sobre el ECM de los casos utilizando Filtersim y el convencional (ECM: 0.474) está por debajo del modificado (ECM: 0.476). Se concluye de lo anterior que el error cuadrático entre los valores simulados mediante Filtersim (convencional y modificado) y los valores reales tiene una mayor dispersión que los generados mediante Simulación Secuencial Gaussiana.

Los mapas de probabilidad llevaron a categorizar de forma adecuada sectores sobre una Ley de corte bajo umbrales de certidumbre entre 85 % y 90 %. En particular, la incorporación de derivas mejora la calidad de la categorización si se compara con la información real (pozos de tronadura) para los tres primeros bancos. No así en el último banco, donde se observó una subestimación de

la ley en el sector Oeste dañando la categorización. Se concluye de lo anterior que la calidad de los resultados al incorporar derivas depende de cuan relacionado está el comportamiento interno de la variable en la imagen de entrenamiento con el comportamiento de la misma variable en el dominio a simular. Este criterio es bastante subjetivo, aunque respaldado y guiado de forma adecuada con información geológica es posible determinar dónde aplicar la nueva técnica.

Sobre la cuantificación de los recursos a través de las curvas Tonelaje-Ley se tiene que la Simulación Secuencial Gaussiana sobrestima los recursos a leyes de corte bajo 1.2 y subestima los recursos a leyes de corte mayores para los bancos 2385, 2375 y 2355, sólo en el banco 2365 las curvas se ajustan a las presentadas en la realidad (pozos de tronadura). Por su parte, Filtersim convencional reproduce adecuadamente el comportamiento de las curvas para todos los bancos dentro de variaciones razonables. Al incorporar derivas con un 40 % de influencia se disminuye las variaciones entre las realizaciones y se mejora el ajuste en sectores de leve sobre y subestimación.

6.3. Conclusiones generales

La metodología de incorporar derivas externas tiene ventajas y desventajas que consideran un conjunto de variables externas tales como: disponibilidad de una T.I. 100 % informada al interior de una grilla específica, consistencia de la imagen de entrenamiento y los datos condicionantes; forma, tamaño y continuidad de los patrones a inferir; tipo de mineralización (vetas, depósitos diseminados, depósitos exóticos, etc) a simular; y todos los parámetros internos. Se observa que tiene buenos resultados al simular cuerpos mineralizados vetiformes, u otros elementos geológicos cuya disposición espacial tenga un carácter continuo en una dirección y acotado (estrecho) en la dirección perpendicular. Esto último principalmente por el contraste generado en los contornos de las estructuras al ser codificadas con valores numéricos. El caso contrario se observa en los depósitos más diseminados, donde los contornos entre sectores de altas leyes y bajas leyes es difuso y poco marcado. Esta diferencia sustancial es resultado del proceso de agrupación de patrones. En el primer caso los patrones inferidos de la imagen de entrenamiento son claramente semejantes o diferentes. Se deduce así que la reproducción de patrones toma mayor relevancia que la reproducción del variograma (debido a la naturaleza del algoritmo) lo cual es beneficioso al algoritmo pero pocas veces ocurre en la gran Minería del Cobre por la naturaleza geológica de sus depósitos.

Dentro de las principales desventajas que presenta el algoritmo es la sensibilidad de los resultados a la consistencia entre la imagen de entrenamiento, los datos condicionantes y la deriva externa. La capacidad de representar el mismo fenómeno es de vital importancia, de otra manera los resultados se ven distorsionados y la capacidad de predicción disminuida. Este fenómeno es recurrente en las técnicas de simulación multipuntos y dificulta la aplicación en minería donde encontrar un sector desde el cual extraer una T.I. capaz de representar completamente el comportamiento de otro sector se vuelve una tarea difícil. Por otro lado, se observó que los resultados mejoran al incorporar derivas externas con un rango de influencia entre el 40 % y el 60 %. Sin embargo las derivas impuestas (en el caso de estudio roca sedimentaria) fueron obtenidas del Ground Truth algo imposible de realizar en minería. Las derivas interpretadas (en ambos casos de estudio) fueron obtenidas sobre muestras del Ground Truth (sondajes, en el caso de estudio Cerro Colorado) entregando mejoras en la calidad de los resultados con diferencia que en el corte pulido de roca sedimentaria son más claras que en los bancos de aplicación en Cerro Colorado. Por todo lo anterior, la ventaja de incluir derivas externas utilizando Filtersim modificado, que presenta buenos resultados, depende en gran medida de la capacidad de obtener una imagen de entrenamiento y una deriva externa que tengan consistencia con los datos condicionantes y en particular sobre el fenómeno que se desea modelar.

6.4. Recomendaciones y trabajo futuro

A nivel de algoritmo, la incorporación de una grilla dual (tal como la usada por el algoritmo Simpat) puede llevar a generar contactos más suaves en la reproducción de las estructuras. Este debería ser incorporado en el proceso interno llevado a cabo por el algoritmo.

Sobre los casos de estudio realizados se recomienda realizar las siguientes modificación y/o trabajos futuros:

Caso de estudio - Roca sedimentaria

Los resultados obtenidos son todos para una condición fija de los parámetros de entrada a Filtersim (convencional e incorporando derivas). Sería interesante aumentar el tamaño de la grilla de búsqueda e interna de 9x9 a 11x11 - 13x13 - 15x15 disminuyendo al mismo tiempo el número de *clusters* de 200 a 170 - 140 - 110. Con ello se capturarían mejor las estructuras a gran escala y se aseguraría la existencia de patrones en el proceso de agrupamiento.

Otra opción interesante es modificar la geometría de la grilla de búsqueda e interna de cuadrada a rectangular. Dado que la estructura presenta continuidad en la vertical (anisotropía), la cual se busca reproducir en las simulaciones, la grilla de búsqueda y la interna podrían cambiarse de 9x9 a 9x13 - 9x17 - 9x21 con el eje mayor en la dirección de la anisotropía.

Cambiar la grilla de búsqueda desde un tipo *no causal* a una *causal* (modificando la estructura interna del algoritmo), alterando además la secuencia de simulación desde un camino aleatorio a uno secuencial. Se argumenta lo anterior a privilegiar la construcción de estructuras en la dirección Norte-Sur.

Respecto a los filtros. Se emplearon en las direcciones Norte-Sur y Este-Oeste, sin embargo en el caso de estudio - Roca sedimentaria las estructuras principales están en dirección Norte-Sur. Se propone utilizar los filtros sólo en esa dirección y no en la dirección Este-Oeste. Esto ya que el interés de capturar estructuras horizontales no es prioridad y es posible que la agrupación de patrones similares sea más sencilla y de mejor calidad.

Realizar un estudio del Error Cuadrático Medio entre el valor real y el esperado punto a punto y no sólo entre las derivas externas y resultantes.

Caso de estudio - Cerro Colorado

Se recomienda cambiar la metodología de simulación. Utilizar la información del banco $N - \textit{ximo}$ para inferir la imagen de entrenamiento y simular el comportamiento del banco $(N - 1) - \textit{ximo}$ inferior condicionado a los datos de sondajes. Repetir el proceso para un número finito de bancos. Repetir el proceso global considerando la información del banco $N - \textit{ximo}$ pero simulando el banco $(N - 2) - \textit{ximo}$, y luego el $(N - 3) - \textit{ximo}$. Con lo anterior cuantificar la repercusión que tiene la aplicación de Filtersim, en particular la incorporación de derivas, a distintos niveles del yacimiento.

Considerar derivas en la dirección Norte-Sur (eje Y) y no solo en la dirección Este-Oeste. Bien podría plantarse una metodología (función en dos dimensiones) para el cálculo de una deriva local más representativa al considerar una combinación entre las derivas en la media Norte-Sur y Este-Oeste.

Por último, realizar un contraste entre Filtersim incorporando derivas y la simulación real que se tenía para el banco estudiado por parte de la Compañía Minera Cerro Colorado.

Bibliografía

- [1] G. ARPAT AND J. CAERS, *A multiple-scale, pattern-based approach to sequential simulation*, Geostatistics Banff 2004-Springer Netherlands, (2005), pp. 255–264.
- [2] AUSTIN, *Sgems short course: filtersim*, in Stanford Geostatistical earth modeling software, Austin, Texas, 2007, pp. 9–10.
- [3] A. BOUCHER, J. F. COSTA, L. G. RASERA, AND E. MOTTA, *Simulation of geological contacts from interpreted geological model using multiple-point statistics*, Mathematical Geosciences, (2013), pp. 1–12.
- [4] F. BOUZARI AND A. H. CLARK, *Anatomy, evolution, and metallogenic significance of the supergene orebody of the Cerro Colorado porphyry copper deposit, I region, northern Chile*, Economic Geology, 97 (2002), pp. 1701–1740.
- [5] J. CAERS AND A. BOUCHER, *Multiple-point geostatistics: mastering training-image based algorithms*, in MPS Short course, Chile, 2008, p. 23.
- [6] C. DEUTSCH AND A. JOURNEL, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, second ed., 1998, pp. 20-24.
- [7] X. EMERY, *Geoestadística*, Universidad de Chile, Facultad de ciencias físicas y matemáticas, Universidad de Chile, Santiago, Chile, pp. 21-28, 60-71, 2011.
- [8] X. EMERY, *Simulación Geoestadística*, Universidad de Chile, Facultad de ciencias físicas y matemáticas, Universidad de Chile, Santiago, Chile, pp. 6, 2011.
- [9] X. EMERY, *Lección 11: simulación no paramétrica*, in Cátedra simulación geoestadística, Universidad de Chile, Santiago, Chile, pp. 62., 2012.
- [10] X. EMERY, *Simulación estocástica y geoestadística no lineal*, Universidad de Chile, Facultad de ciencias físicas y matemáticas, Universidad de Chile, Santiago, Chile, pp. 1-7, 2012.
- [11] X. EMERY AND C. LANTUÉJOUL, *Can a training image be a substitute for a random field model?*, Mathematical Geosciences, (2013), pp. 1–15.
- [12] F. GUARDIANO AND M. SRIVASTAVA, *Multivariate geostatistics: Beyond bivariate moments*, Geostatistics-Troia: Kluwer Academic, Dordrecht, (1993), pp. 133–144.
- [13] J. HARTIGAN, *Clustering algorithms*, John Wiley and Sons, Inc, (1975).
- [14] M. HONARKHAH AND J. CAERS, *Stochastic simulation of patterns using distance-based pattern modeling*, Mathematical Geosciences, 42 (2010), pp. 487–517.
- [15] M. HUYSMANS AND A. DASSARGUES, *Direct multiple-point geostatistical simulation of edge properties for modeling thin irregularly shaped surfaces*, Mathematical Geosciences, 43 (2011), pp. 521–536.

- [16] M. HUYSMANS, P. ORBAN, E. COCHET, M. POSSEMIERS, B. RONCHI, K. LAURIKS, O. BATELAAN, AND A. DASSARGUES, *Using multiple-point geostatistics for tracer test modeling in a clay-drape environment with spatially variable conductivity and sorption coefficient*, Mathematical Geosciences, (2013), pp. 1–19.
- [17] A. IZENMAN, *Modern multivariate statistical techniques: regression, Classification, and Manifold Learning* Springer Texts in Statistics, New York, (2008), pp. 423–424.
- [18] G. MARIETHOZ, P. RENARD, AND J. STRAUBHAAR, *The direct sampling method to perform multiple-point geostatistical simulations*, Water Resources Research, 46 (2010), pp. 1–5.
- [19] D. E. MYERS, *To be or not to be... stationary? that is the question*, Mathematical Geology, 21 (1989), pp. 347–362.
- [20] Á. PARRA AND J. M. ORTIZ, *Adapting a texture synthesis algorithm for conditional multiple point geostatistical simulation*, Stochastic Environmental Research and Risk Assessment, 25 (2011), pp. 1101–1111.
- [21] J. STRAUBHAAR, P. RENARD, G. MARIETHOZ, R. FROIDEVAUX, AND O. BESSON, *An improved parallel multiple-point algorithm using a list approach*, Mathematical Geosciences, 43 (2011), pp. 305–328.
- [22] S. STREBELLE, *Sequential simulation drawing structures from training images*, tesis de doctorado, Universidad de Stanford, California, pp. 34-53, 2000.
- [23] S. STREBELLE, *Conditional simulation of complex geological structures using multiple-point statistics*, Mathematical Geology, 34 (2002), pp. 3–12.
- [24] S. ÁVALOS, *Simulación multi-puntos utilizando el algoritmo filtersim: análisis de sensibilidad y casos de estudio*, tesis de ingeniería, Universidad de Chile, Santiago, Chile, pp. 109-114, 2013.
- [25] L. Y. WEI, *Texture Synthesis by fixed neighborhood searching*, tesis de doctorado, Universidad de Stanford, Stanford, CA, pp. 9-25, 36-40, 2001.
- [26] J. WU, A. BOUCHER, AND T. ZHANG, *A sgems code for pattern simulation of continuous and categorical variables: Filtersim*, Computers & Geosciences, 34 (2008), pp. 1863–1876.
- [27] T. ZHANG, P. SWITZER, AND A. JOURNEL, *Filter-based classification of training image patterns for spatial simulation*, Mathematical Geology, 38 (2006), pp. 63–80.

Capítulo 7

ANEXOS

7.1. Anexo A

Implementación de Filtersim en Python

El algoritmo *Filter-Based Simulation* (Filtersim) fue desarrollado por Zhang, et. al. en el año 2006 y luego implementado en el software SGeMS¹ en 2007. Actualmente solo se encuentra implementado en lenguaje C++ cuyo código fuente es de libre acceso. La presente tesis pretende tomar los principios del algoritmo y sus procesos tanto globales como particulares con la finalidad de incorporar modificaciones en sus procesos internos. Para ello se optó por implementar el algoritmo en Python 2.7 dada su simpleza en escritura y estructura. A continuación se detallan los procesos generales llevados a cabo.

Fijar parámetros externos

- Declaración de las variables externas determinadas por el usuario.
 - (a) Tamaño de grilla de búsqueda (**ngsx**, **ngsy**, **ngsz**)
 - (b) Tamaño de grilla interna (**innerx**, **innery**, **innerz**)
 - (c) Número de realizaciones (**repetición**)
 - (d) Número de multigrillas (**nmgrid**)
 - (e) Número de cluster en el proceso de agrupamiento, *K-mean partition* (**ncluster**)
 - (f) Pesos datos condicionantes (**w1**, **w2**, **w3**)
 - (g) Mínimo de datos condicionantes (**minimumdc**)

Importaciones

- Importación de los parámetros asociados a la imagen de entrenamiento desde el archivo de parámetros *parmb.txt*.
 - (a) Número de bloques (**numx**, **numy**, **numz**)
 - (b) Tamaño de los bloques (**tamx**, **tamy**, **tamz**)
 - (c) Origen de la imagen de entrenamiento (**minx**, **miny**, **minz**)
- Importación de la imagen de entrenamiento desde el archivo *mb.txt*. (**TI**)
- Importación de los parámetros asociados a la grilla a simular desde el archivo de parámetros *simgrid.txt*.

¹SGeMS, *Stanford Geostatistical Modeling Software*, es un software de análisis geoestadístico desarrollado por la Universidad de Stanford en 2009 e integra algoritmos MPS y herramientas de pre y post procesamiento.

- (a) Número de bloques (**snumx, snumy, snumz**)
- (b) Tamaño de los bloques (**sgtamx, sgtamy, sgtamz**)
- (c) Origen de la grilla a simular(**sgminx, sgminy, sgminz**)
- Importación de los parámetros asociados a los datos condicionantes desde el archivo de parámetros *parcd.txt*.
 - (a) Número de bloques (**dnumx, dnumy, dnumz**)
 - (b) Tamaño de los bloques (**dctamx, dctamy, dctamz**)
 - (c) Origen de la grilla de datos condicionantes(**dcmix, dcminy, dcminz**)
 - (d) Total de datos condicionantes (**totaldc**)
 - (e) Importación de los datos condicionantes desde el archivo *cd.txt*.
- Importación de los datos condicionantes a la grilla a simular.
 - (a) Asociar el ponderador **w1** al bloque (o punto) en donde se encuentra el dato condicionante. (Datos condicionantes que se encuentren fuera de la grilla a simular no son tomados en consideración)

Creación de filtros

- Generación de los filtros. Desde el archivo principal (*main.py*) llama a la función “Filters” (*Filters.py*) de la forma: $F_i = Filters(ngsx, ngsy, ngsz)$ con la finalidad de generar una matriz F_i que contiene todos los filtros. Dentro de la función es posible definir tantos filtros como se desee y con la estructura que usuario defina.

Creación de patrones

- Se comienza por generar una matriz denominada Patrones que contendrá el resultado de aplicar, para cada nivel de multigrilla o en una grilla simple, cada uno de los filtros en un punto específico de la imagen de entrenamiento.

$$par1 = (numx - ngsx + 1); par2 = (numy - ngsy + 1); par3 = (numz - ngsz + 1)$$

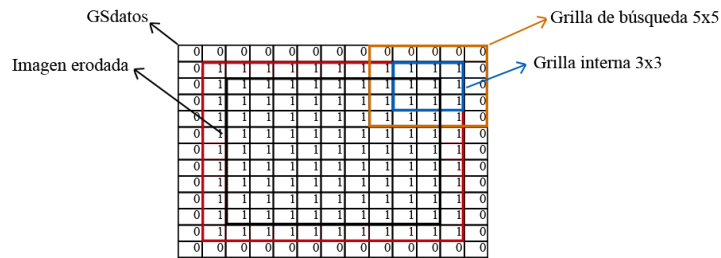
- $Patrones = -99.0 \cdot np.ones((par1, par2, par3, nmgrid, 10))$
- Se define la matriz *maps* llamando a la función *Scoremap.py*
- $maps = Scoremap(par1, par2, par3, nmgrid)$
- Se comienza el proceso de agrupamiento. Dada la cantidad de clases definidas en **ncluster** se eligen **ncluster** patrones de manera aleatoria en la matriz *Patrones* y se comienza con el proceso de agrupamiento basado en *K-mean partition*.
- Al tener las clases definidas se establece el prototipo representativo de la clase como el promedio de todos los patrones internos.

Simulación

- Se comienza con un ciclo (*for*) de tamaño igual al número de realizaciones definidas por el usuario en la variable **repetición**.
- Se crean 3 matrices idénticas de tamaño igual a la grilla a simular, ellas son:

$$\begin{aligned} Sdatos &= np.zeros((nmgrid, snumx, snumy, snumz)) \\ Listos &= np.zeros((nmgrid, snumx, snumy, snumz)) \\ GSind &= -99.0 \cdot np.ones((nmgrid, snumx, snumy, snumz)) \end{aligned}$$

Sobre *GSdatos*, como se observa en la figura inferior, se pasa de valor 0 a 1 dentro de los bloques que serán simulados, es decir, todos los que están dentro de la grilla a simular salvo los externos entre la grilla de búsqueda y la grilla interna en los contornos.



- Sobre *GSind*, cuyos valores por defecto son -99.0, se definen números que van de 1 a $(sgnumx - (ngsx - 1)) \cdot (sgnumy - (ngsy - 1)) \cdot (sgnumz - (ngsz - 1))$ de manera aleatoria. En otras palabras, se define un camino aleatorio dentro de la imagen erodada definido por números que van de 1 al número total de bloques dentro de ella. Este camino aleatorio define la ruta a seguir en el proceso de simulación. Cabe destacar que para grillas superiores hay bloques internos en *GSdatos* que se mantienen en 0 dado que son bloques no simulados a ese nivel.
- Se establece si va a simular en una grilla simple ($nmgrid = 1$) o en multi-grillas ($nmgrid > 1$). Cabe destacar que en el proceso de multigrillas se contempla intrínsecamente el caso de la grilla simple salvo sutiles diferencias, respecto a los datos de condicionamiento tal como será explicado más adelante.
- Se genera una variable denominada **com** definida como la suma total de la matriz *GSdatos* (asociado al nivel de grilla en el proceso de simulación). El valor definido por **com** entrega la cantidad de bloques que aún quedan por simular. Por lo tanto, se genera un nuevo proceso que tiene como restricción no detenerse hasta que **com** sea igual a cero, es decir, que todos los bloques estén simulados. Cabe notar que al utilizar multi-grillas el valor de **com** viene dado por la suma de los elementos dentro de la matriz *Listos* definida con valores 1 para las celdas a simular y 0 el resto para cada uno de los niveles de multi-grilla.
- Se comienza a recorrer el camino aleatorio dentro de la grilla a simular. Se define en cada instante la posición espacial del centro de la grilla de búsqueda como (**menx**, **meny**, **menz**).
- A través de la variable **com1** se cuantifica la cantidad de bloques que restan para simular dentro de la grilla de búsqueda de tamaño $(ngsx, ngsy, ngsz)$. Para ello se suman los correspondientes elementos dentro de *GSdatos* que define la cantidad buscada. Si el valor de **com1** es igual a cero entonces la matriz *Listos* se cambia el valor del bloque centrado en (**menx**, **meny**, **menz**).
- De tener un valor $com1 > 0$ se da paso a la siguiente etapa. Se define la variable **vueltas** inicialmente con valor 1 y al recorrer por completo el camino aleatorio incrementa de uno en uno. La relevancia de definir esta variable es que en la primera vuelta se considera la restricción de un número mínimo de datos condicionantes dentro de la grilla de búsqueda. Por lo tanto, al tener $com1 > 0$ se verifica que al interior de la grilla de búsqueda existan al menos (**minimumdc**) datos condicionantes.
- Al cumplirse los requisitos anteriores se establece luego el prototipo más cercano al evento condicionante. Para ello se utiliza la función distancia *Template Pixels* la cual establece una métrica de distancia como la suma de las diferencias absolutas pixel a pixel entre una celda informada dentro del evento condicionante y su homóloga en el prototipo. Esta diferencia es luego ponderada por w_1, w_2 o w_3 dependiendo el tipo de dato utilizado. Al

determinar el prototipo más cercano se elige de manera aleatoria un patrón dentro de la clase caracterizada por el prototipo.

- Teniendo en consideración el patrón seleccionado se procede a congelar sobre la grilla a simular los valores que están dentro de la grilla interna definida por el tamaño (innerx, innery, innerz). Los datos condicionantes en la grilla a simular que están dentro de la grilla interna no son reemplazados.
- A los valores simulados congelados se les asocia un ponderador w_2 mientras que a los datos condicionantes se les mantiene el valor w_1 . En el caso de estar utilizando multigrillas y no estar en la grilla más fina se procede de igual forma, sin embargo, al pasar de una grilla gruesa a una fina (de mayor a menor nivel) el ponderador de los datos simulados pasa de w_2 a w_3 ya que pasan a ser “datos condicionantes previamente simulados en grillas superiores”.
- Sobre la matriz $GSdatos$ se cambian los valores 1 por 0 dentro de la grilla interna, es decir, se informan los bloques simulados para ir cuantificando la cantidad restante de bloques.
- El proceso se repite hasta completar toda la zona a simular.
- El proceso completo se repite tantas veces como se definió en **repetición**.

Modificación a la función distancia y nuevo filtro

Esta sección presenta la implementación de derivas en la media local. El usuario debe definir previamente la ponderación que desea asignar a la media local impuesta externamente a través del ponderador w_0 . Tal como se observa en la ecuación (4.1) cuando el valor de w_0 es 1 la elección del prototipo más cercano solo toma en consideración la diferencia entre la media local impuesta y la media del prototipo. Dado que se decide utilizar una media local impuesta externamente y no desde los datos del evento condicionante es que la ecuación final queda:

$$D_l(u_0) = \left(\sum_{d=1}^3 w(d) \sum_{i=1}^{n_d} \left(dat_d(u_0 + u_i) - prot^{(l)}(u) \right) / n_i \right) (1 - w_0) + w_0 \left(\left| \bar{M}_{imp}(u_0) - \bar{M}_{prototipo}^{(l)} \right| \right), \text{ con } w_0 \in [0, 1] \quad (7.1)$$

En donde $\bar{M}_{prototipo}^{(l)}$ representa el valor promedio del prototipo l -ésimo y $\bar{M}_{imp}(u_0)$ la media local impuesta en el punto u_0 .

Finalmente el prototipo que, junto al evento condicionante, entregue el mínimo valor para $D_l(u_0)$ será el escogido y se escogerá aleatoriamente un patrón dentro del grupo representado por ese prototipo.

Creación de los archivos de salida

- El archivo *Scoremaps.txt* contiene los mapas de puntuación (enmarcados en la imagen erodada) para cada uno de los filtros definidos en la función “Filters”. El formato de salida es del tipo GSLib [6].
- El archivo *FinalPrototype.txt* contiene los prototipos asociados a cada clase para cada multigrilla. Al igual que *Scoremaps.txt* guarda el archivo en formato GSLib.
- El archivo *ResultadosUmine.txt* contiene el total de las realizaciones e incluye el promedio de ellas. Su formato de salida es GSLib y debe el sufijo *Umine* dado que fue el software utilizado para realizar el estudio estadístico de las realizaciones.
- El archivo *Results.txt* contiene los resultados de las simulaciones sin formato GSLib y sin el promedio de ellas, simplemente la posición en el espacio y los valores obtenidos para cada realización. Este archivo fue ideado para los post-procesamientos relacionados a las simulaciones.

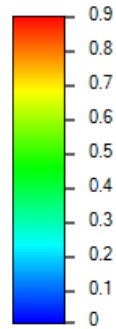
Finalmente, y con el espíritu de generar un código abierto, se deja público en la red tanto el código principal como las Clases extras necesarias para su funcionamiento. El link² se encuentra al pie de página.

²URL: http://github.com/SebastianAvalos/Thesis/blob/master/Main_2d.py

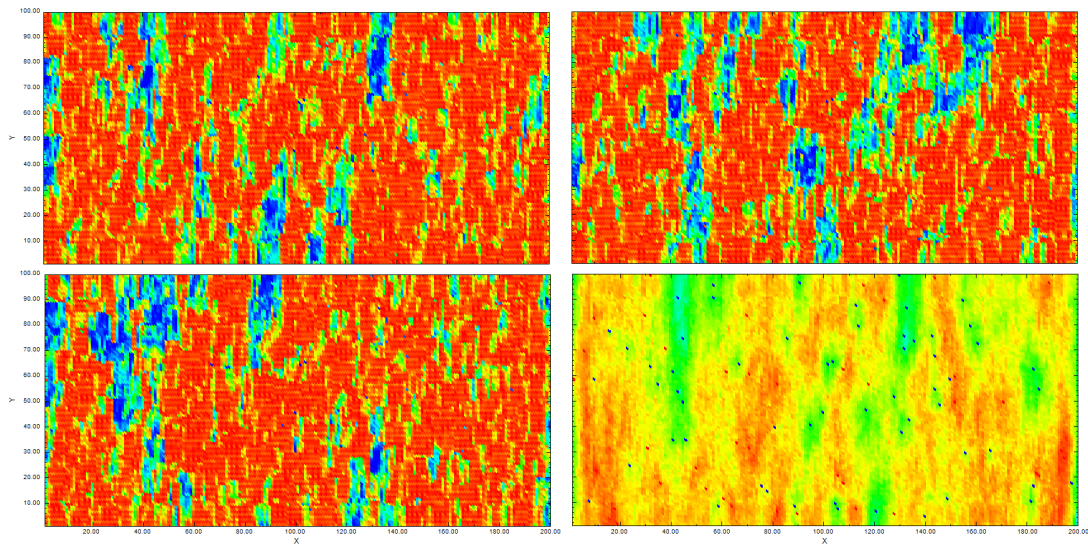
7.2. Anexo B

La presente sección presenta los resultados de las simulaciones del primer caso de estudio (depósito sedimentario). En esta se ilustran 3 de 100 realizaciones escogidas de manera aleatoria y el promedio de las 100 realizaciones.

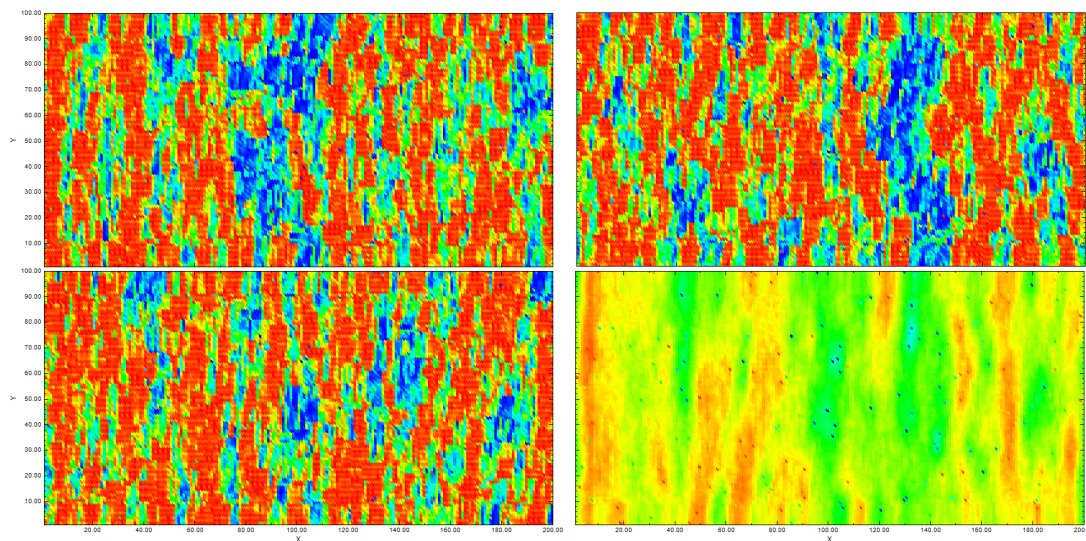
La escala de colores para el despliegue tanto de las realizaciones como de los promedios viene dada por:



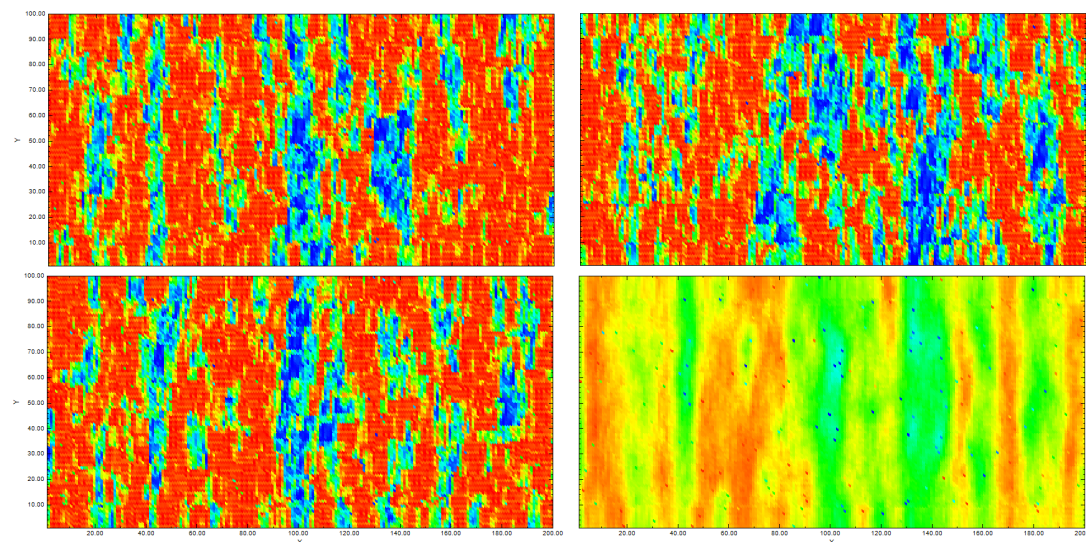
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 0.0$



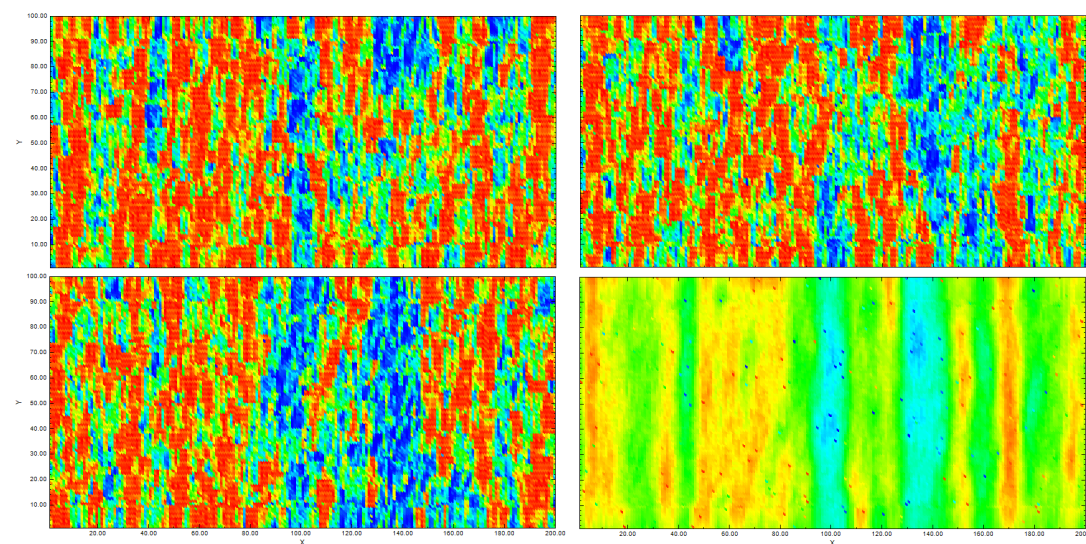
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 0.2$



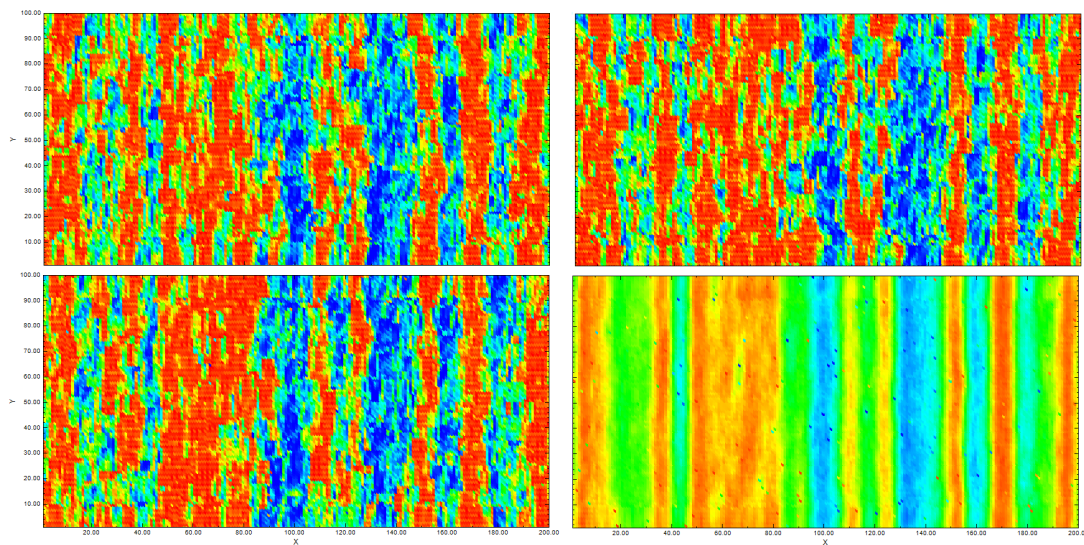
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 0.4$



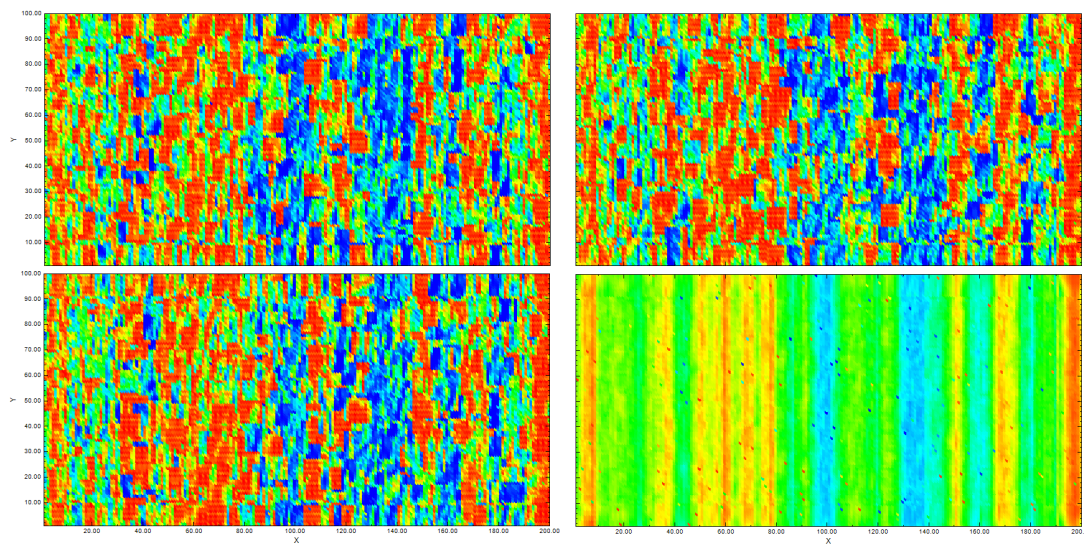
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 0.6$



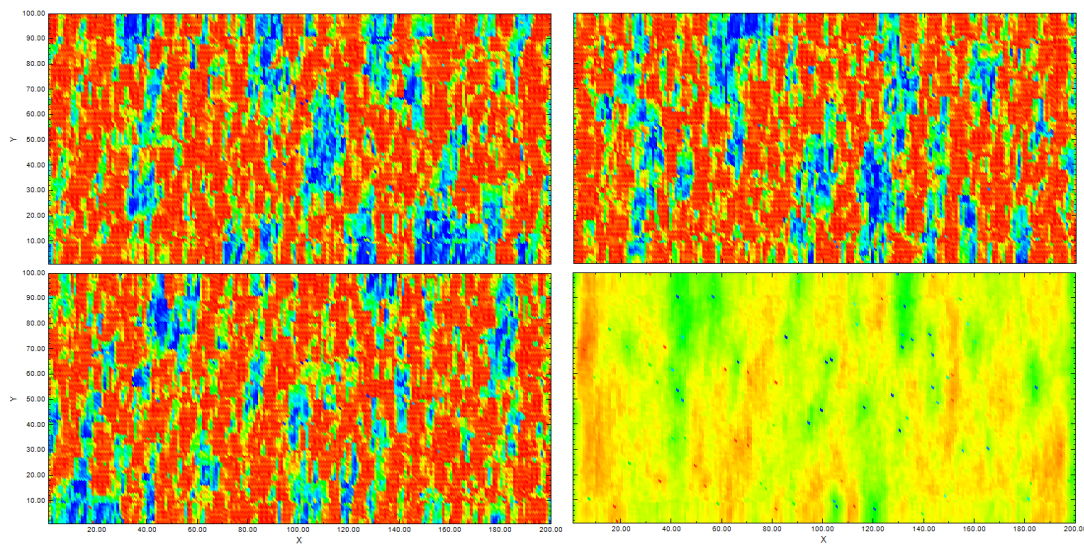
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 0.8$



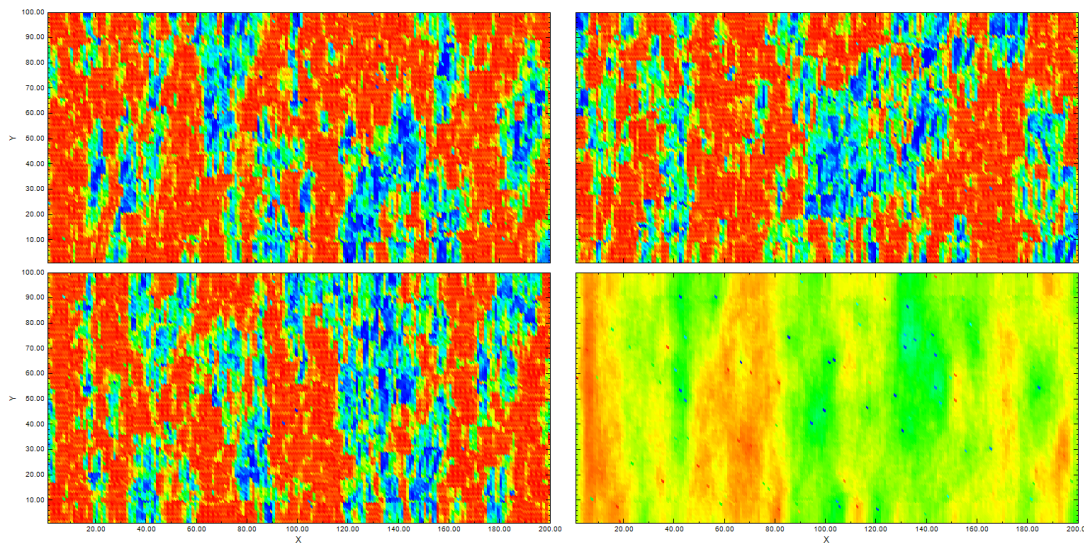
200 D.C. - 100 % D.C. - Media local Impuesta desde GT - $w_0 : 1.0$



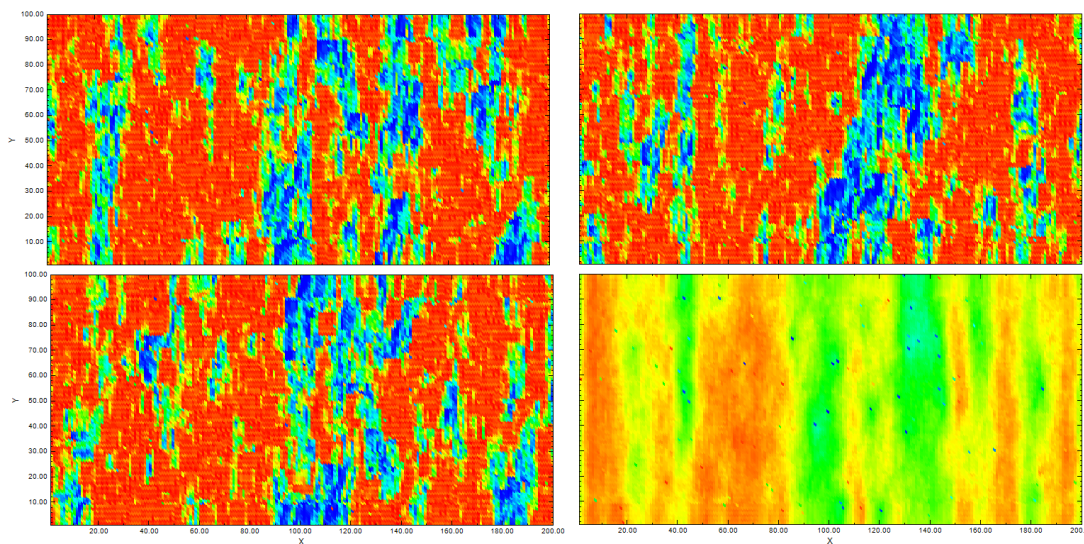
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 0.0$



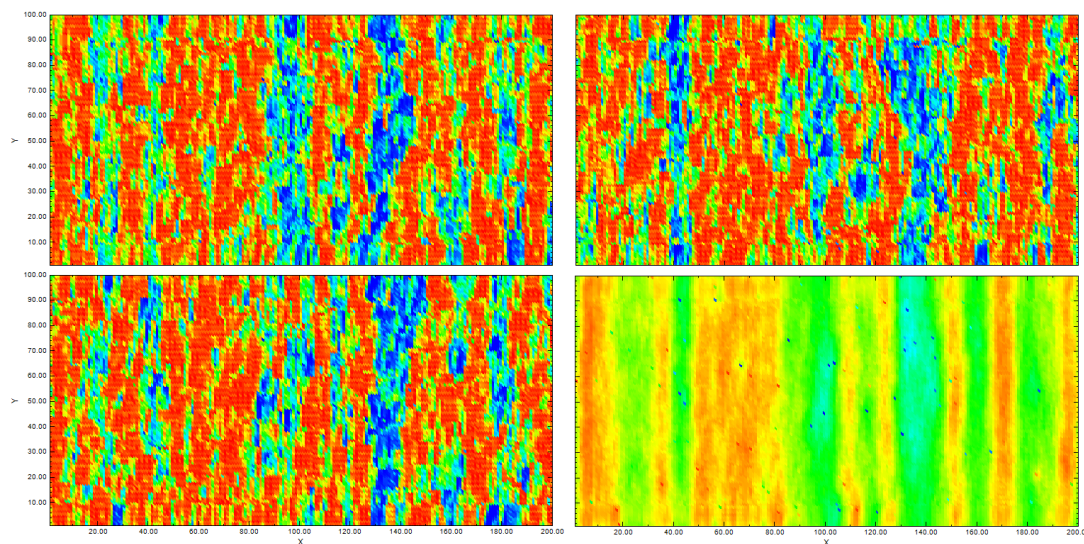
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 0.2$



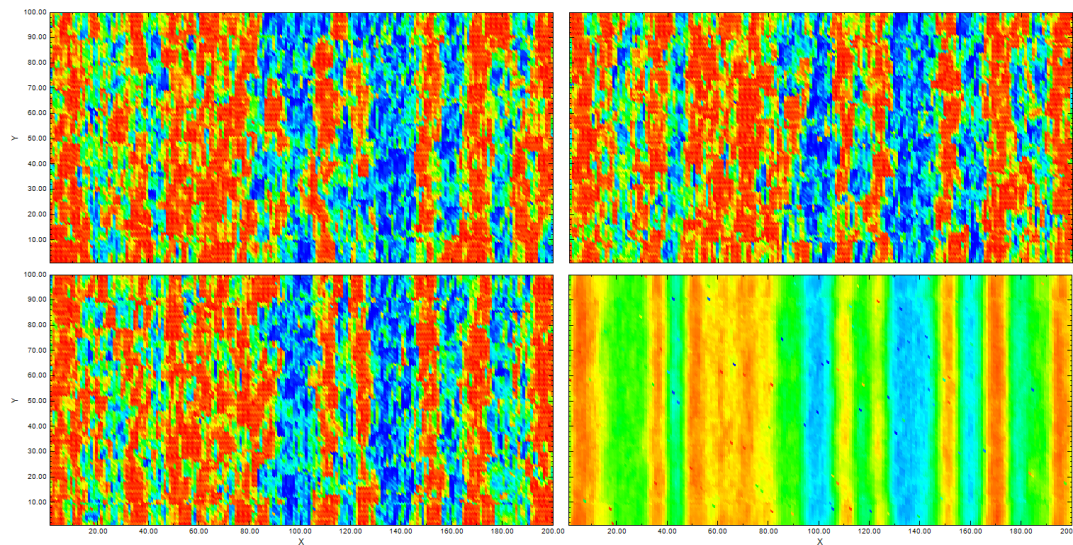
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 0.4$



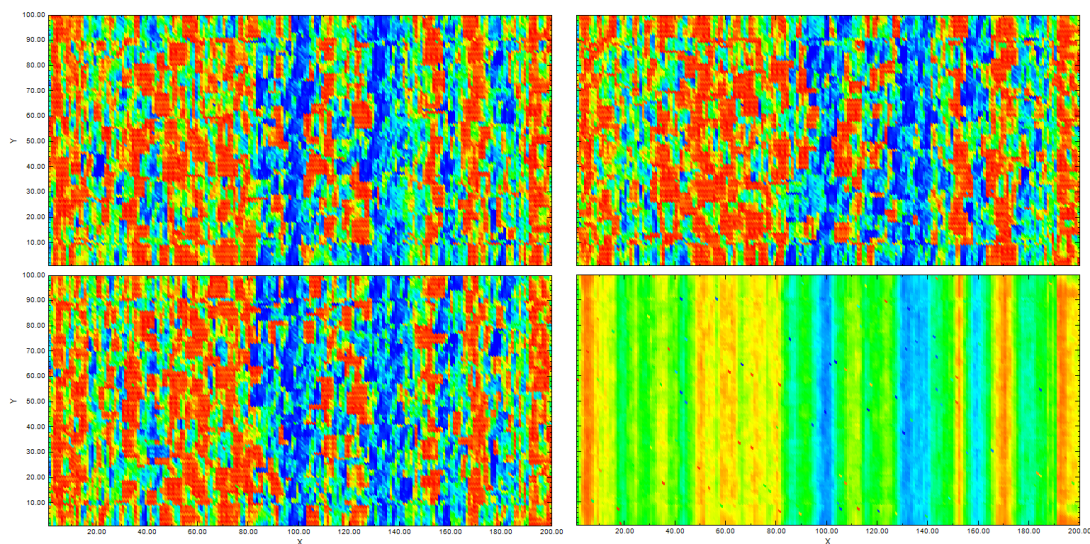
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 0.6$



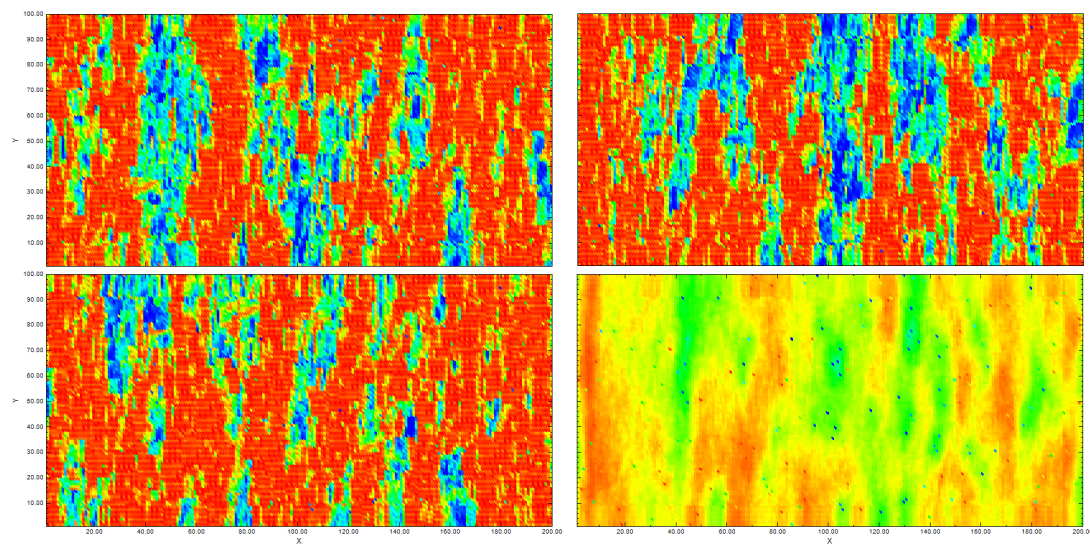
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 0.8$



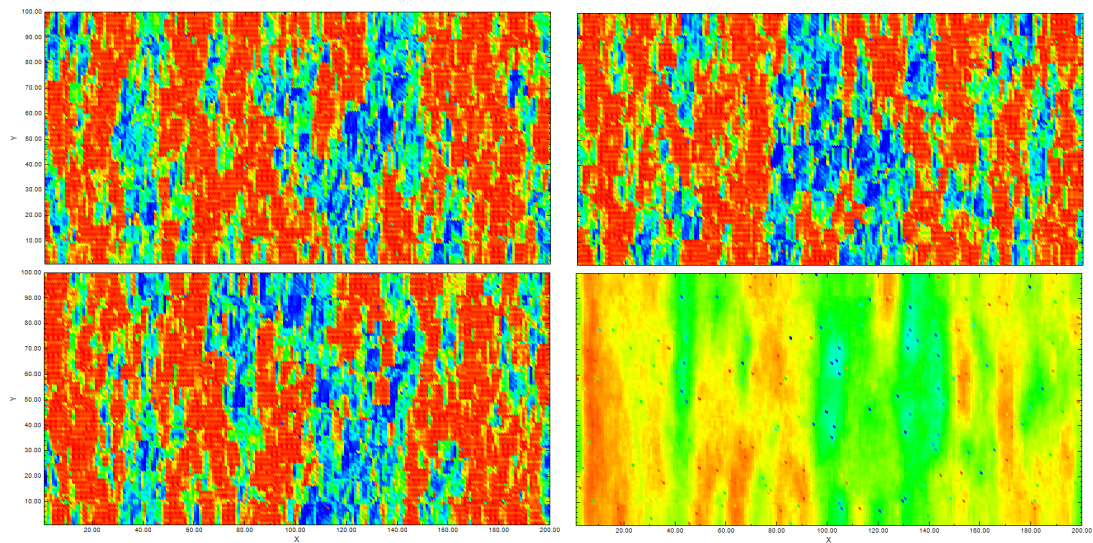
200 D.C. - 50 % D.C. - Media local Impuesta desde GT - $w_0 : 1.0$



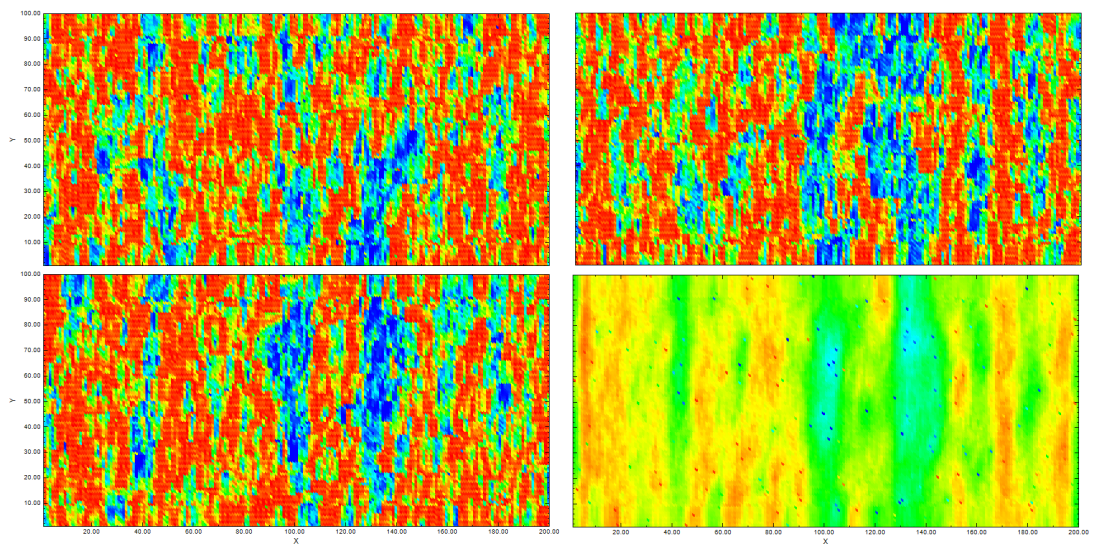
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 0.0$



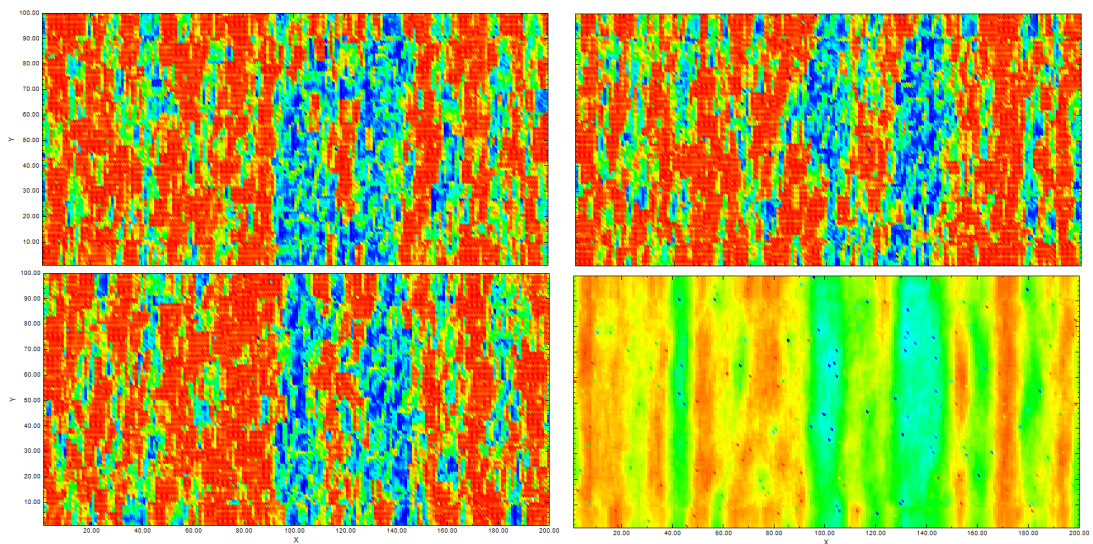
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 0.2$



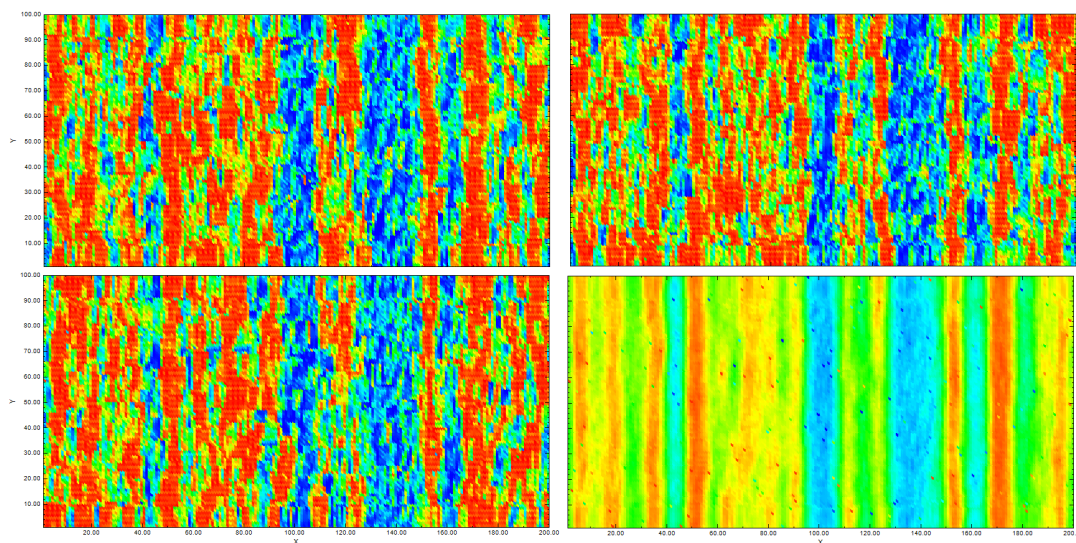
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 0.4$



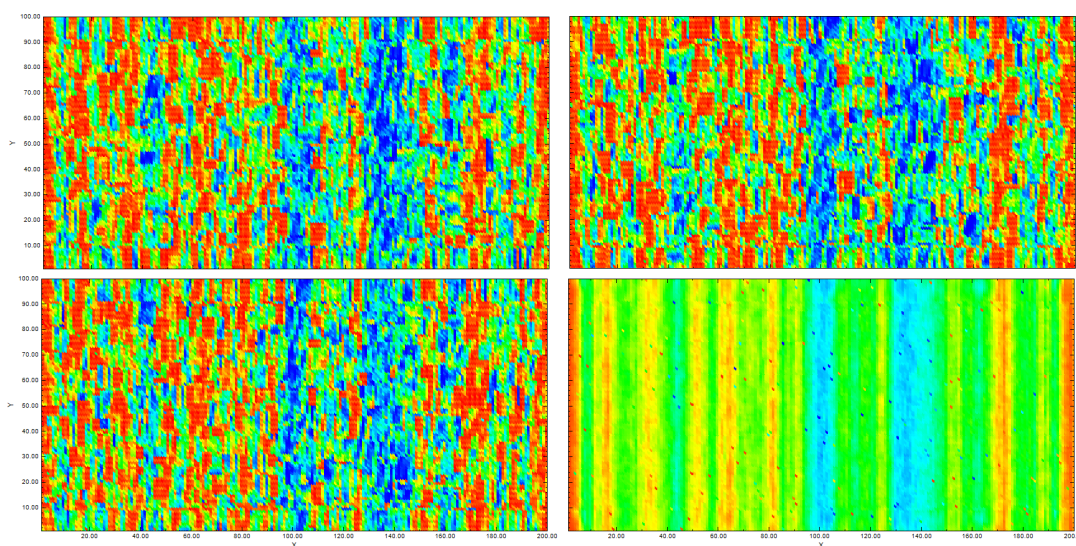
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 0.6$



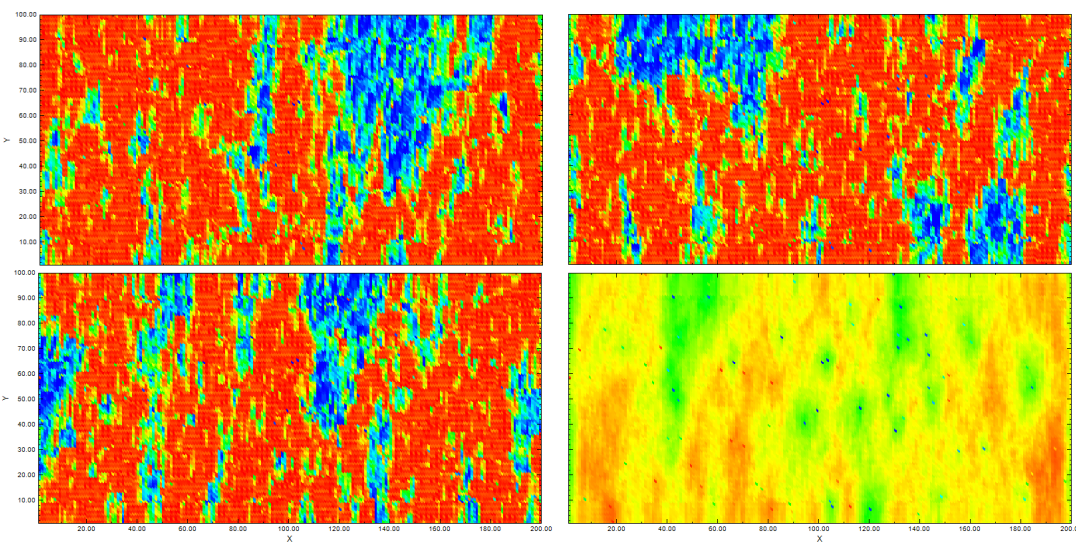
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 0.8$



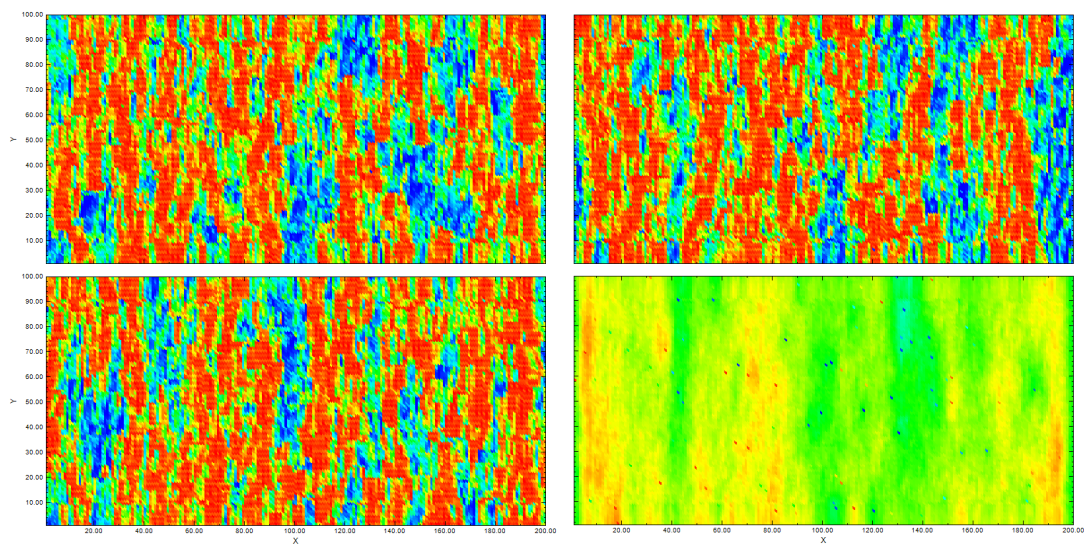
200 D.C. - 100 % D.C. - Media local Inferida desde GT - $w_0 : 1.0$



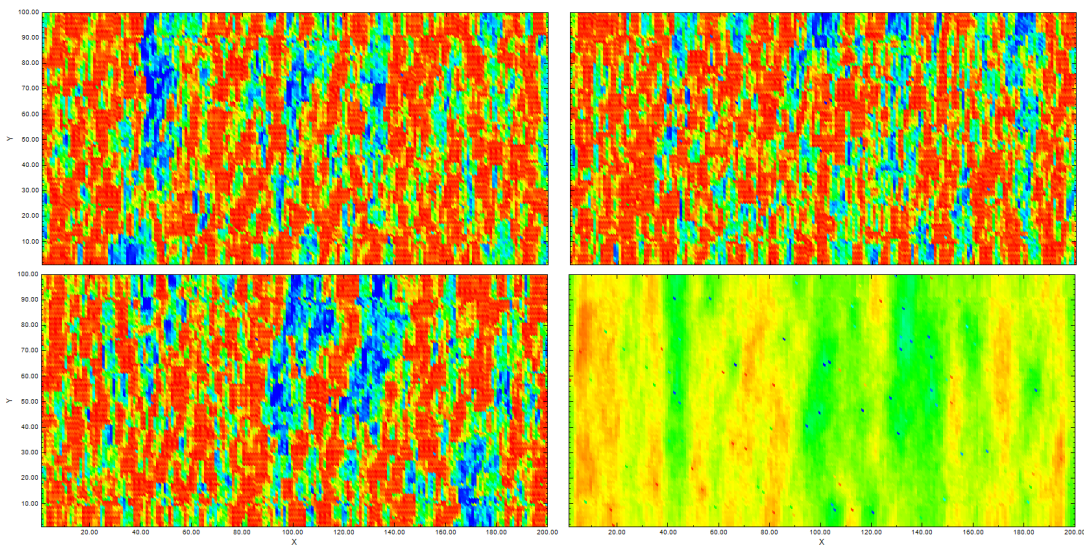
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 0.0$



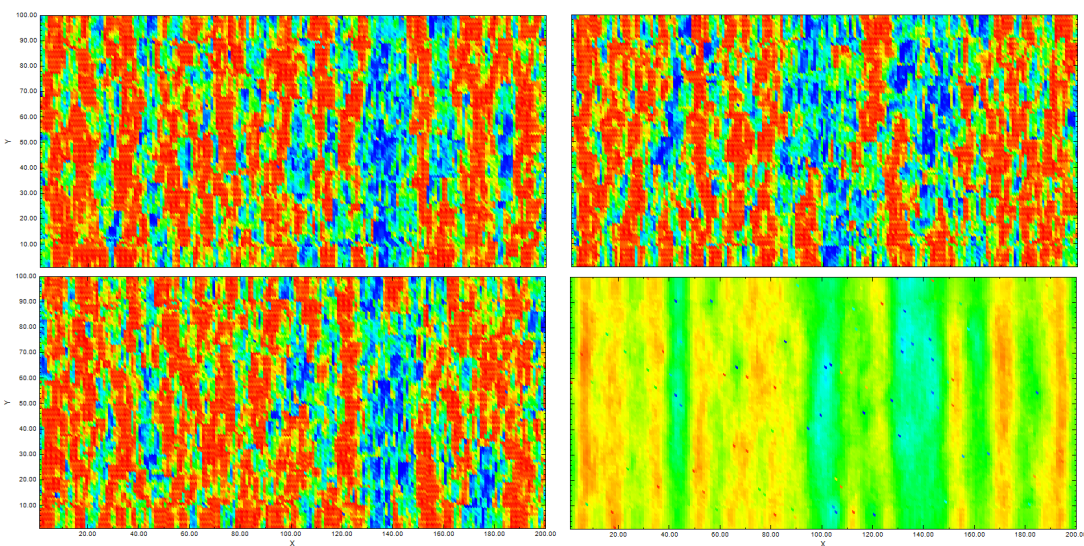
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 0.2$



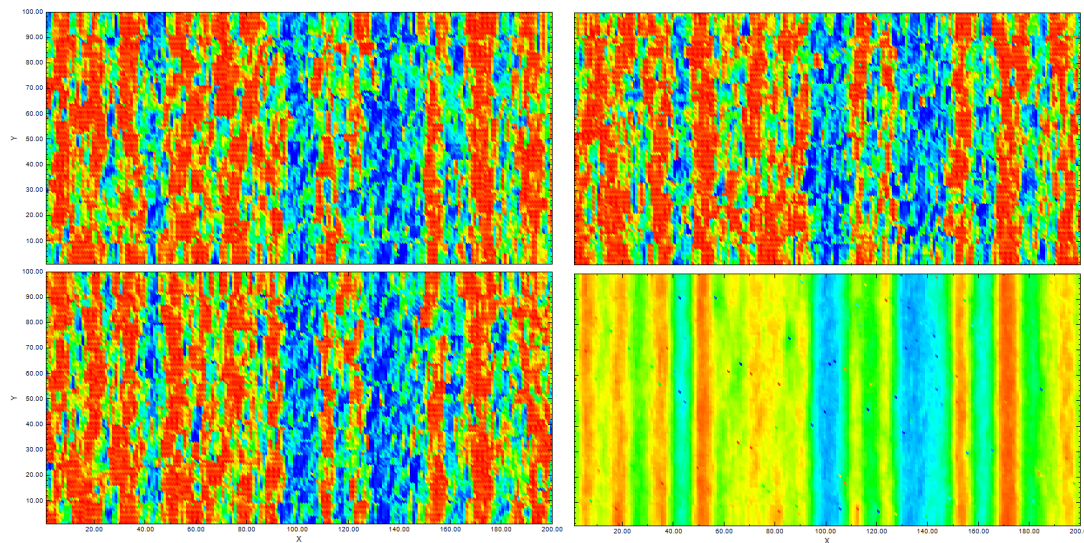
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 0.4$



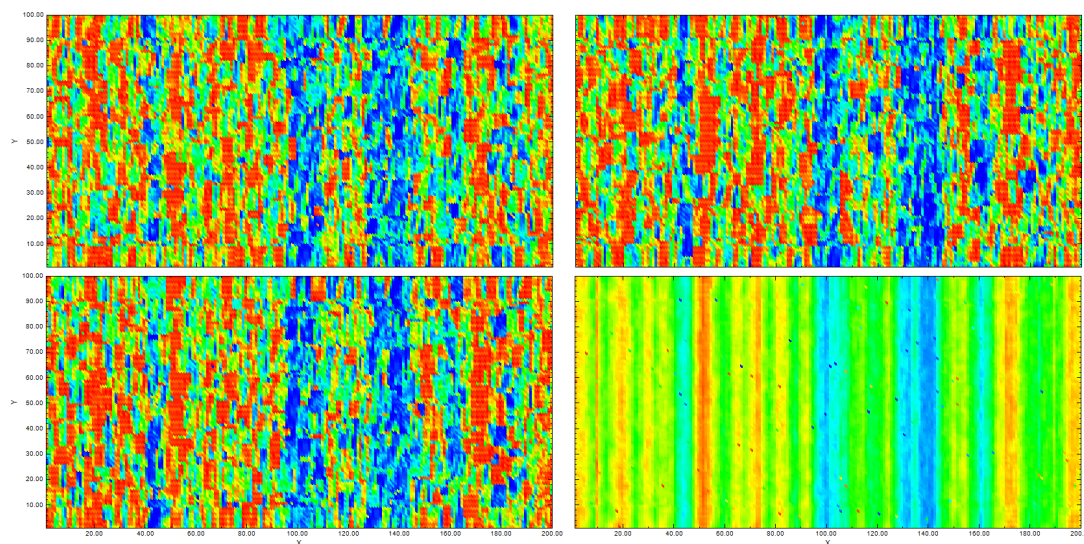
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 0.6$



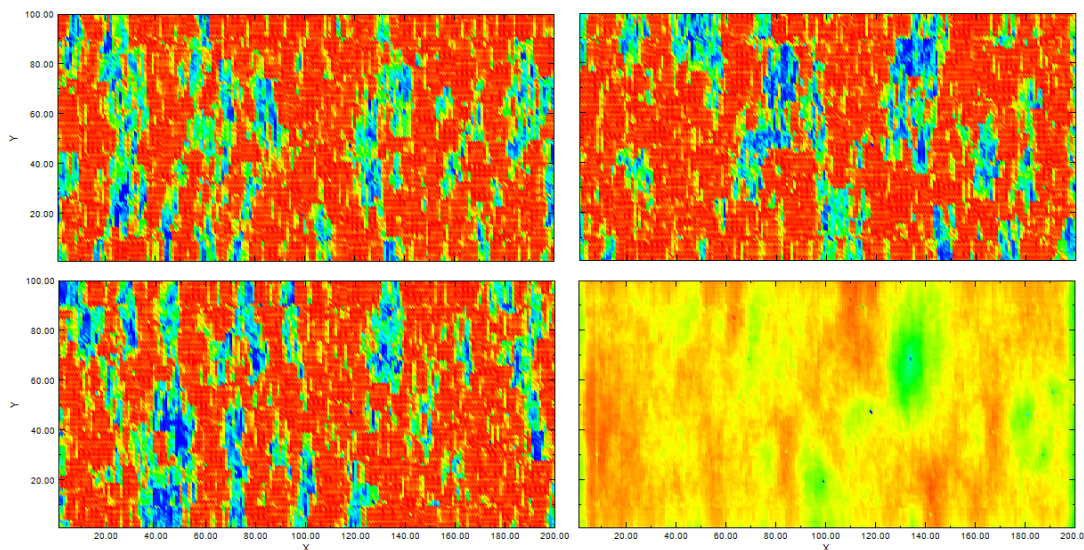
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 0.8$



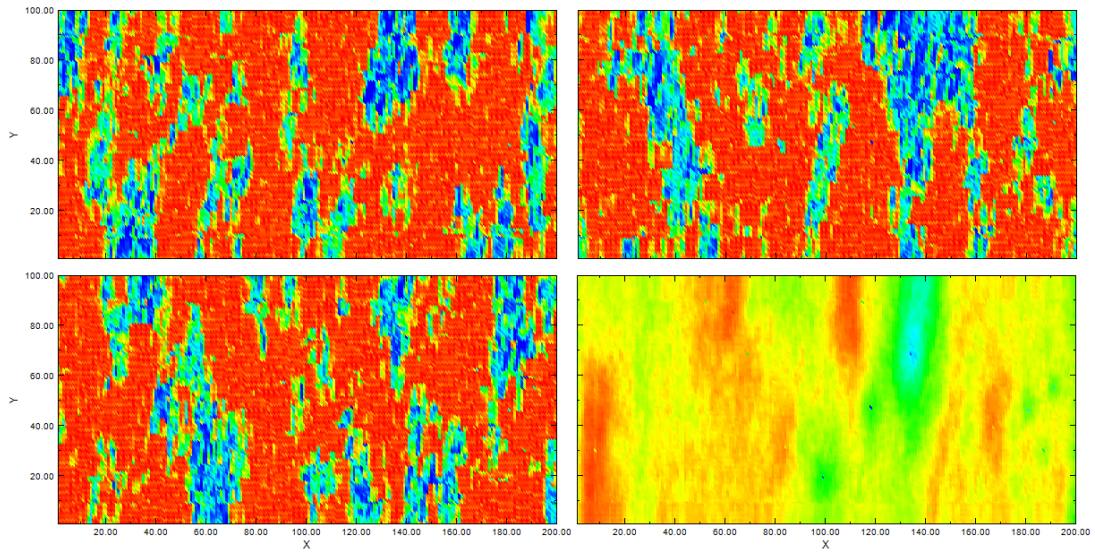
200 D.C. - 50 % D.C. - Media local Inferida desde GT - $w_0 : 1.0$



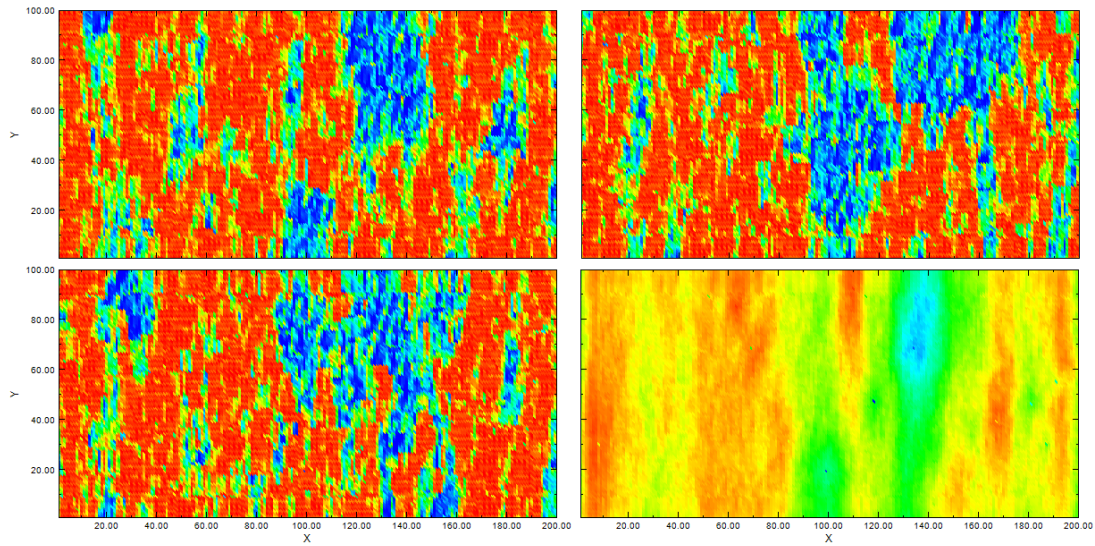
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.0$



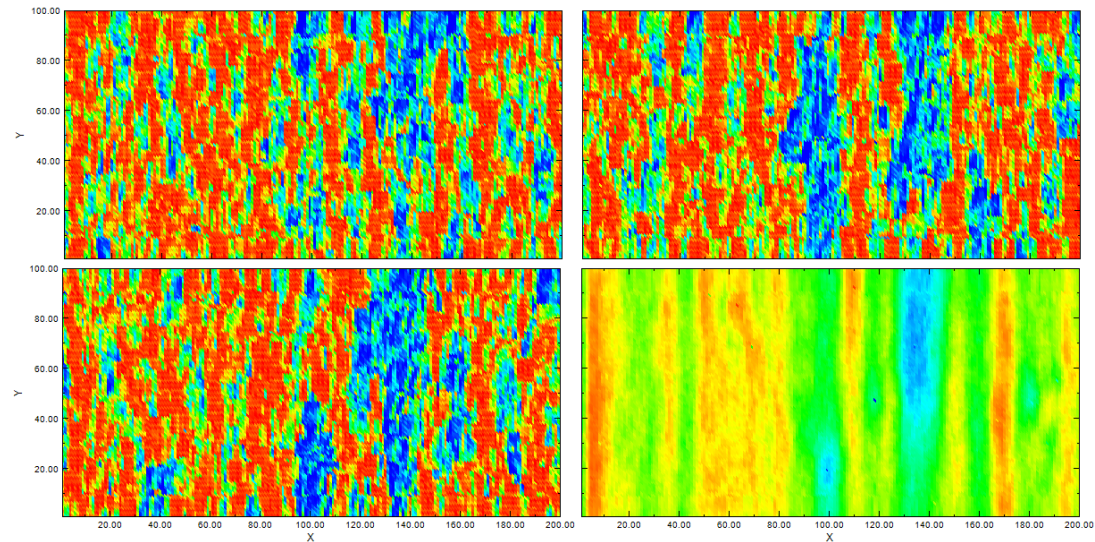
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.2$



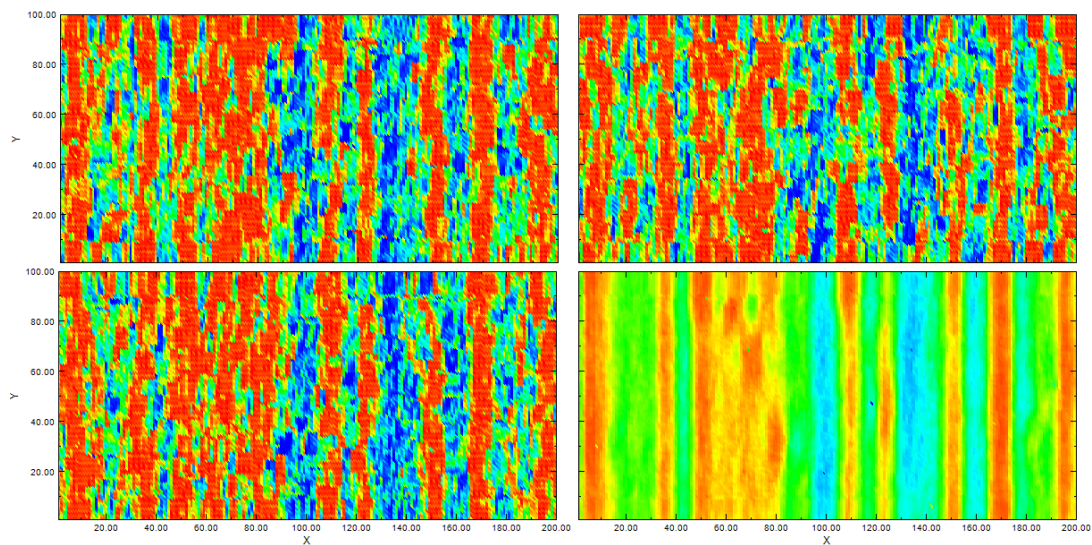
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.4$



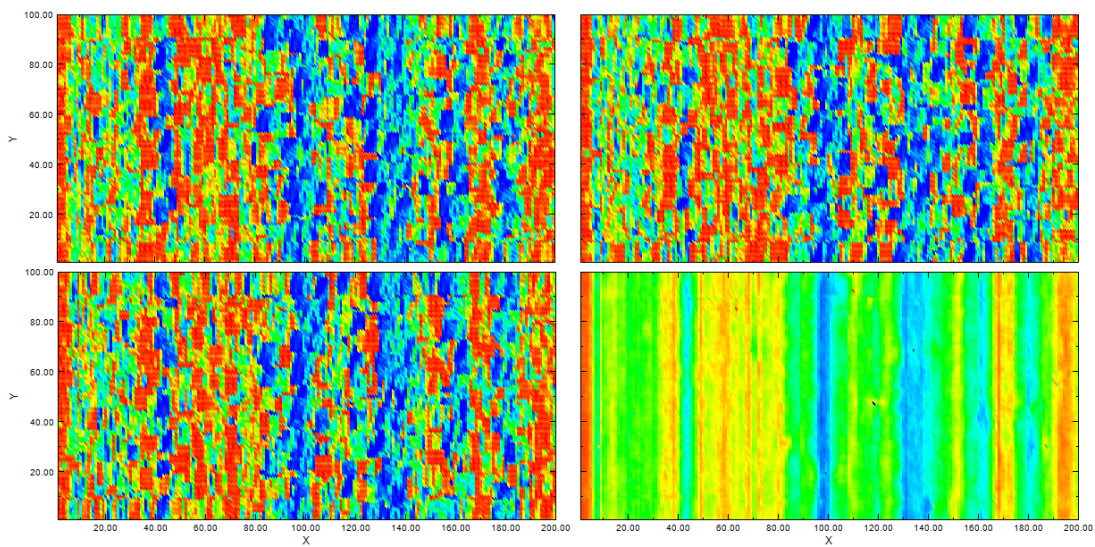
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.6$



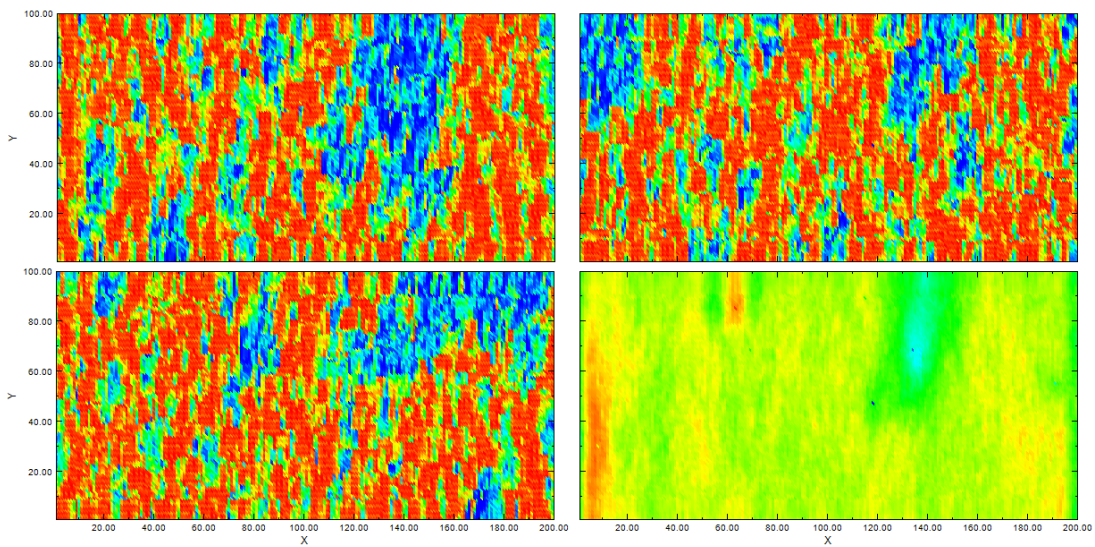
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.8$



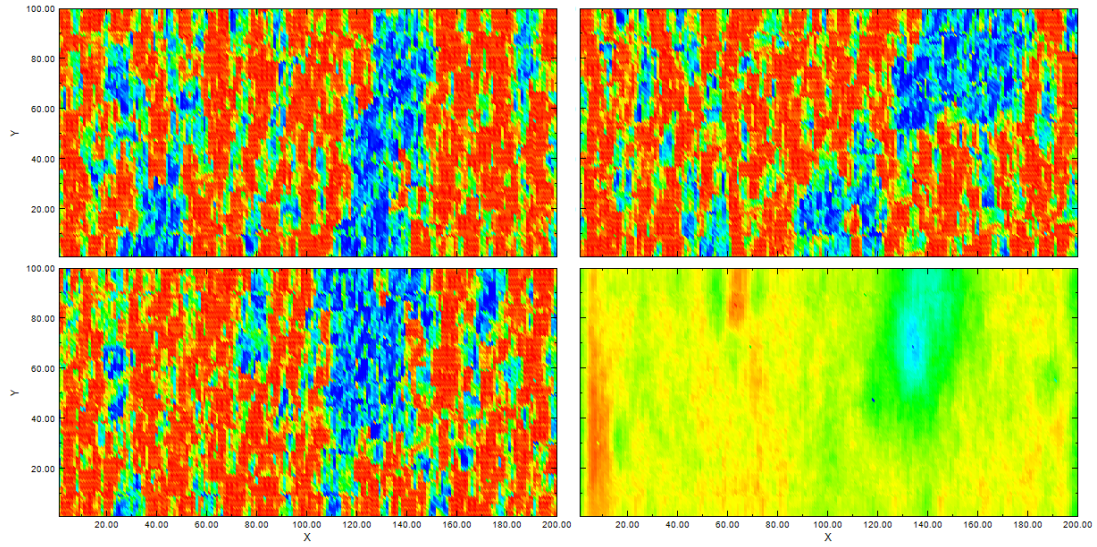
15 D.C. - 100 % D.C. - Media local Impuesta desde los DC - $w_0 : 1.0$



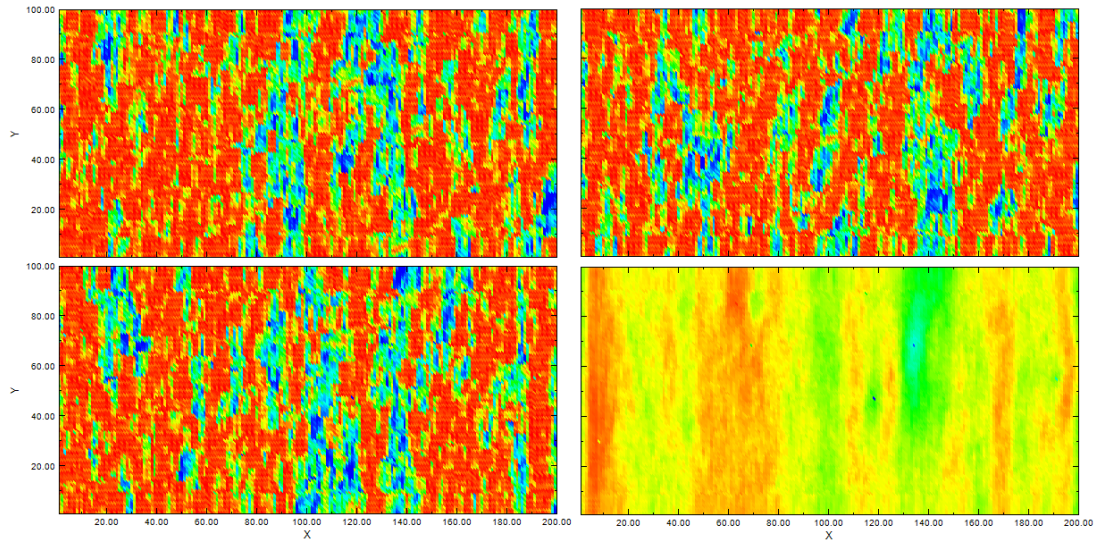
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.0$



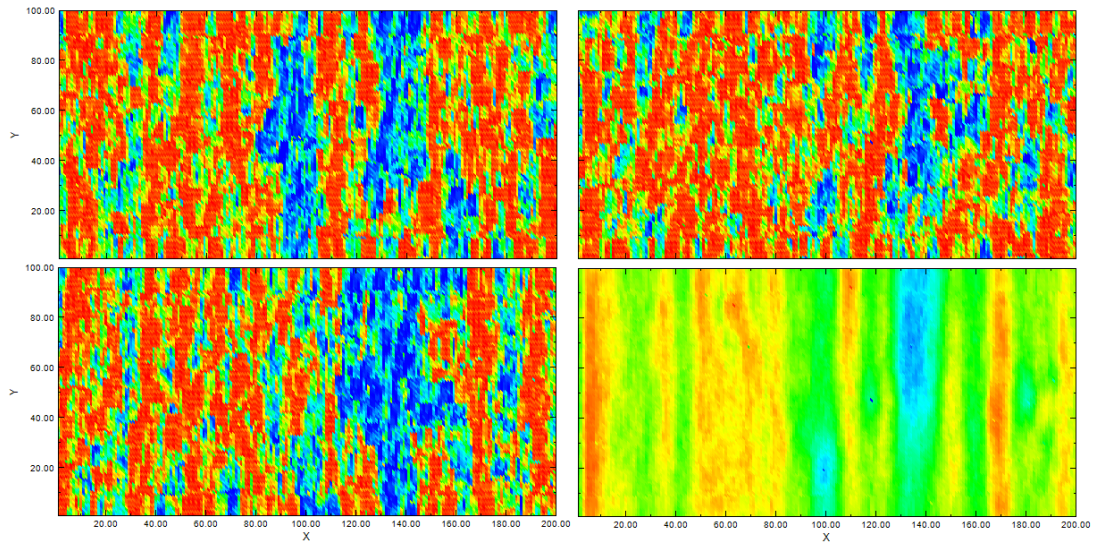
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.2$



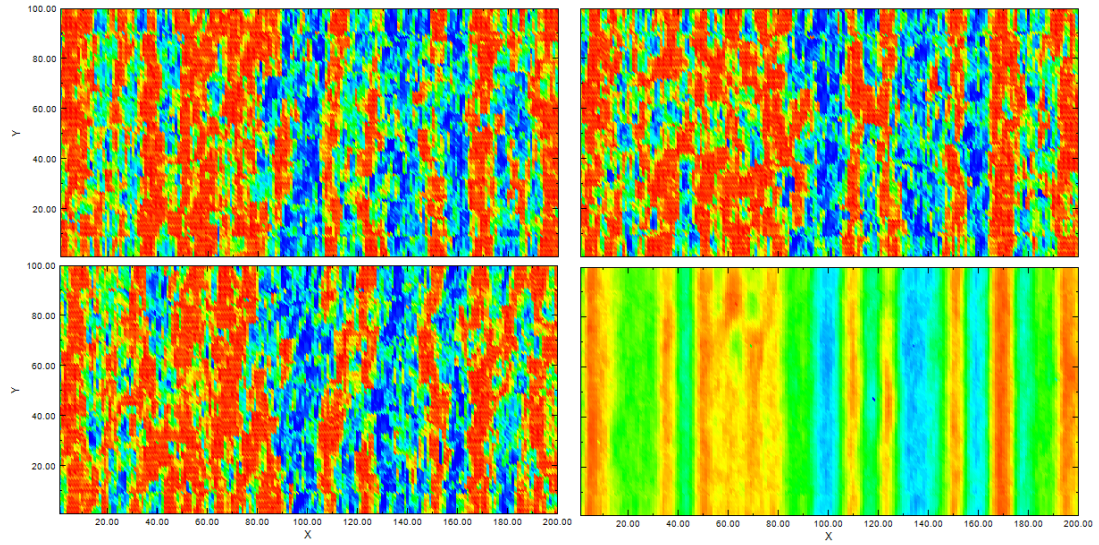
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.4$



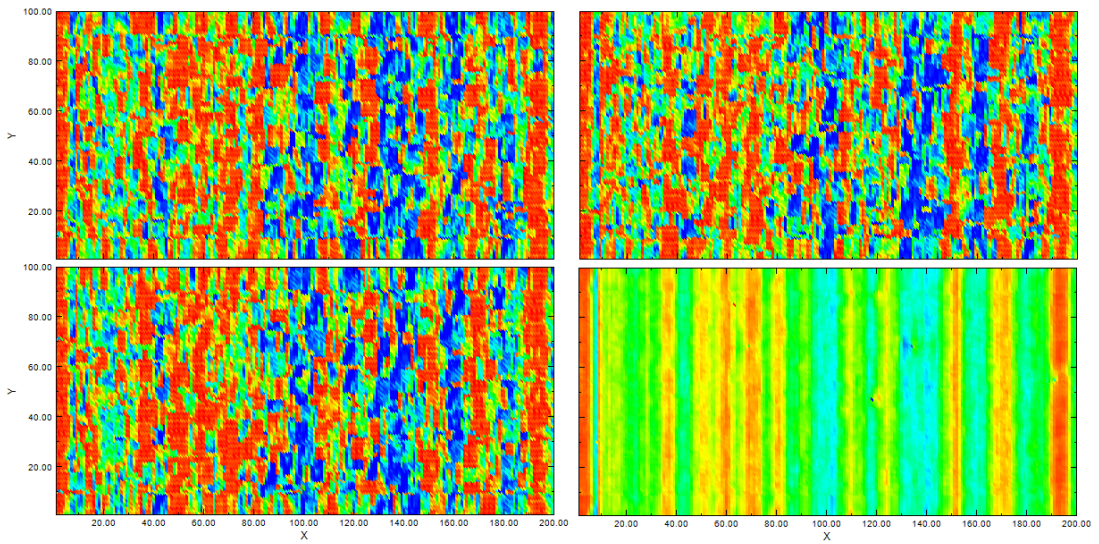
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.6$



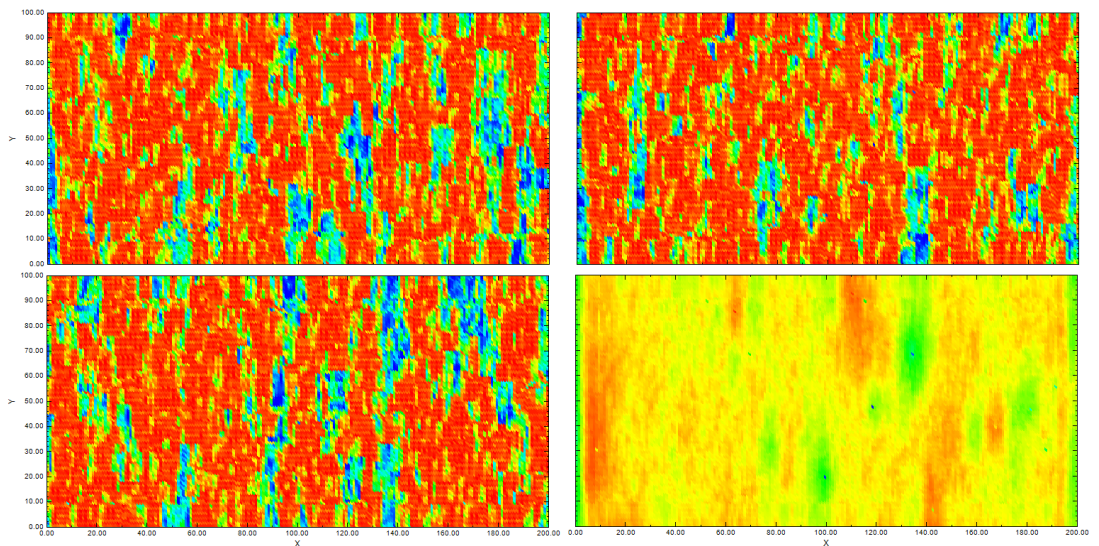
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 0.8$



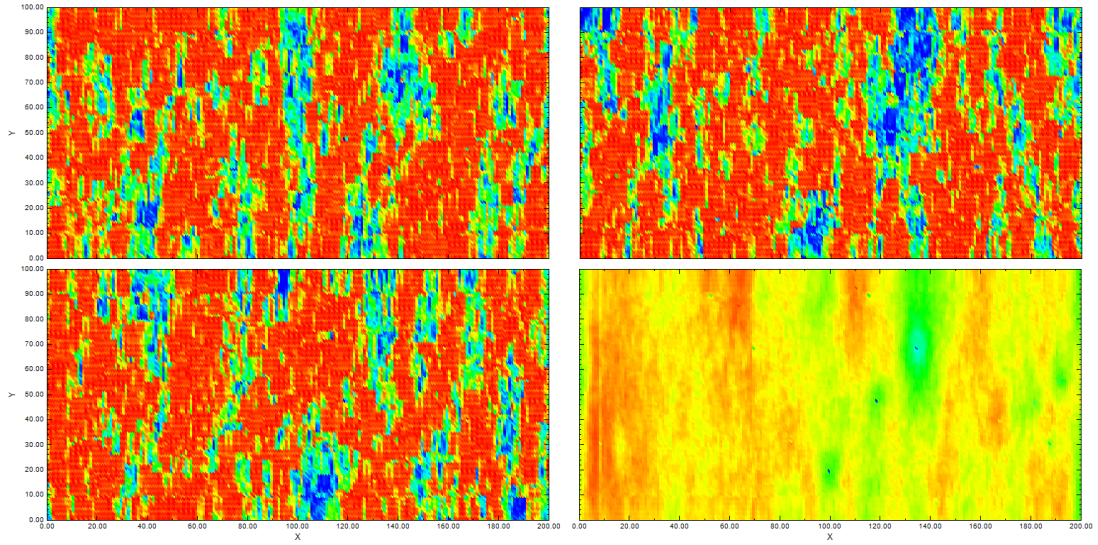
15 D.C. - 50 % D.C. - Media local Impuesta desde los DC - $w_0 : 1.0$



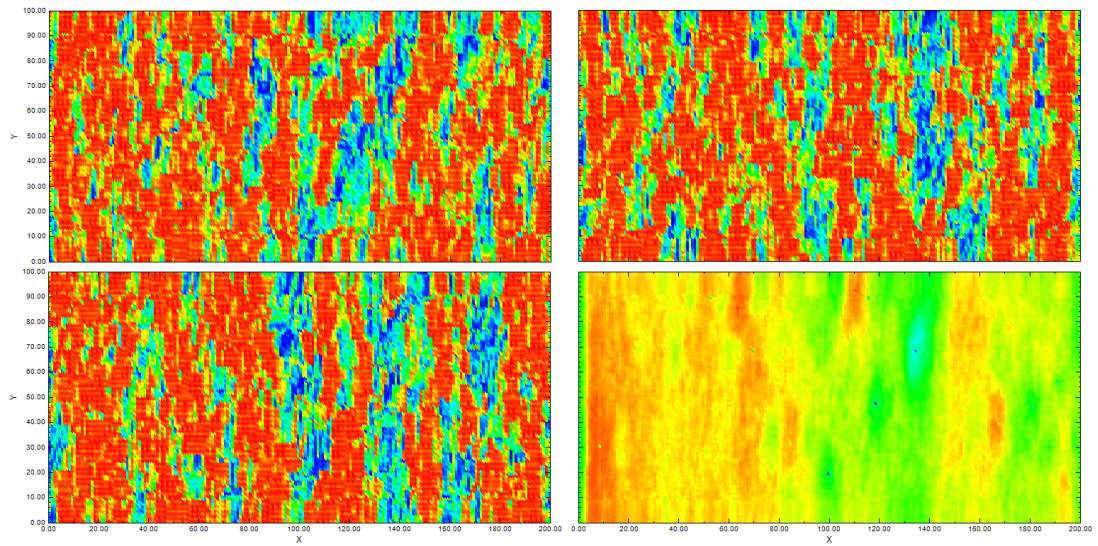
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 0.0$



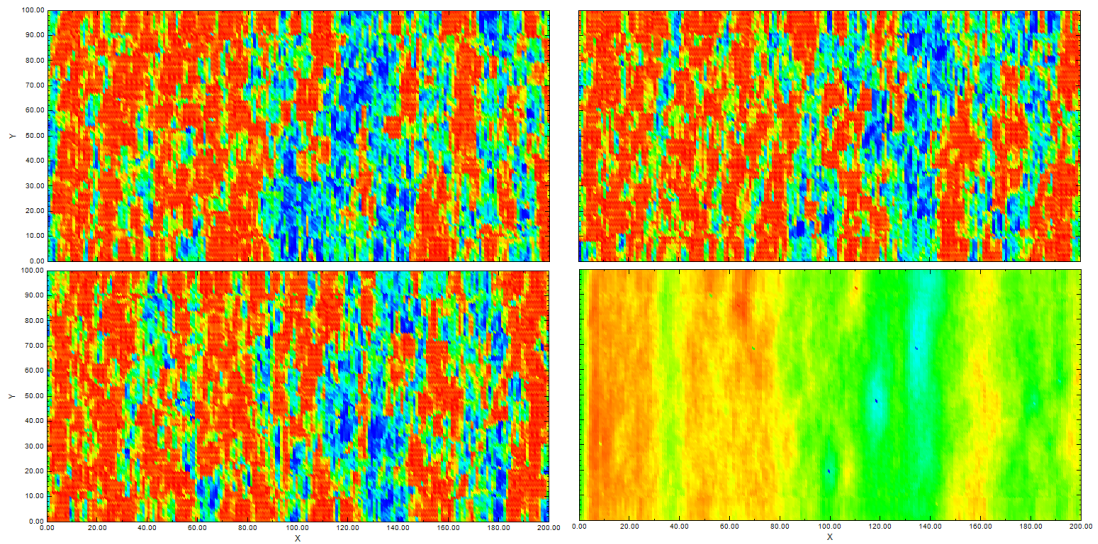
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 0.2$



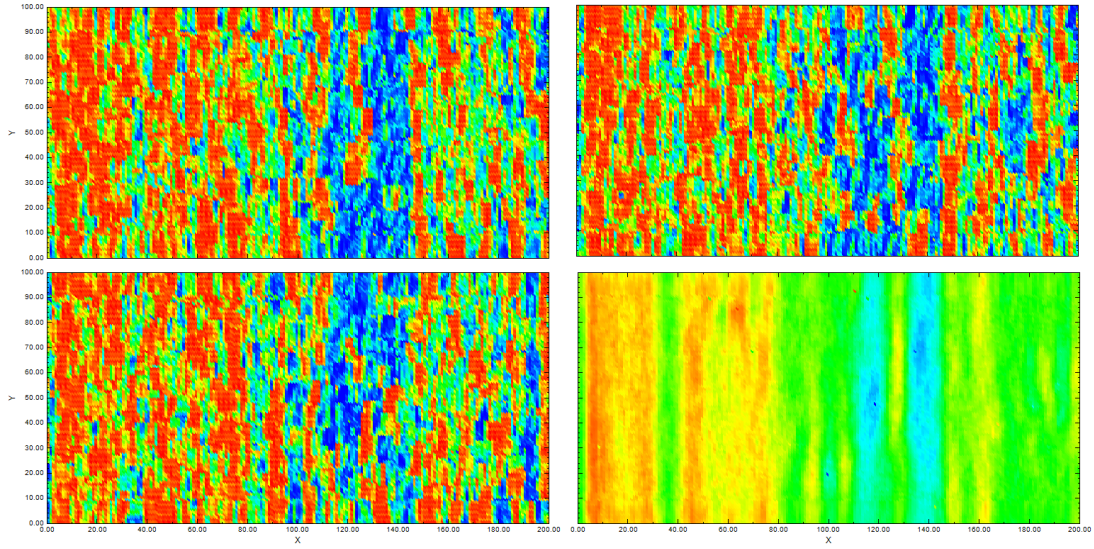
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 0.4$



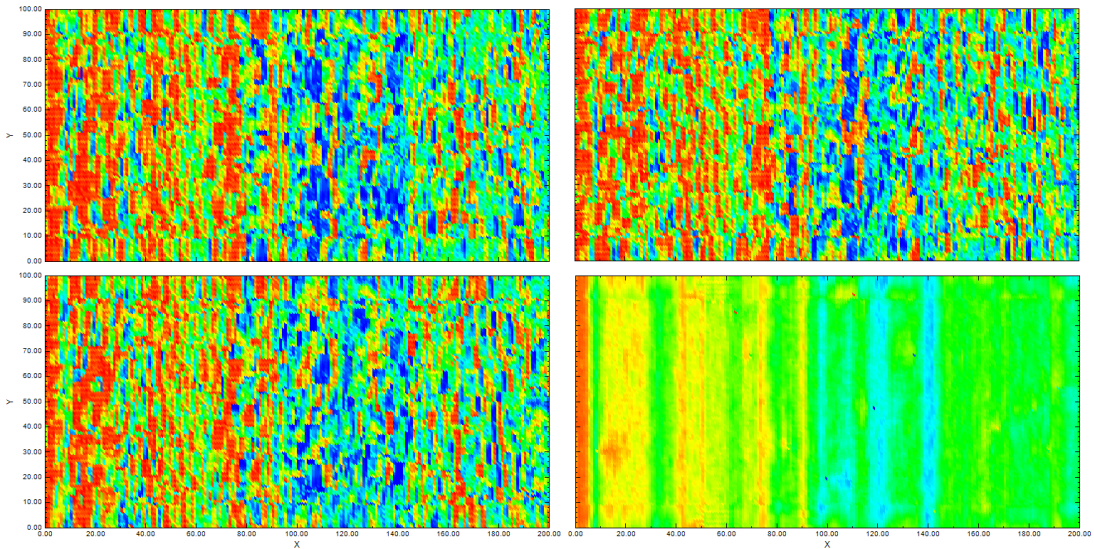
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 0.6$



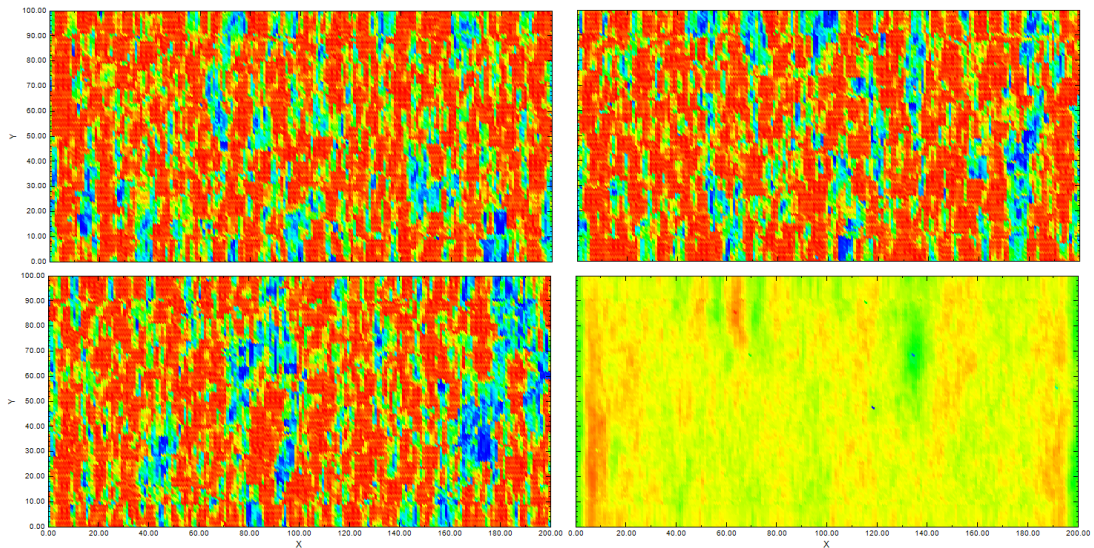
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 0.8$



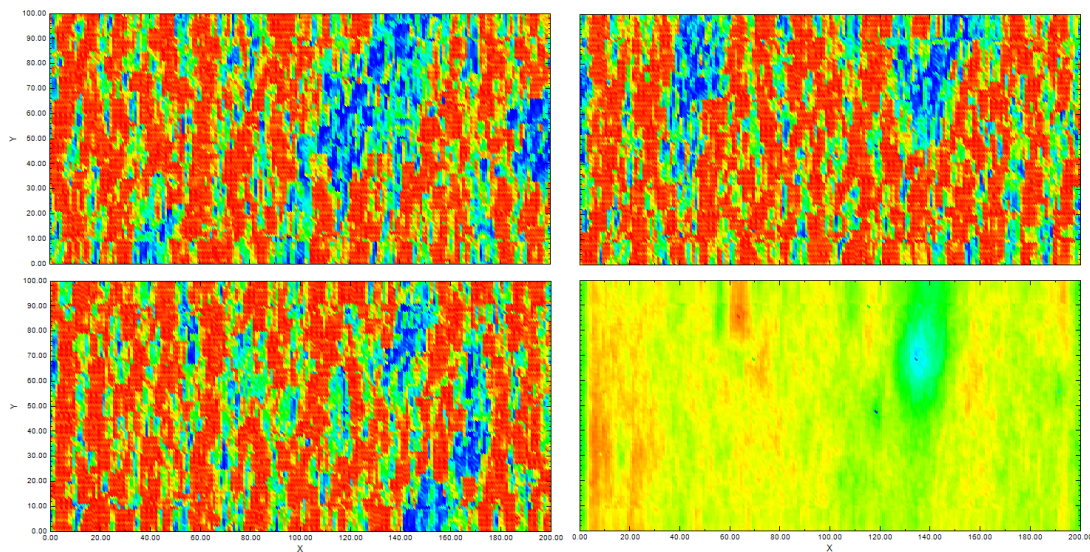
15 D.C. - 100 % D.C. - Media local Inferida desde los DC - $w_0 : 1.0$



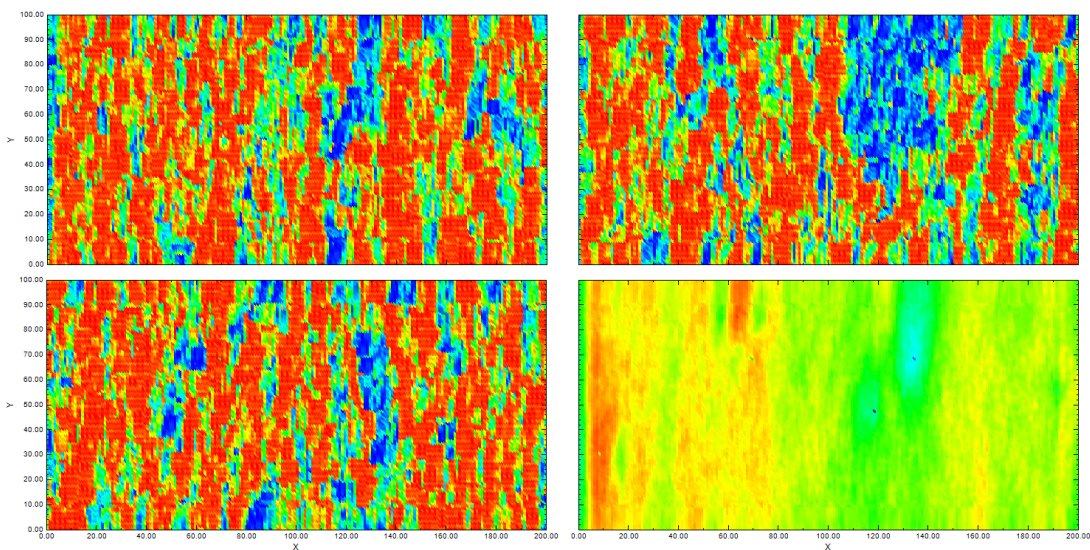
15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 0.0$



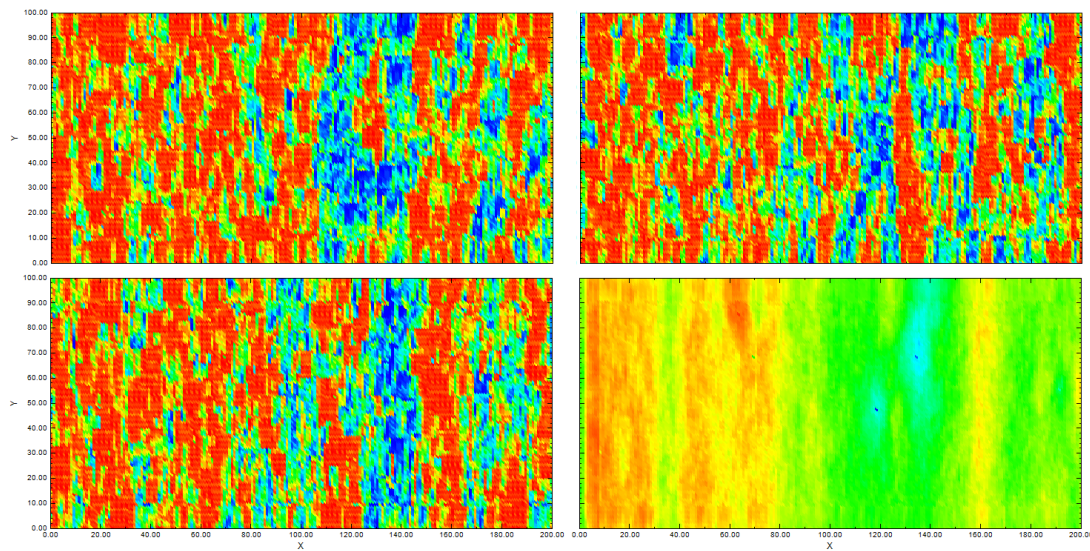
15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 0.2$



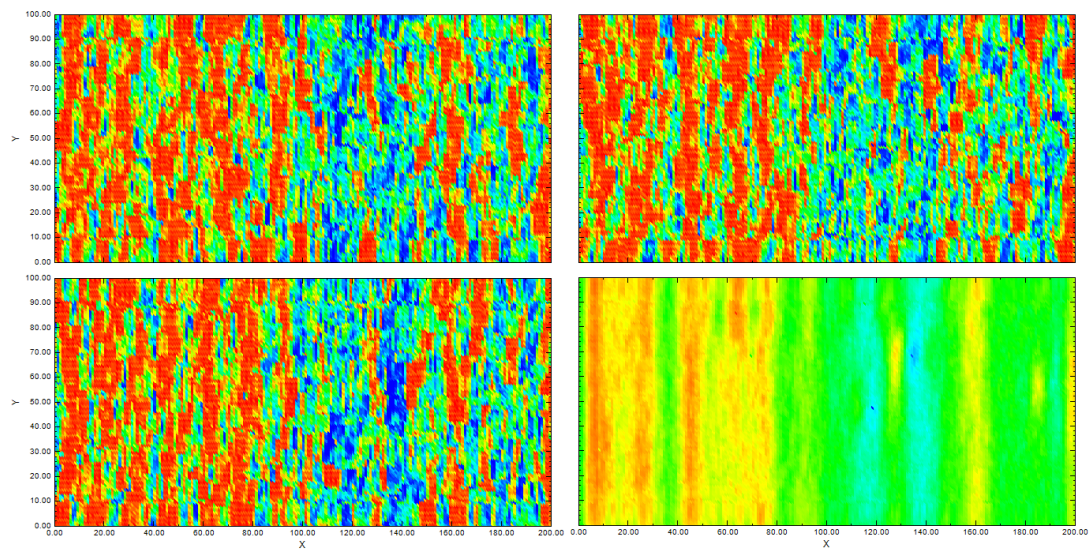
15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 0.4$



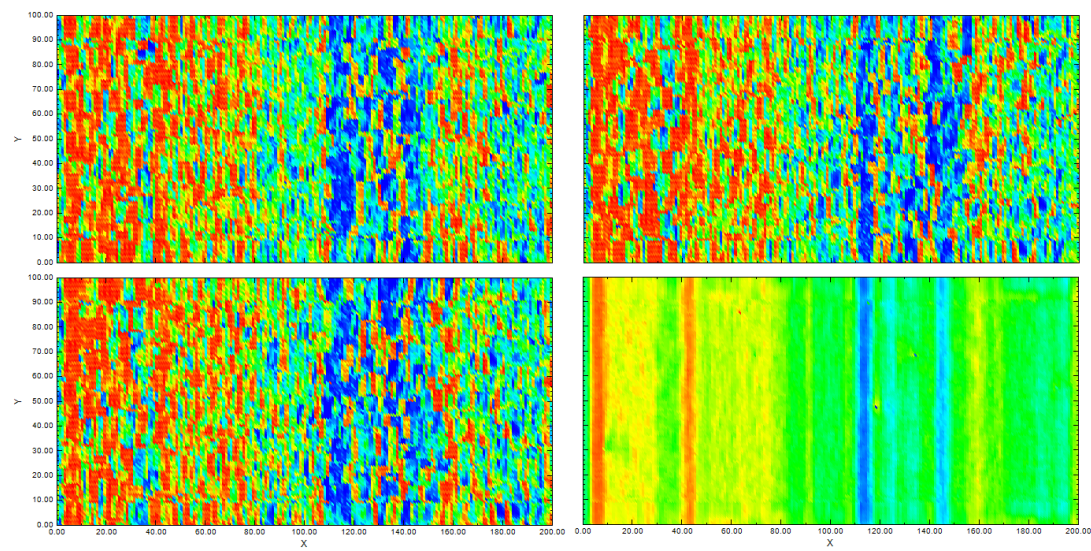
15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 0.6$



15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 0.8$



15 D.C. - 50 % D.C. - Media local Inferida desde los DC - $w_0 : 1.0$



7.3. Anexo C

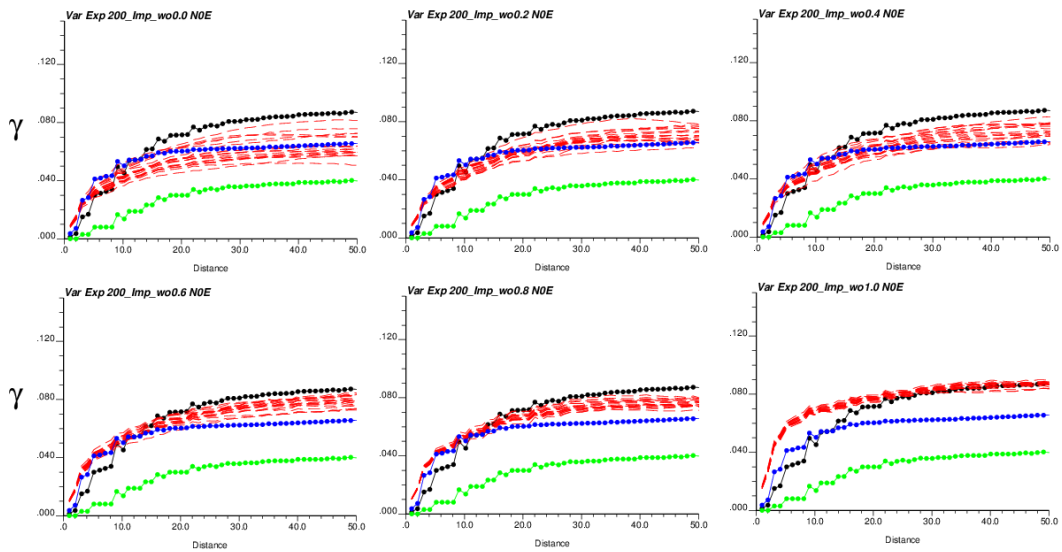
La presente sección entrega los variogramas experimentales de las simulaciones en el depósito sedimentario. Se presenta en color negro el variograma experimental sobre los valores del Ground Truth, en color azul sobre la Imagen de entrenamiento, en color verde sobre la media local impuesta/inferida y en color rojo sobre 20 de las 100 realizaciones seleccionadas de manera aleatoria.

Variogramas experimentales dirección N0° E

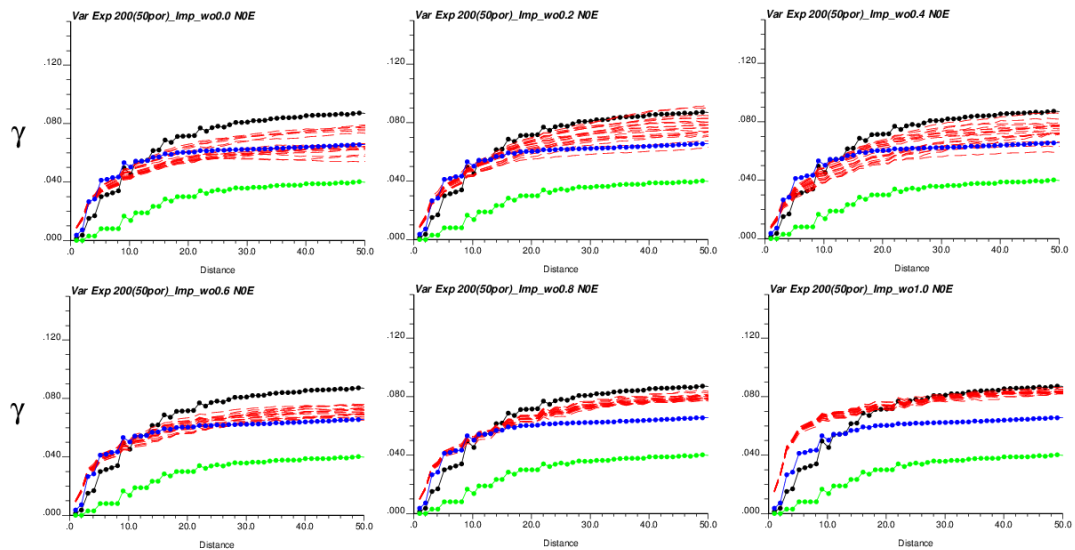
Los parámetros de entrada utilizados son los siguientes:

Azimut	0°	Dip	0°	Ancho Vertical	20.0
Tol. Azimut	22.5°	Tol. Dip	22.5°	Ancho Horizontal	20.0
N° pasos	50	Dist. Paso	1	Tol. Paso	0

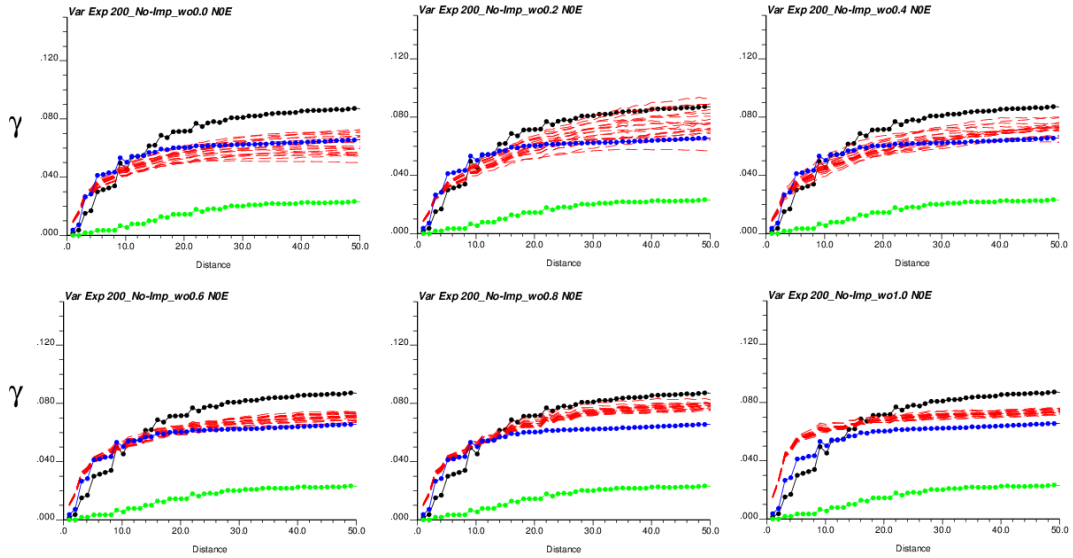
200 D.C. – 100% D.C. – Media local Impuesta desde GT



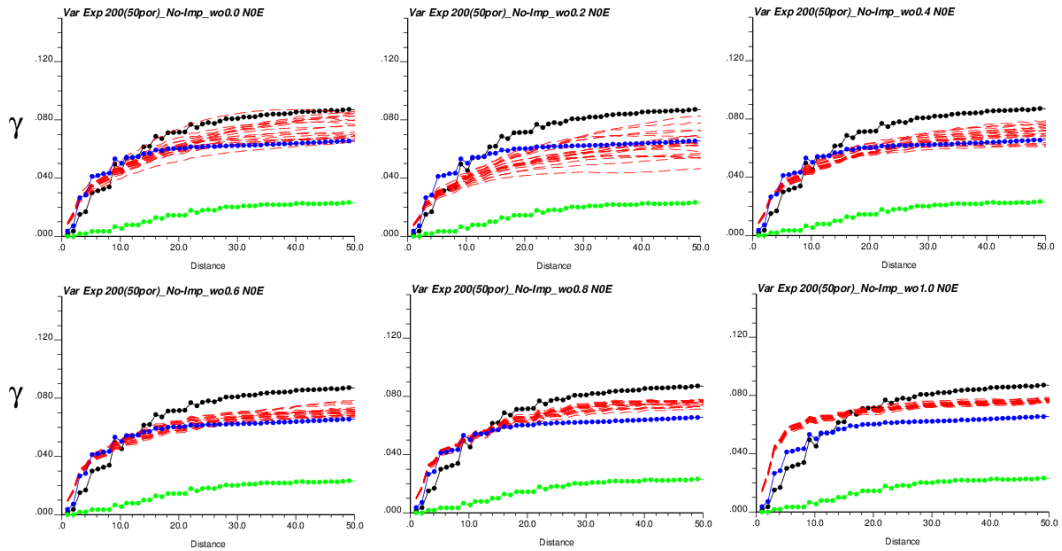
200 D.C. – 50% D.C. – Media local Impuesta desde GT



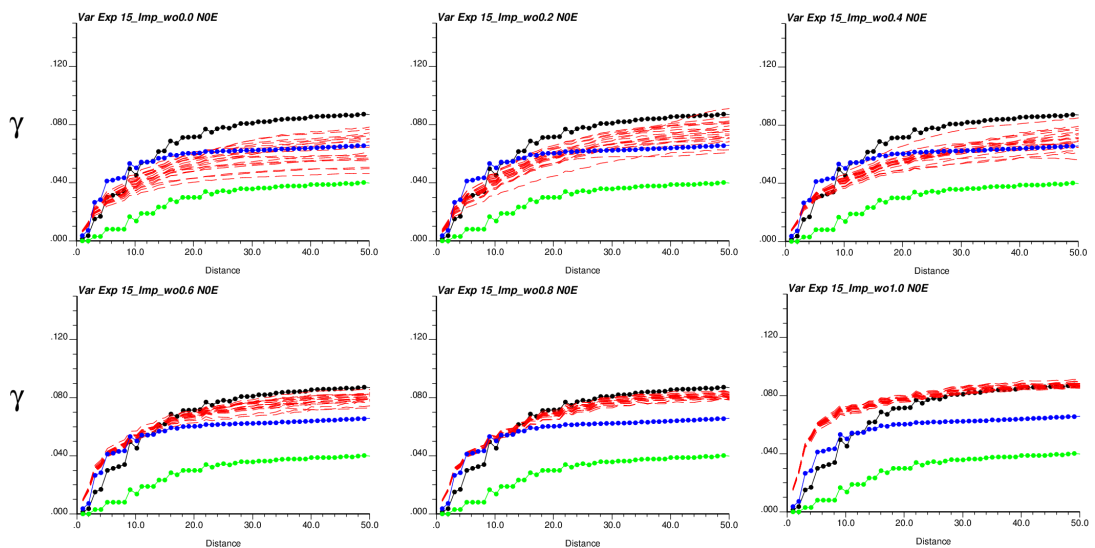
200 D.C. - 100% D.C. - Media local Inferida desde los DC



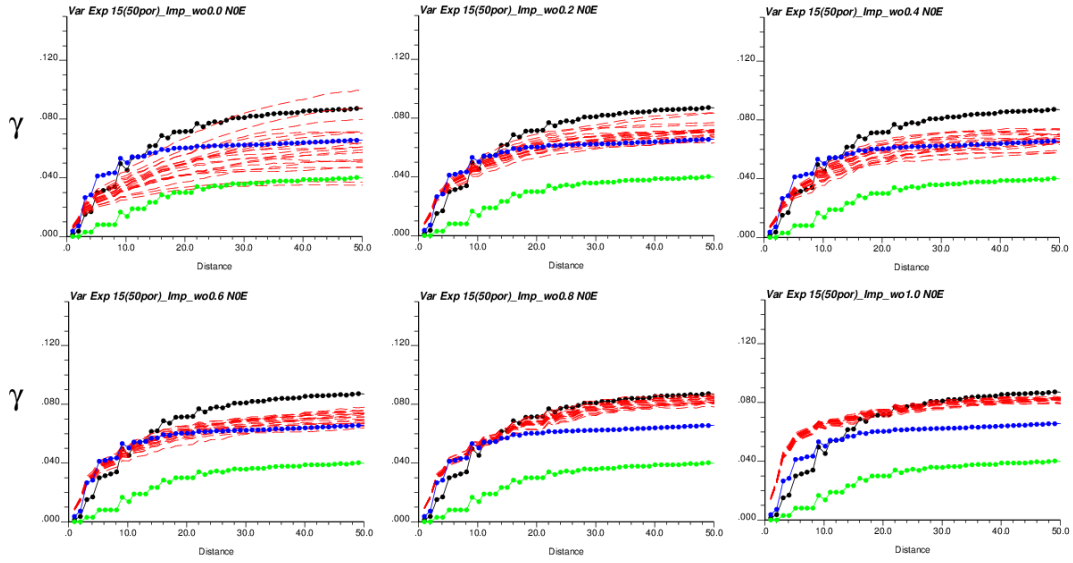
200 D.C. - 50% D.C. - Media local Inferida desde los DC



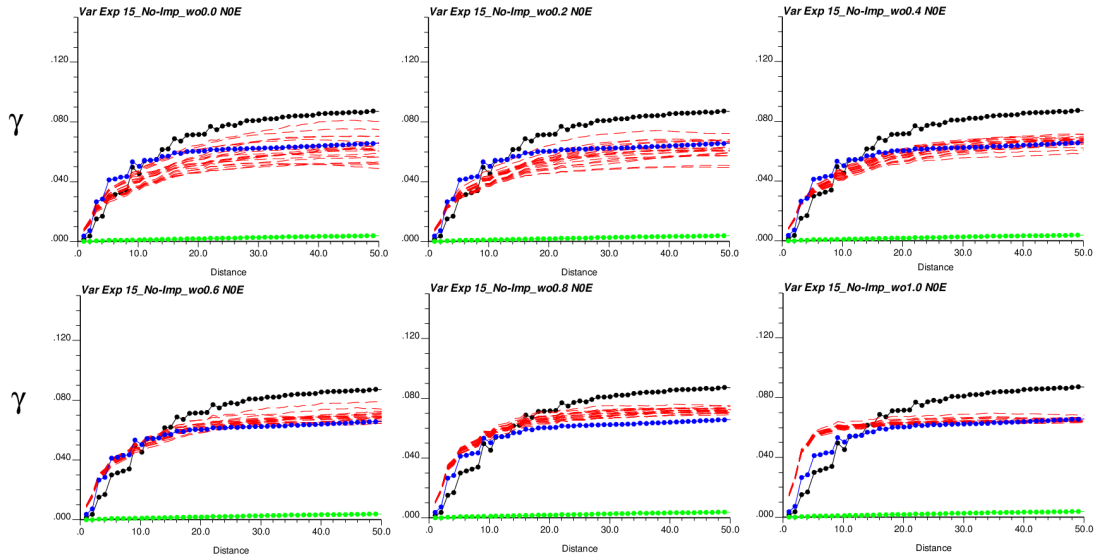
15 D.C. - 100% D.C. - Media local Impuesta desde GT



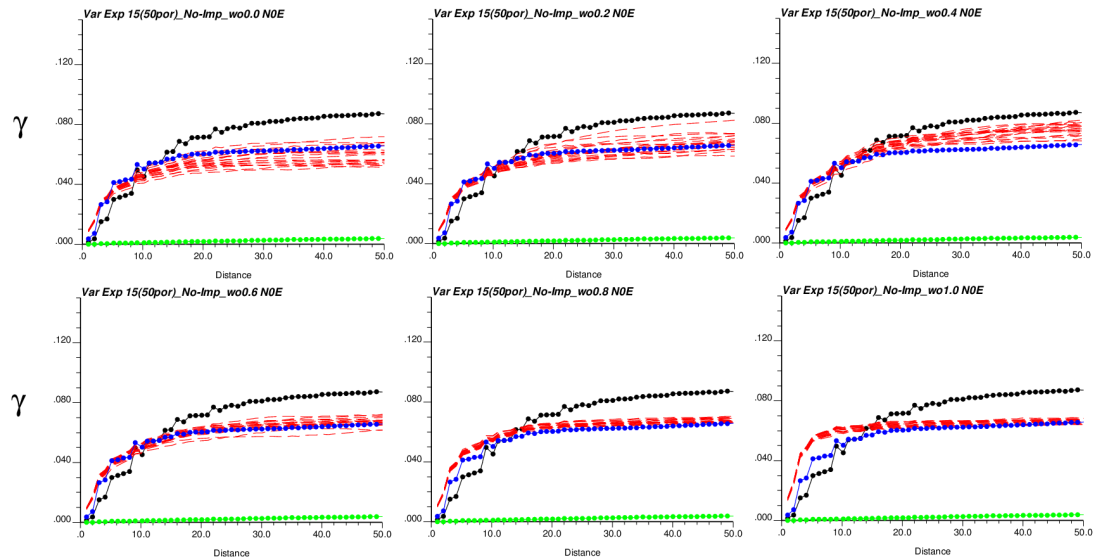
15 D.C. - 50% D.C. - Media local Impuesta desde GT



15 D.C. - 100% D.C. - Media local Inferida desde los DC



15D.C. - 50% D.C. - Media local Inferida desde los DC

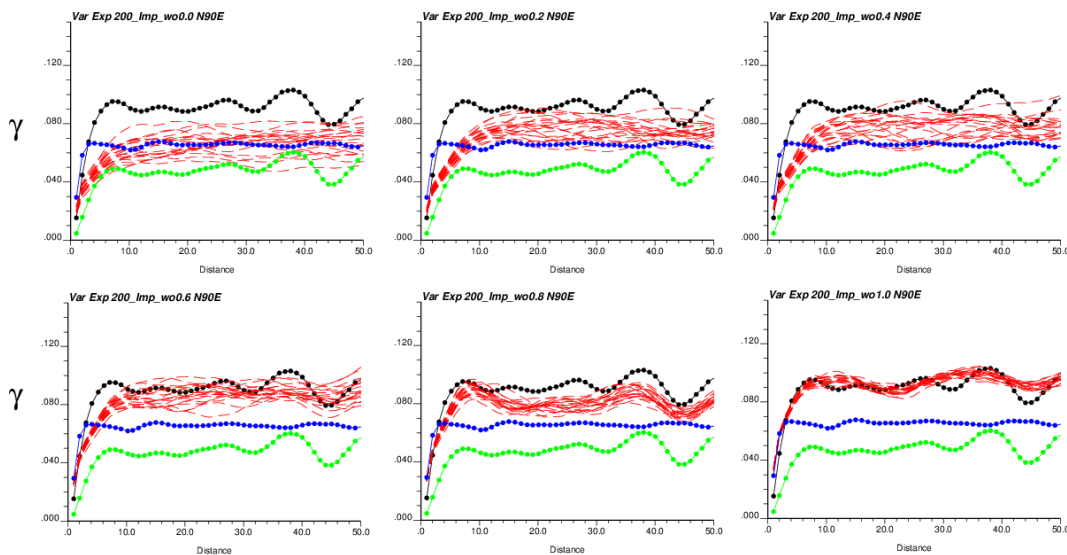


Variogramas experimentales dirección N90° E

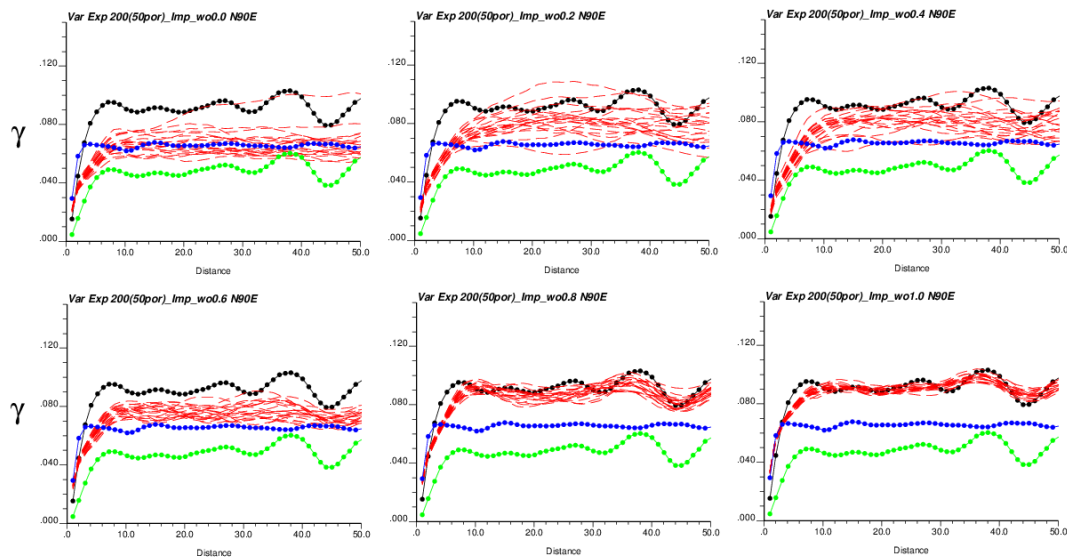
Los parámetros de entrada utilizados son los siguientes:

Azimut	90°	Dip	0°	Ancho Vertical	20.0
Tol. Azimut	22.5°	Tol. Dip	22.5°	Ancho Horizontal	20.0
N° pasos	50	Dist. Paso	1	Tol. Paso	0

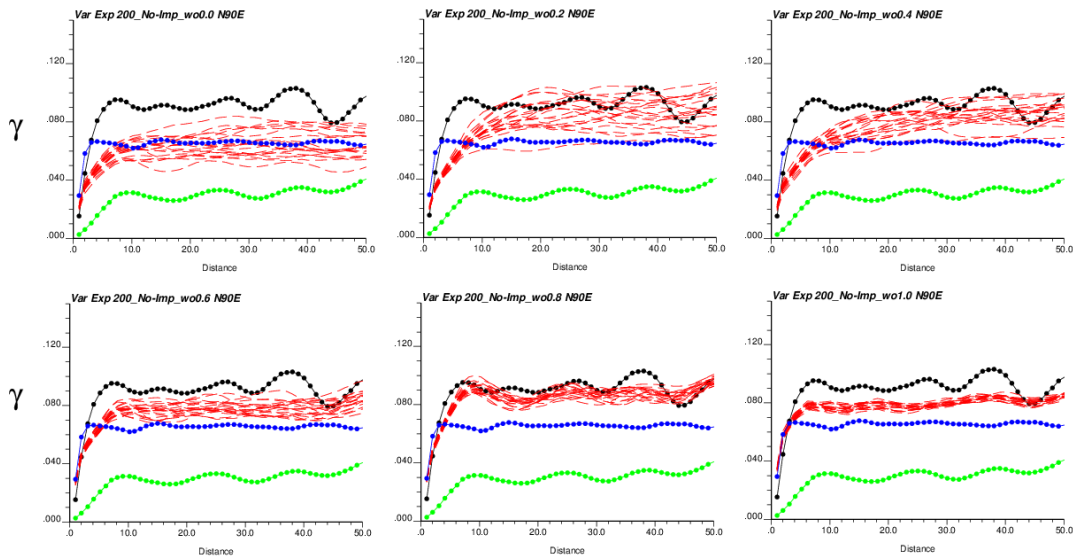
200 D.C. – 100 % D.C. – Media local Impuesta desde GT



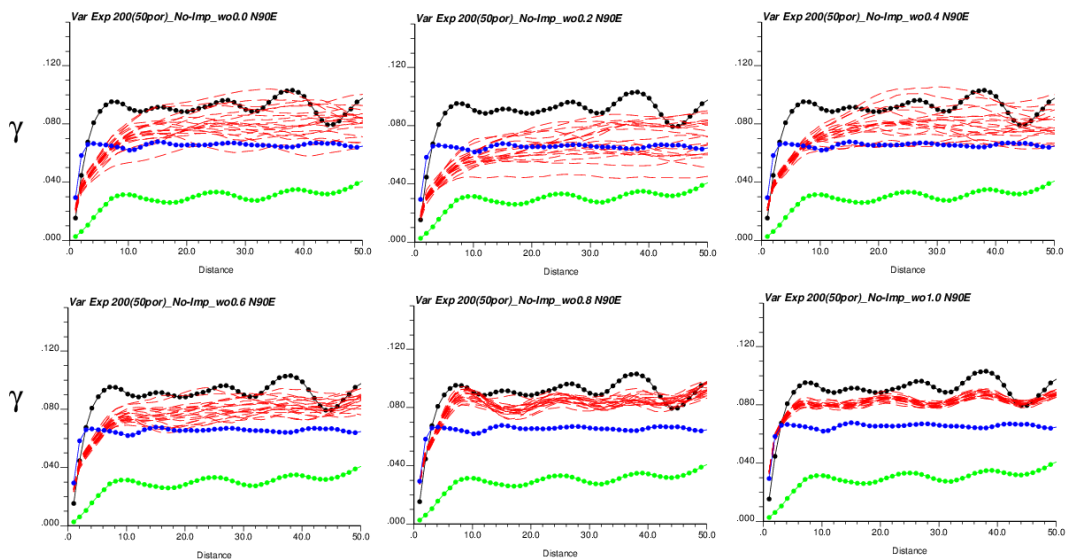
200 D.C. – 50 % D.C. – Media local Impuesta desde GT



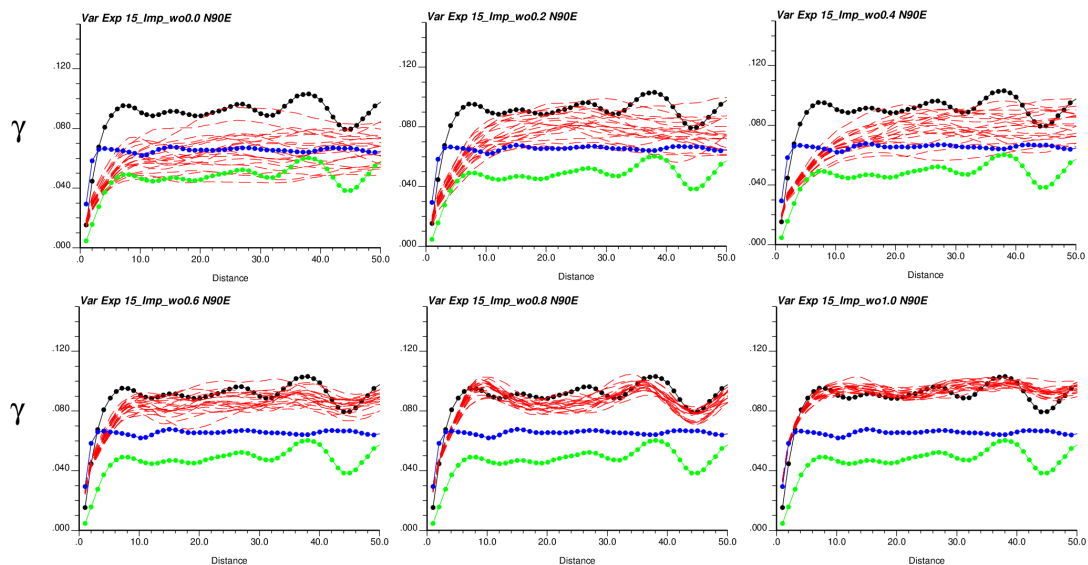
200 D.C. - 100% D.C. - Media local Inferida desde los DC



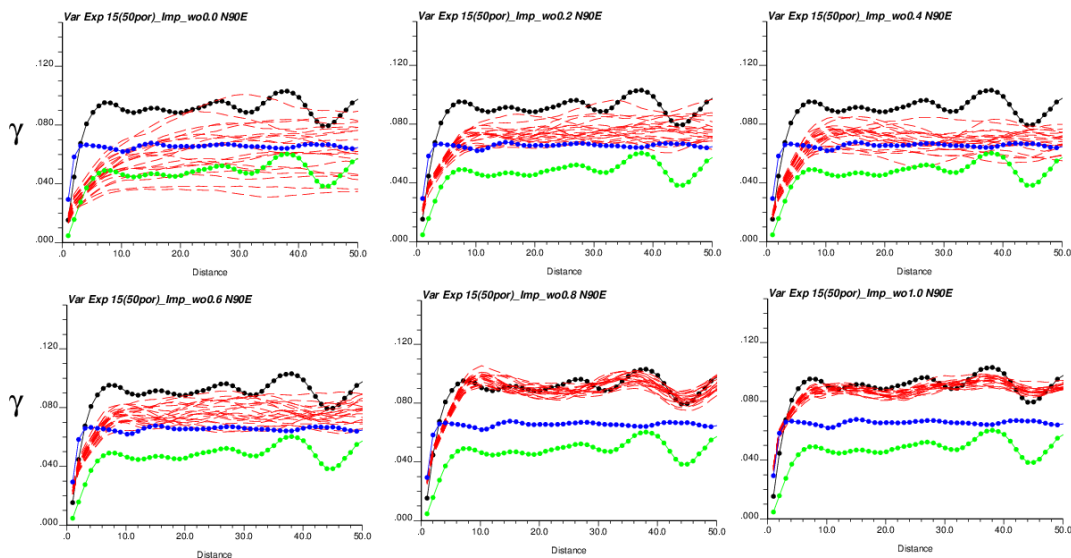
200 D.C. - 50% D.C. - Media local Inferida desde los DC



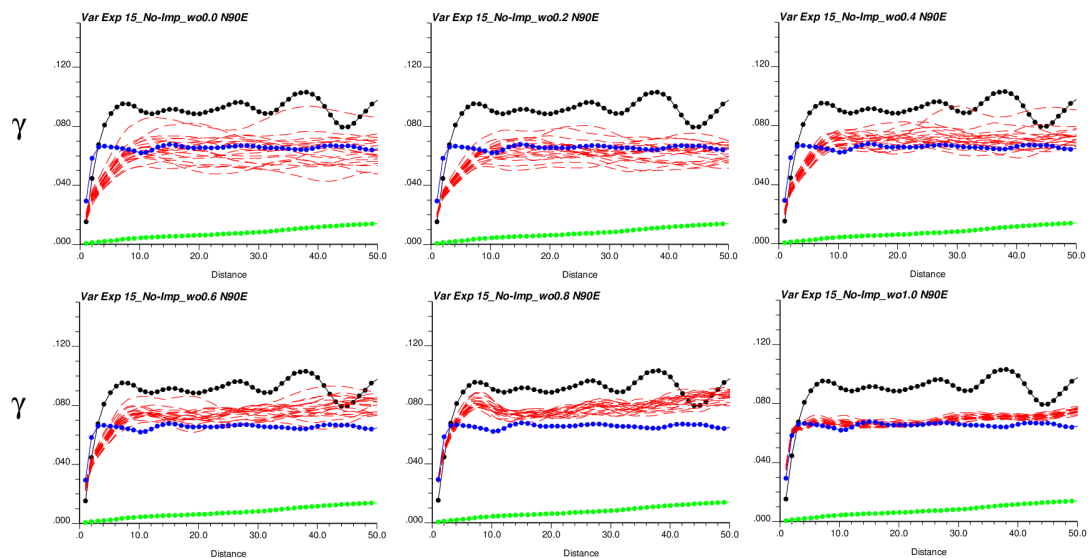
15 D.C. - 100% D.C. - Media local Impuesta desde GT



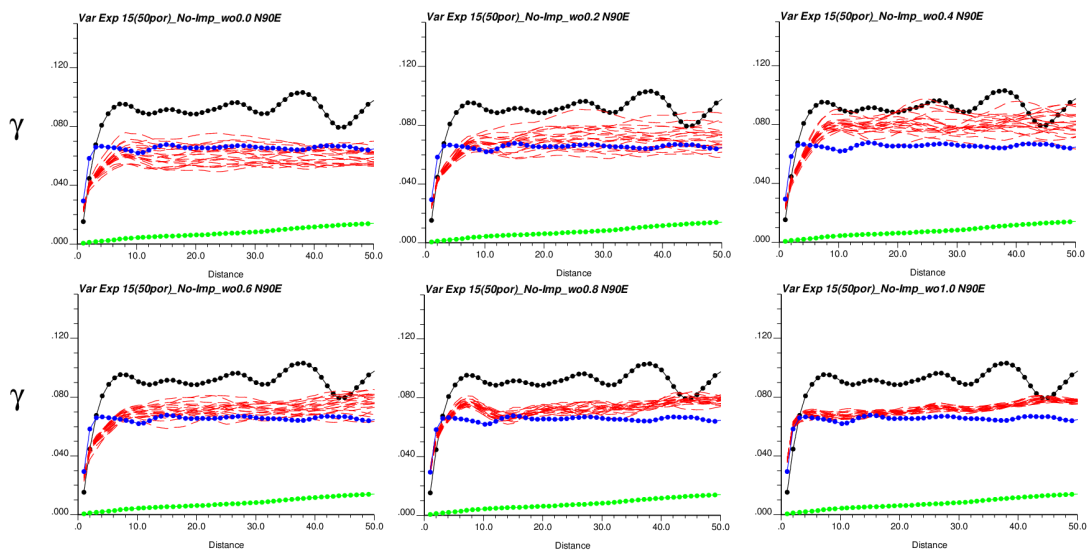
15 D.C. - 50 % D.C. - Media local Impuesta desde GT



15 D.C. - 100 % D.C. - Media local Inferida desde los DC

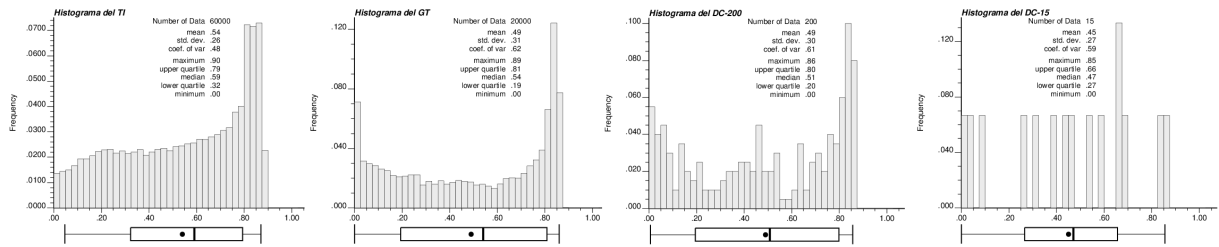


15D.C. - 50 % D.C. - Media local Inferida desde los DC



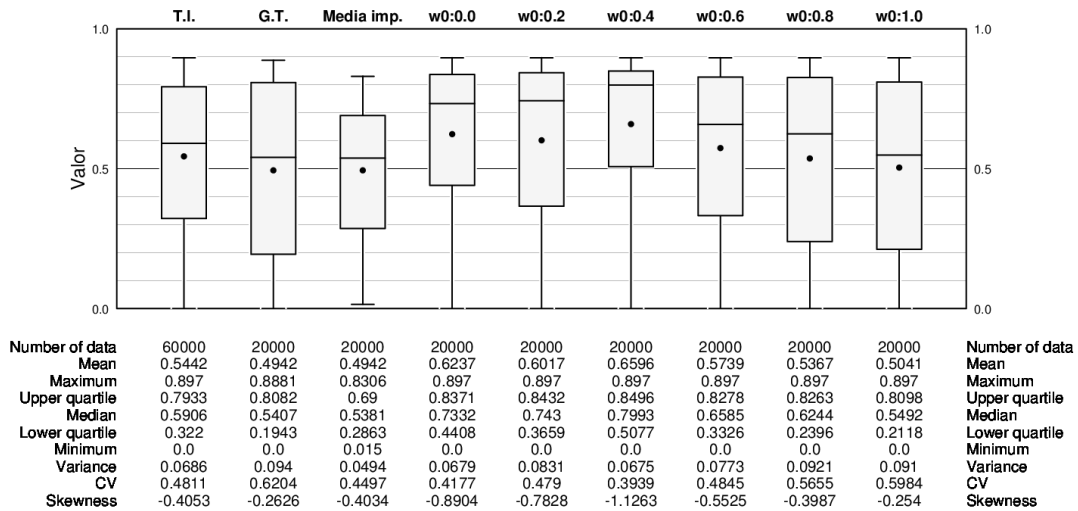
7.4. Anexo D

La presente sección presenta los histogramas de la Imagen de entrenamiento, el Ground truth, y las medias tanto impuesta como inferidas desde los datos condicionantes.

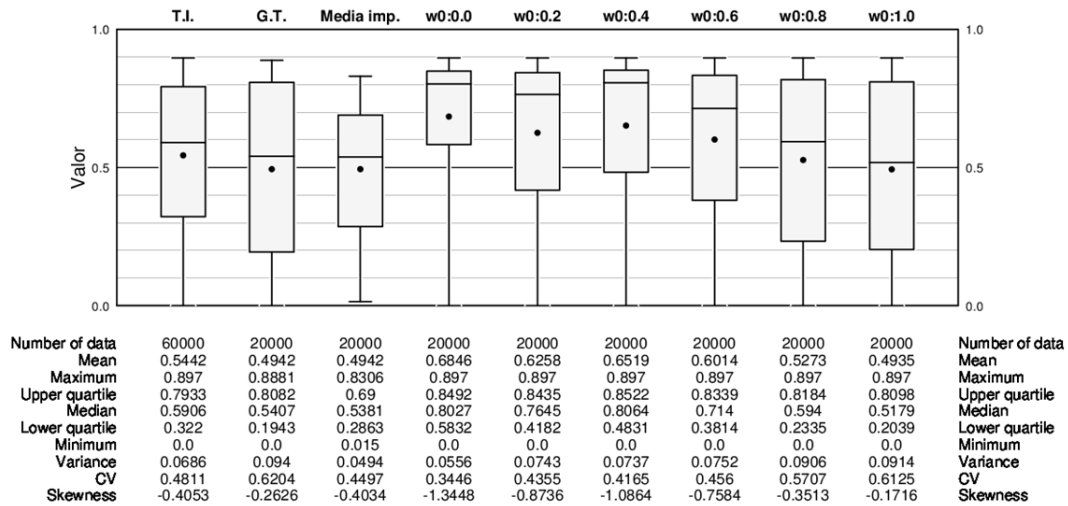


A continuación se presentan solo los diagramas de cajas (box plots) para cada una de las simulaciones. Se decide presentar solo 1 de las 100 realizaciones escogida de manera aleatoria. Cabe destacar que existen fluctuaciones en los histogramas entre una realización y otra, sin embargo, esta fluctuación es leve en comparación a las generadas al modificar el ponderador de influencia de la media local, w_0 . Por lo tanto, se presentan los diagramas de caja para todos los w_0 y se incluyen la imagen de entrenamiento, el ground truth y la media local, según el caso, como referencias.

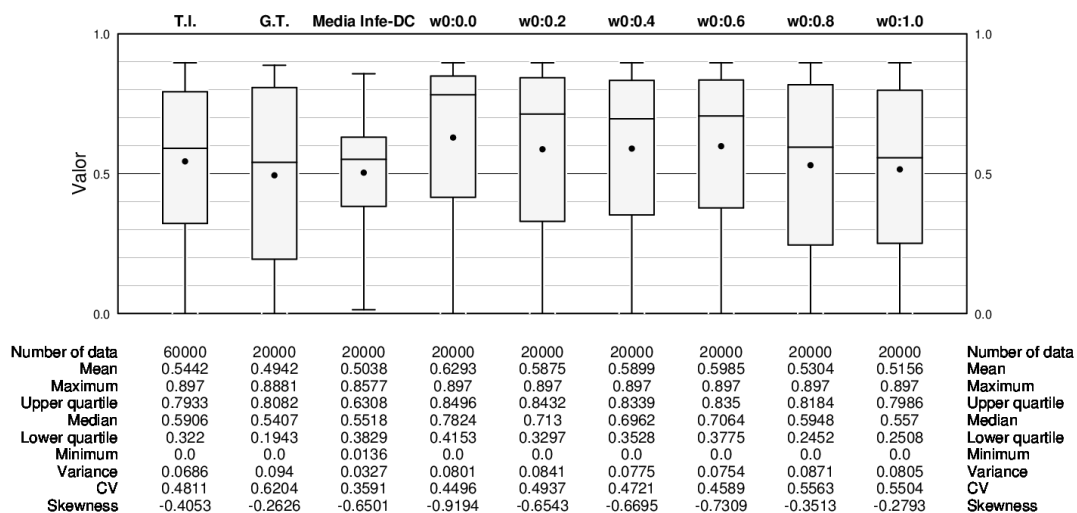
200 D.C. - 100% D.C. - Media local Impuesta desde GT



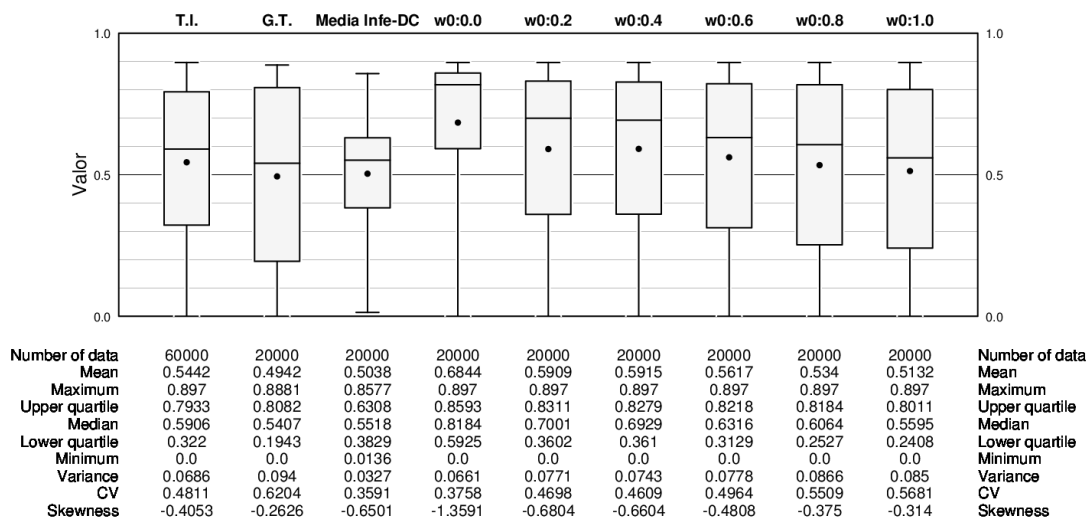
200 D.C. - 50% D.C. - Media local Impuesta desde GT



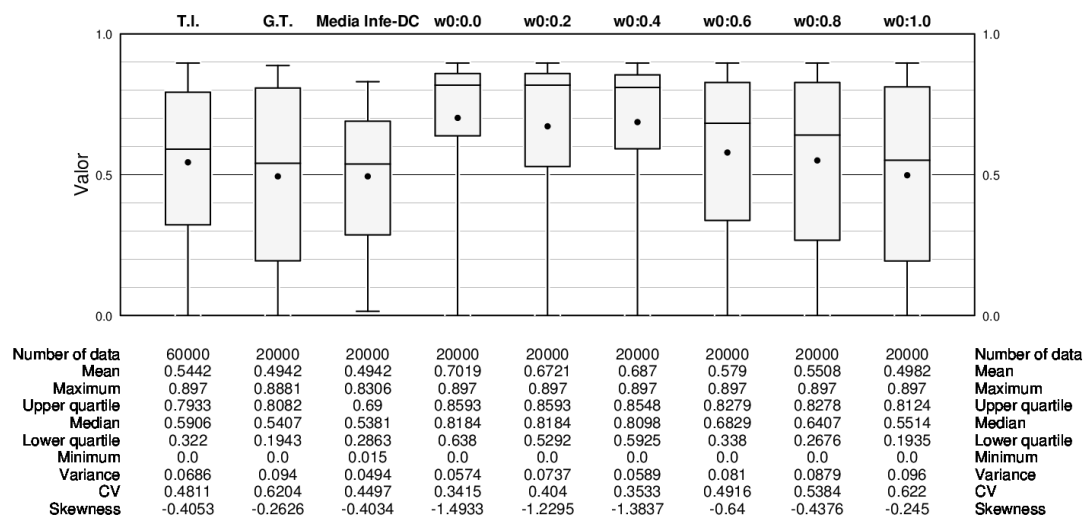
200 D.C. - 100 % D.C. - Media local Inferida desde los DC



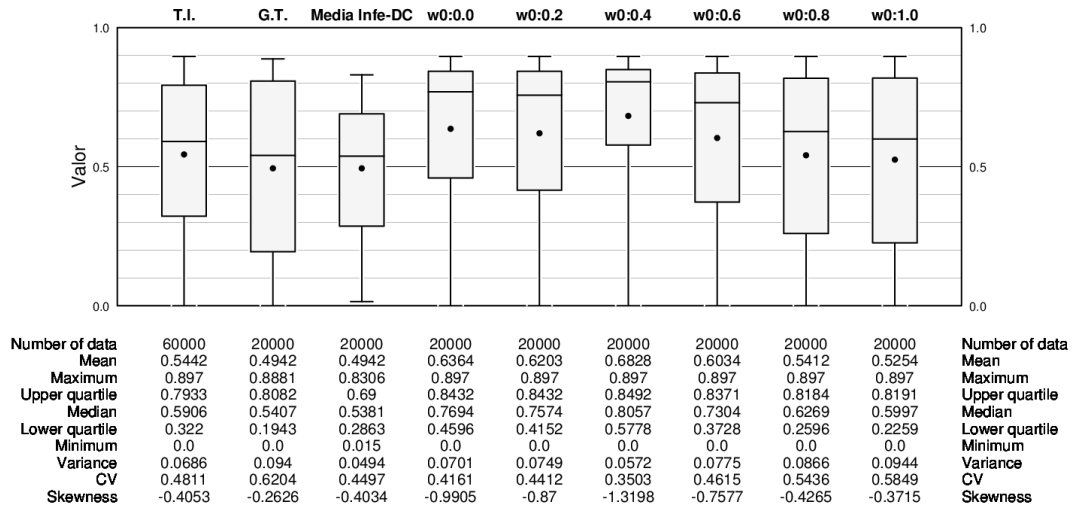
200 D.C. - 50 % D.C. - Media local Inferida desde los DC



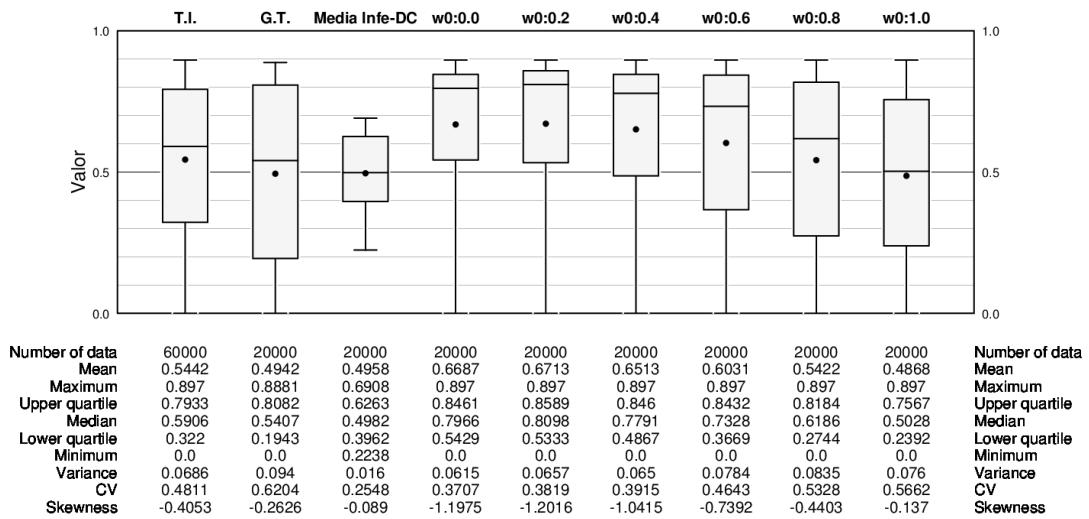
15 D.C. - 100 % D.C. - Media local Impuesta desde GT



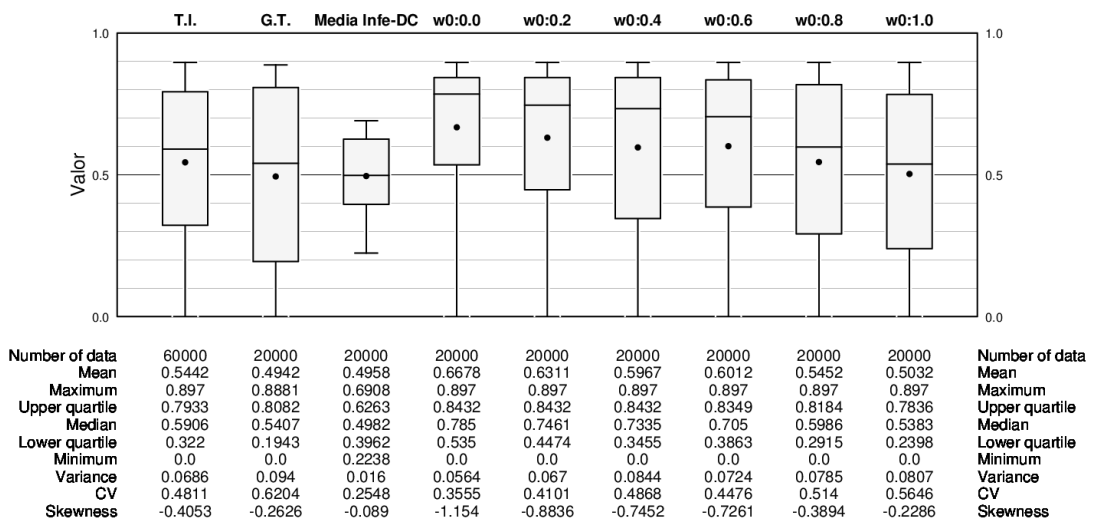
15 D.C. - 50 % D.C. - Media local Impuesta desde GT



15 D.C. - 100 % D.C. - Media local Inferida desde los DC



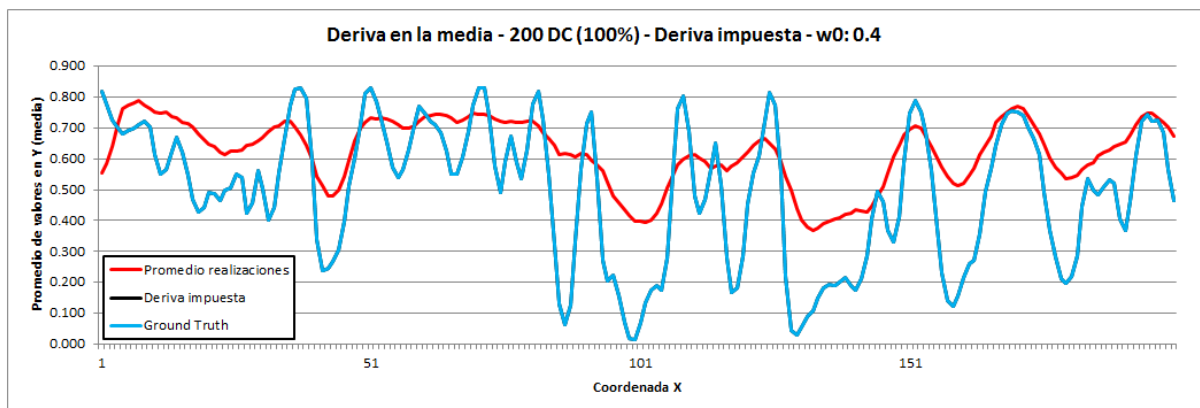
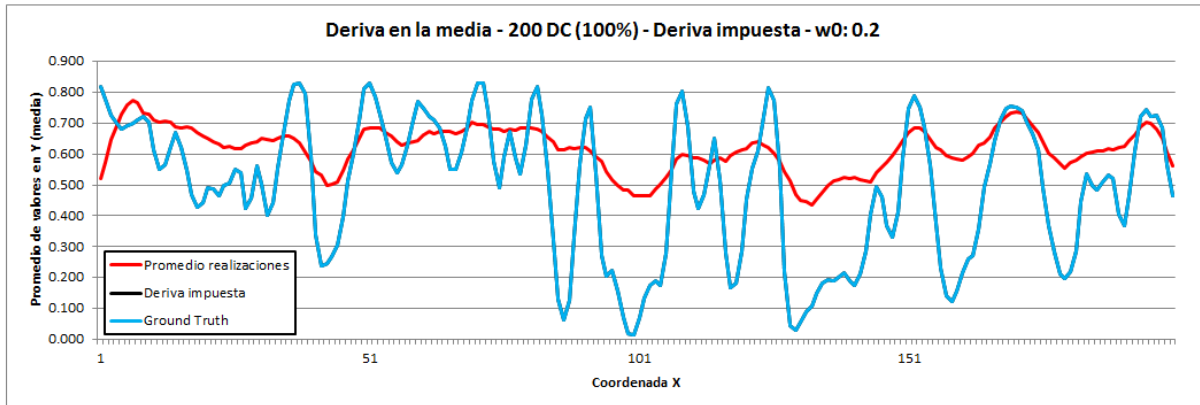
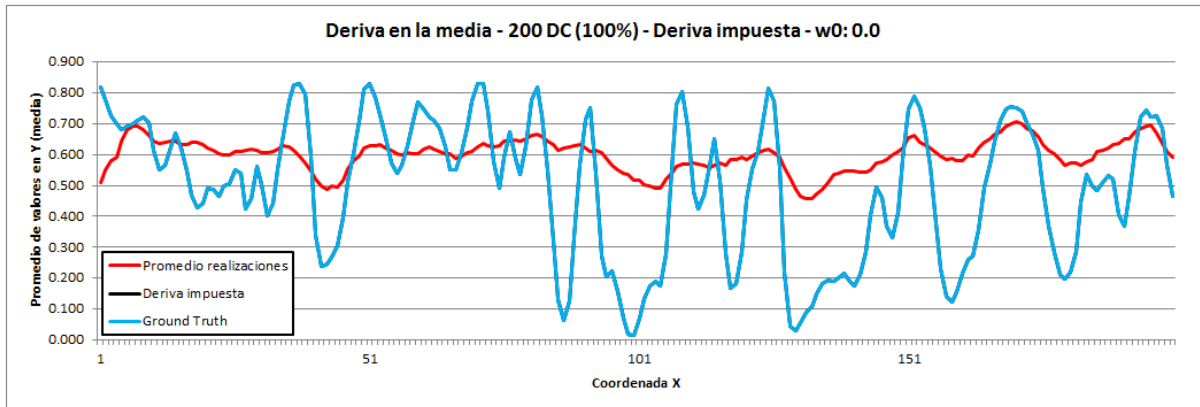
15D.C. - 50 % D.C. - Media local Inferida desde los DC

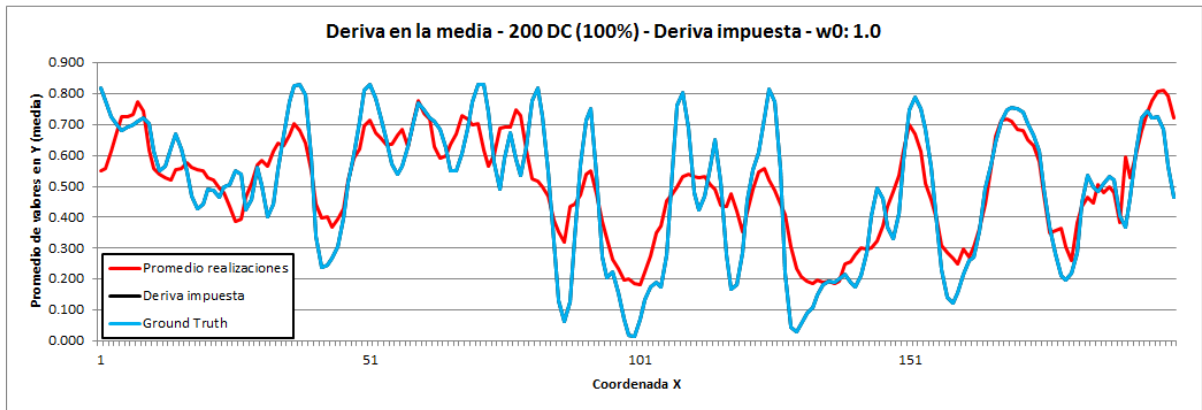
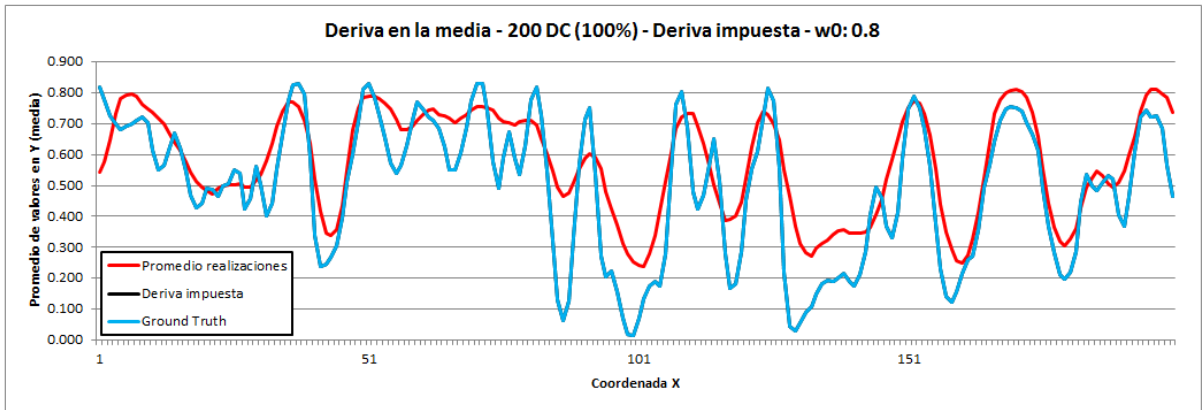
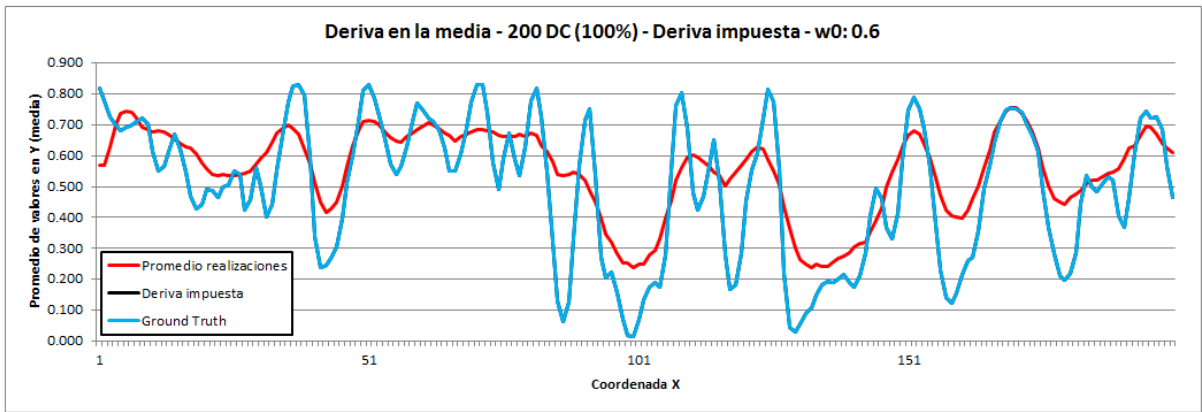


7.5. Anexo E

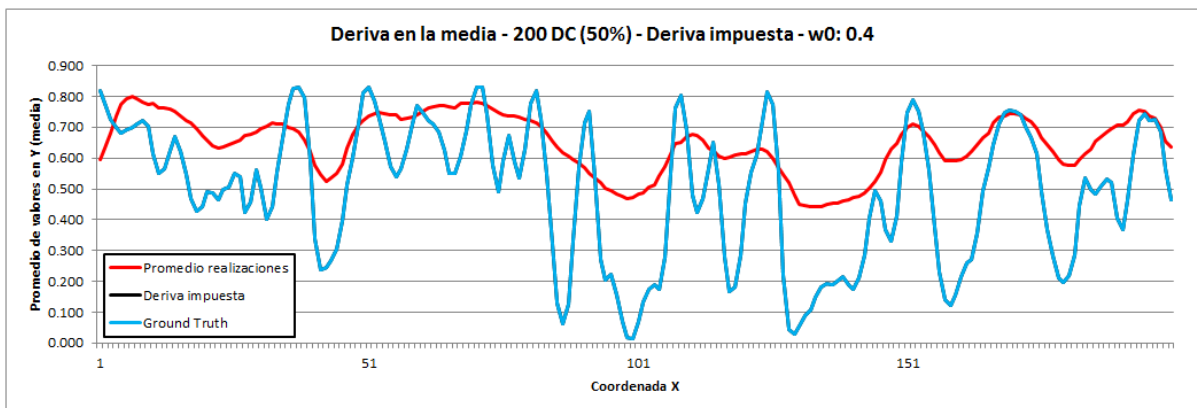
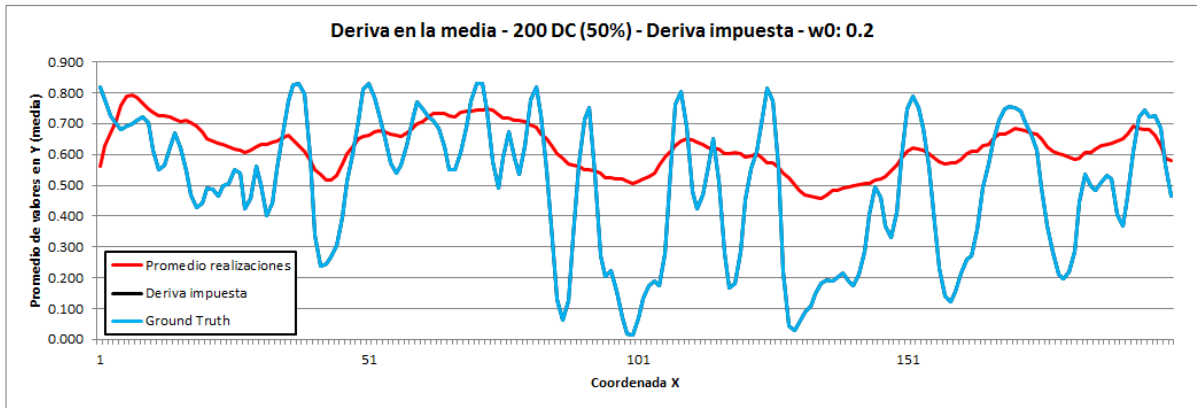
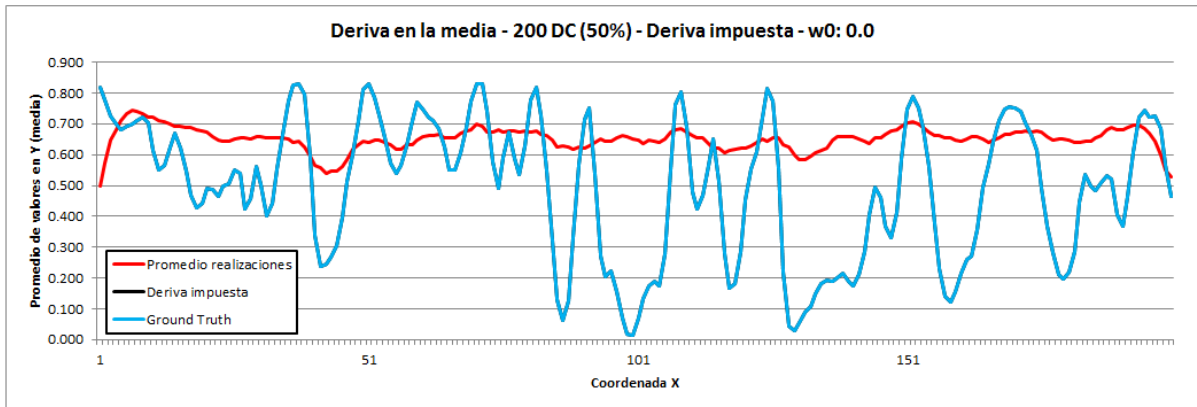
El presente anexo contiene las derivas en la media local a lo largo del eje X para cada simulación del Caso de estudio - Roca sedimentaria. La deriva en la media para el Ground Truth viene en **azul**, la deriva interpretada en **negro** (la impuesta coincide con la deriva del Ground Truth por formulación inicial), y en **rojo** la deriva de 20 realizaciones.

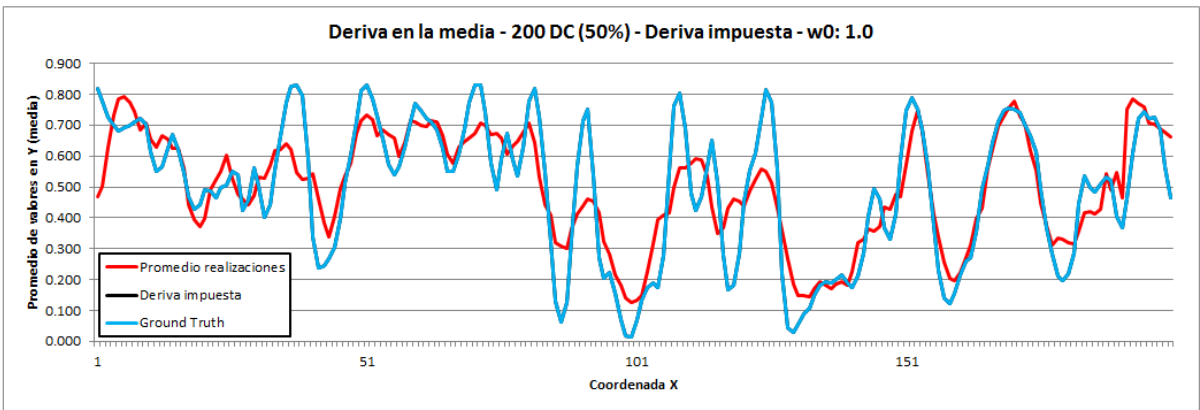
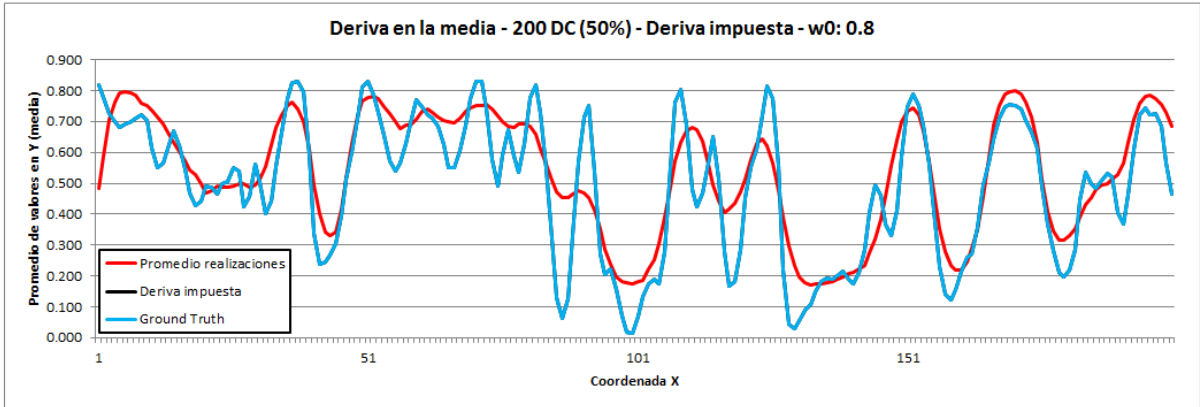
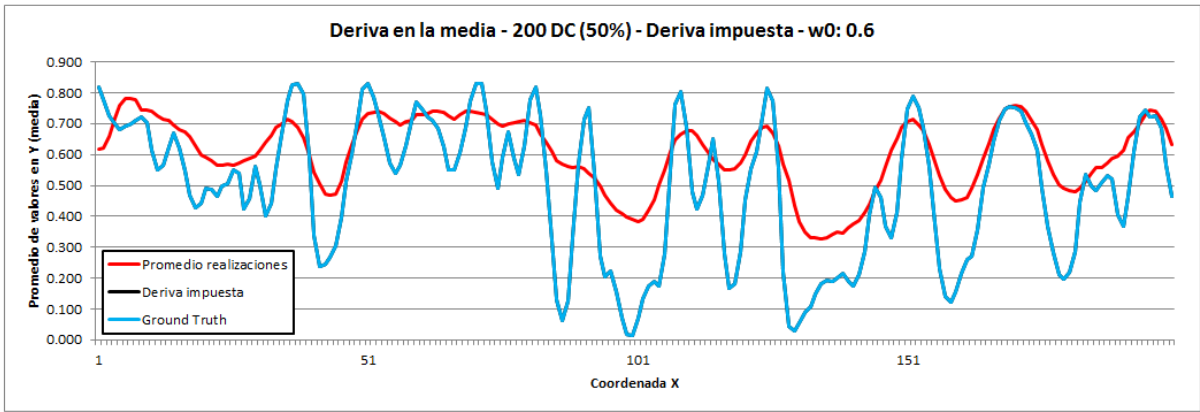
200 D.C. - 100 % D.C. - Media local Impuesta desde GT



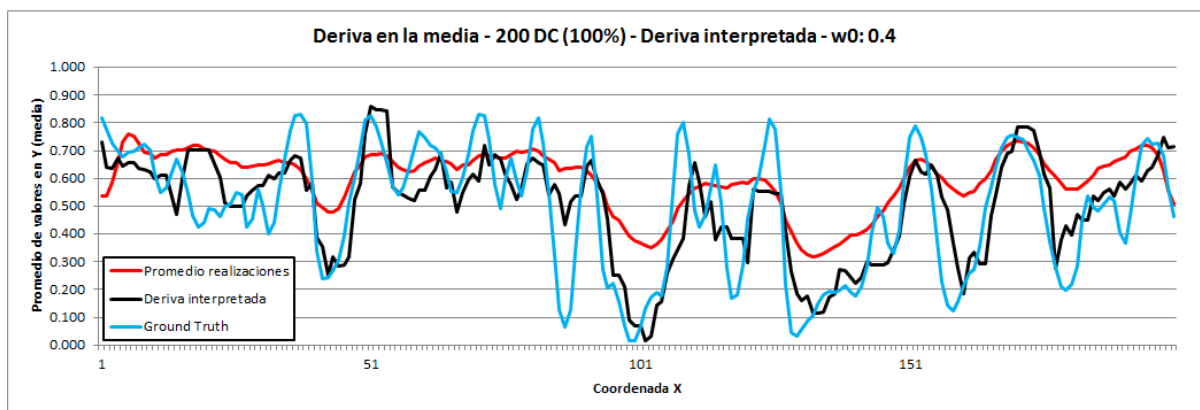
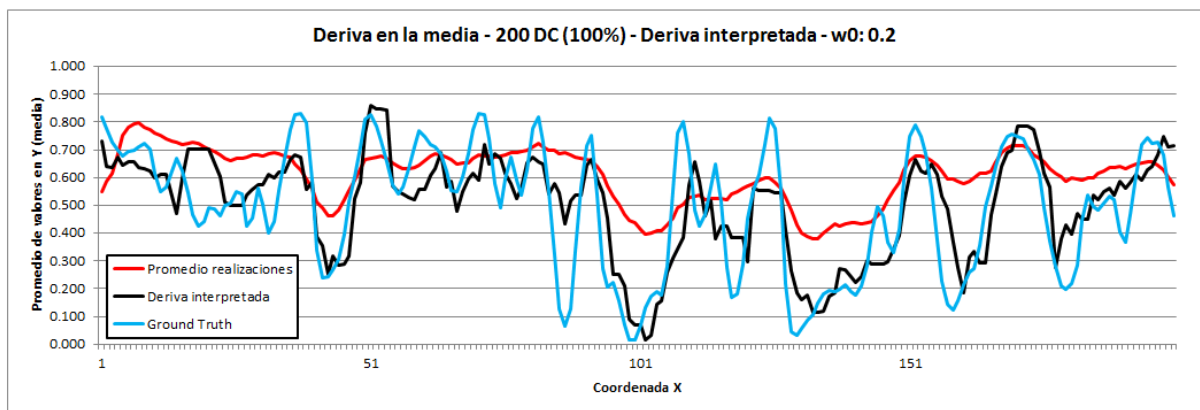
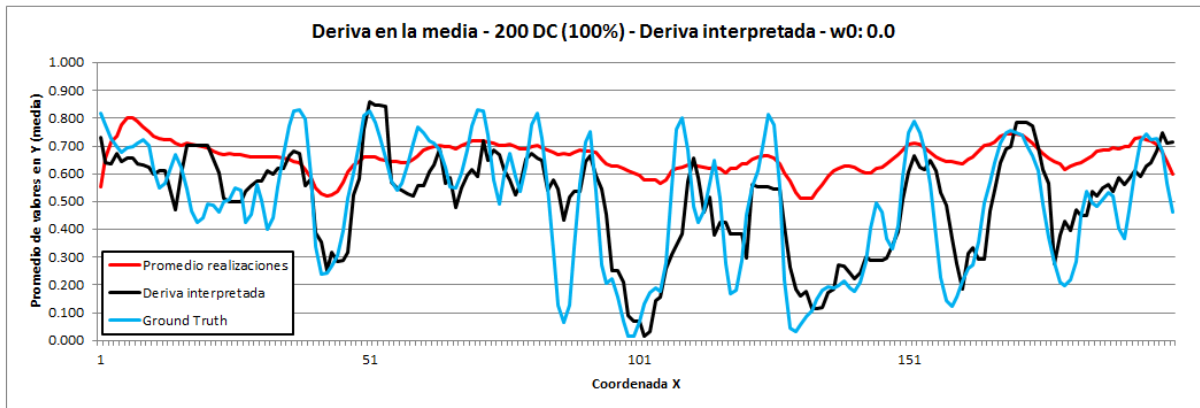


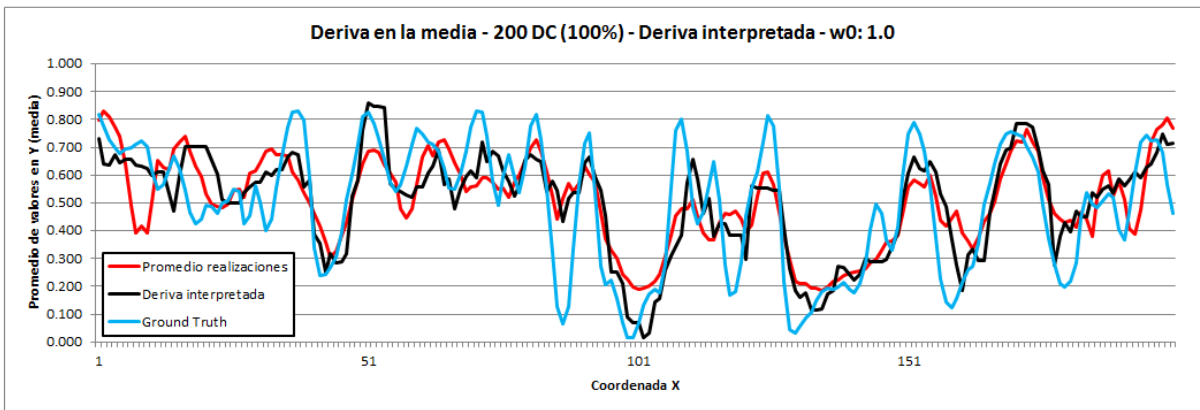
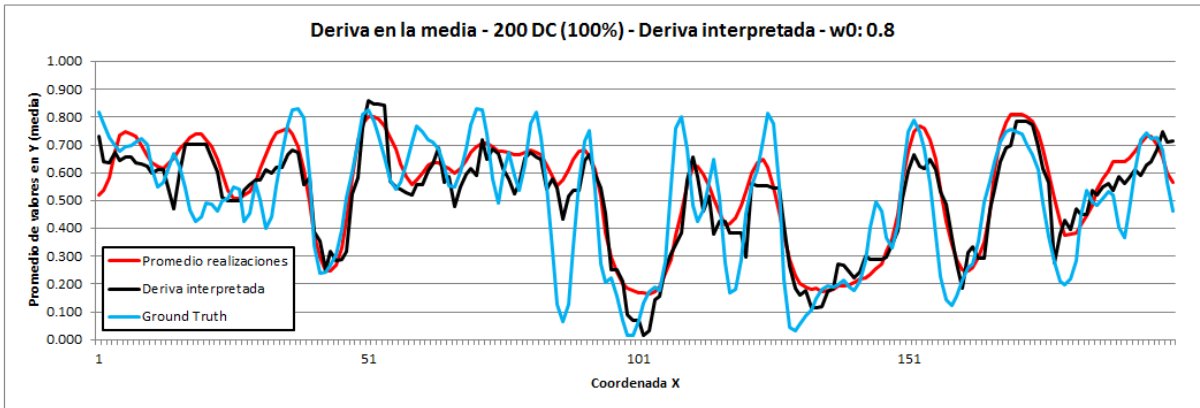
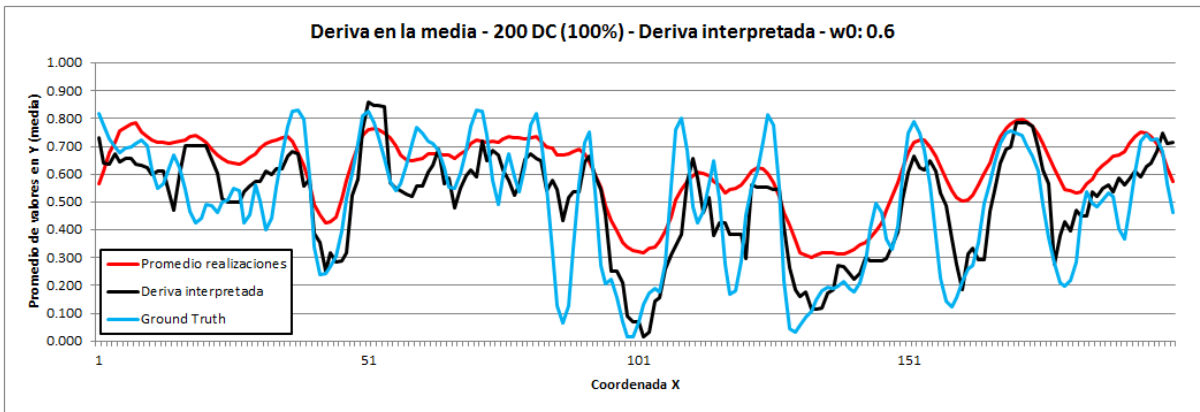
200 D.C. - 50 % D.C. - Media local Impuesta desde GT



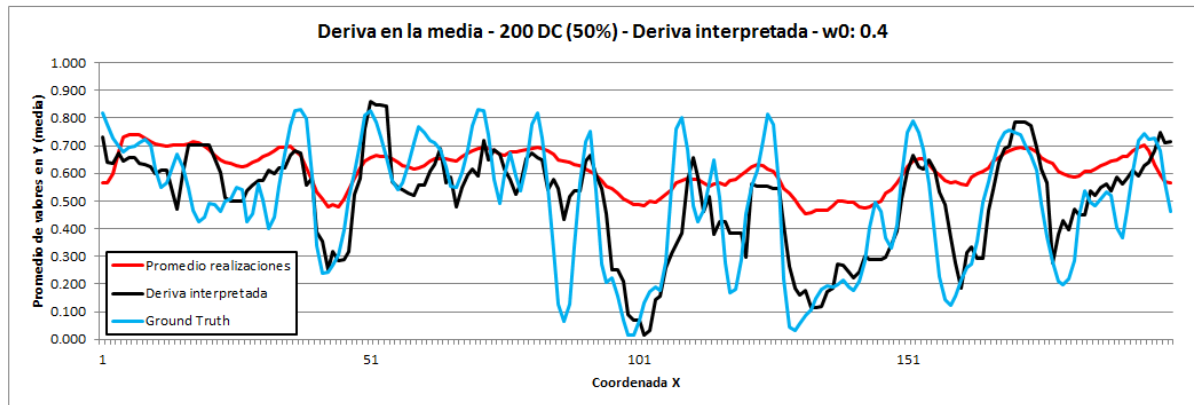
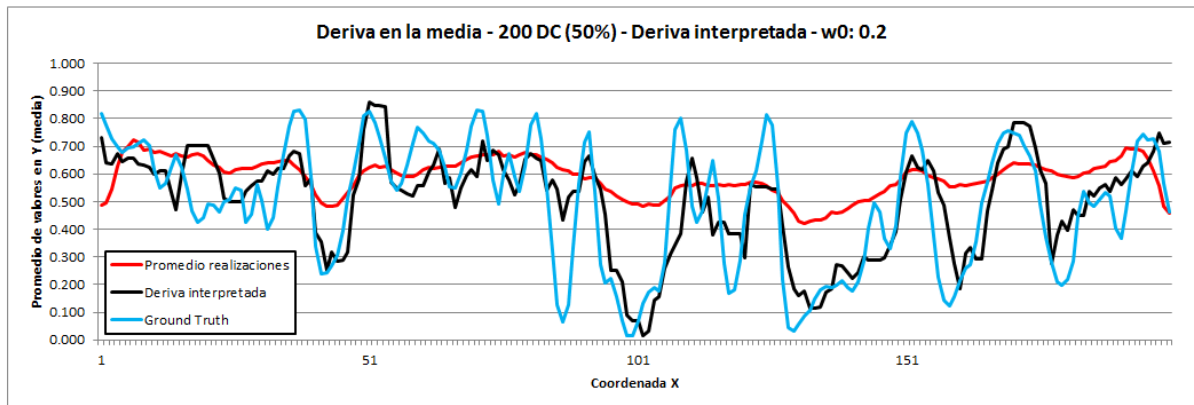
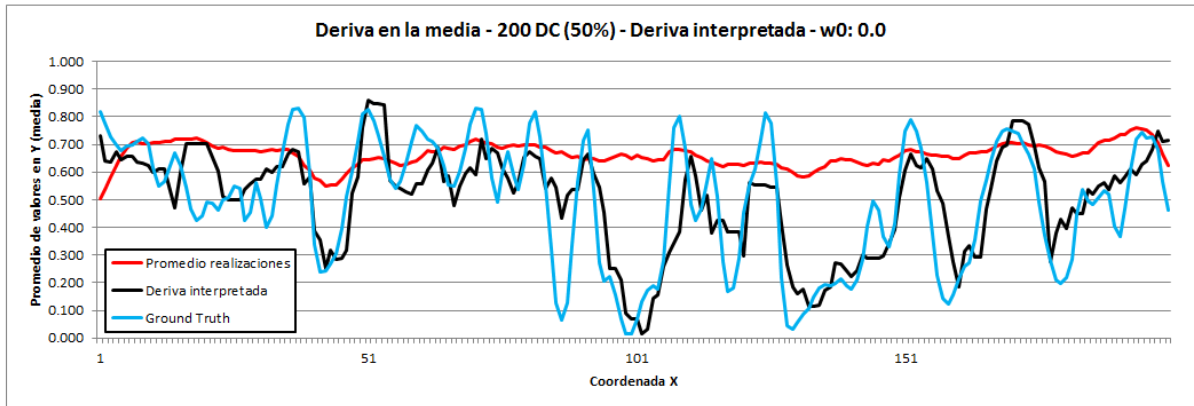


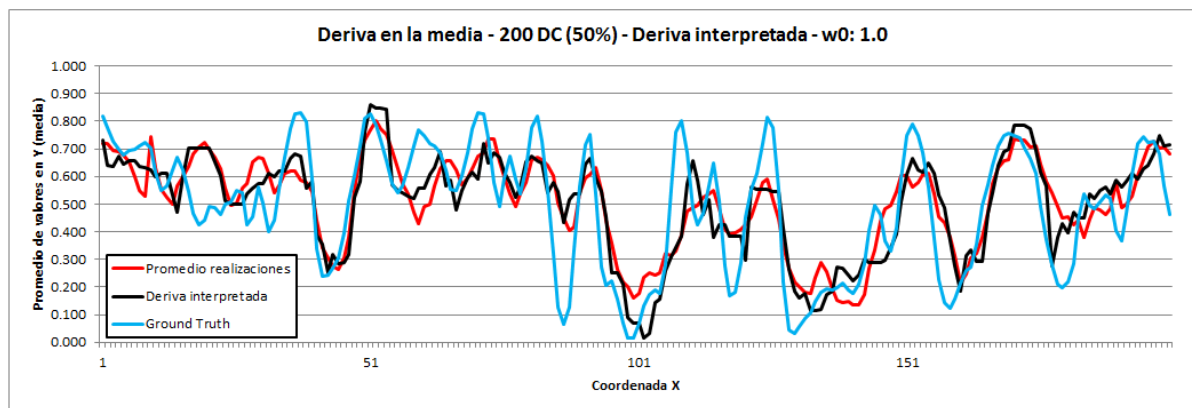
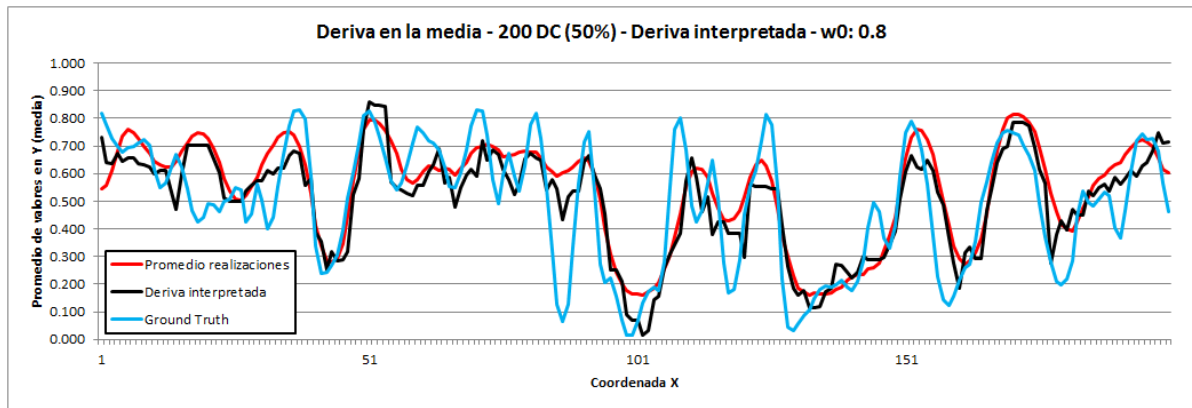
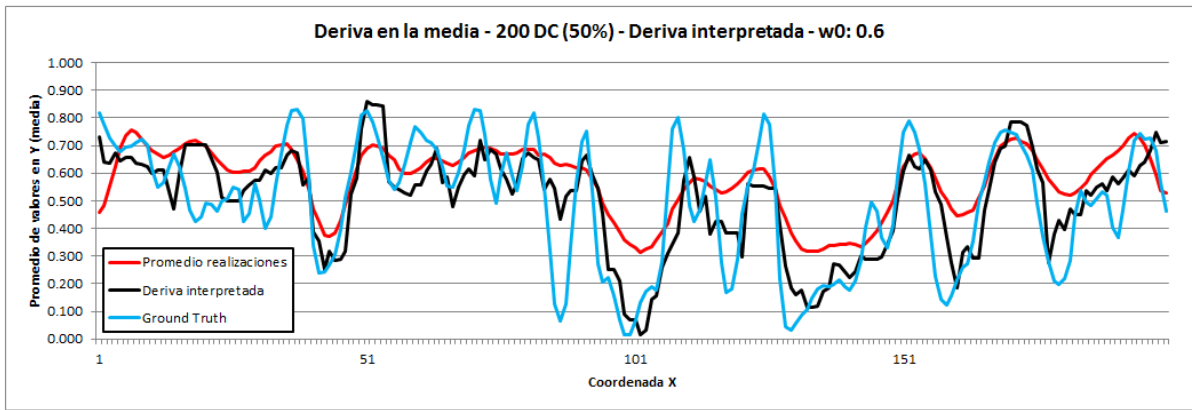
200 D.C. - 100% D.C. - Media local Inferida desde los DC



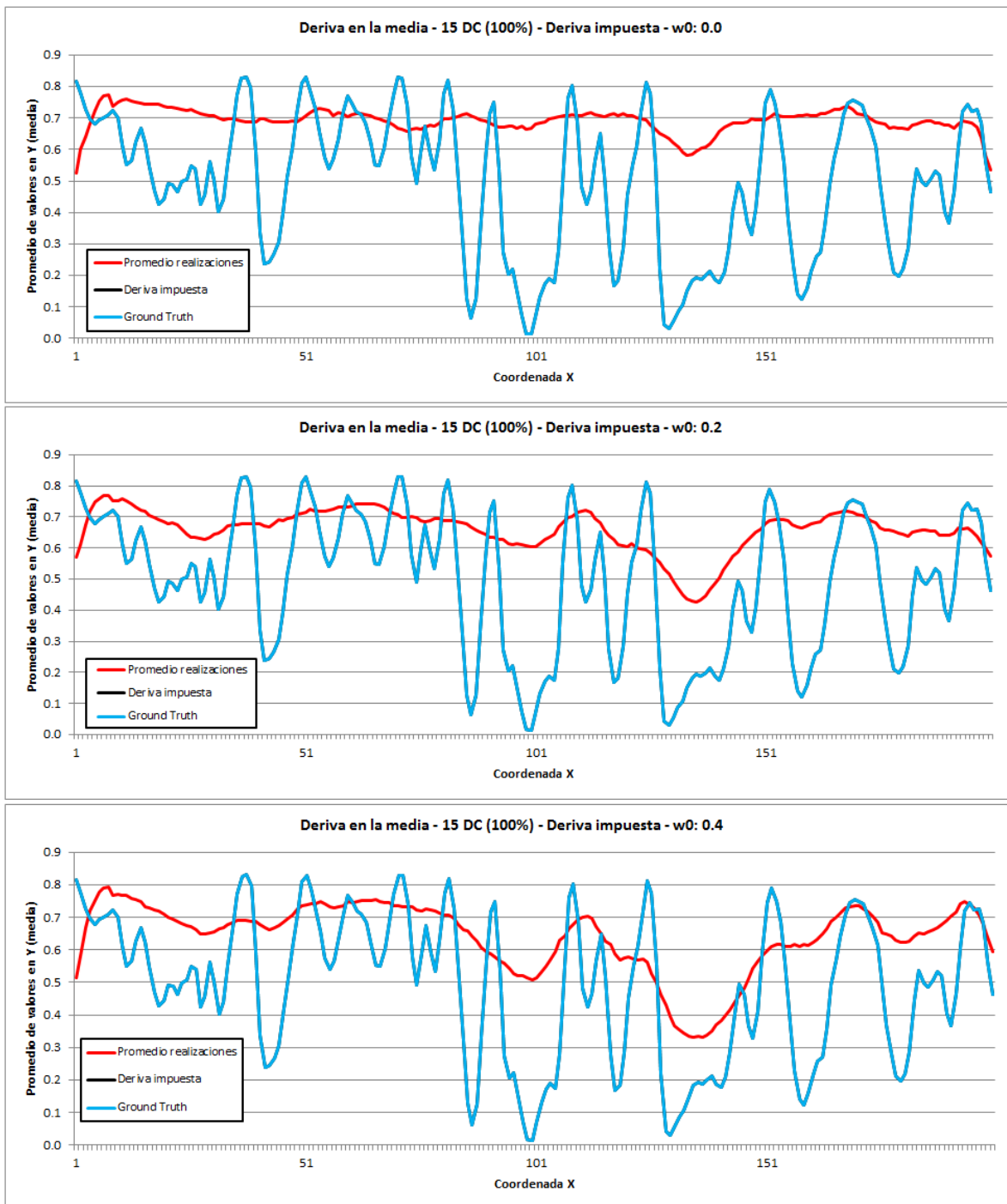


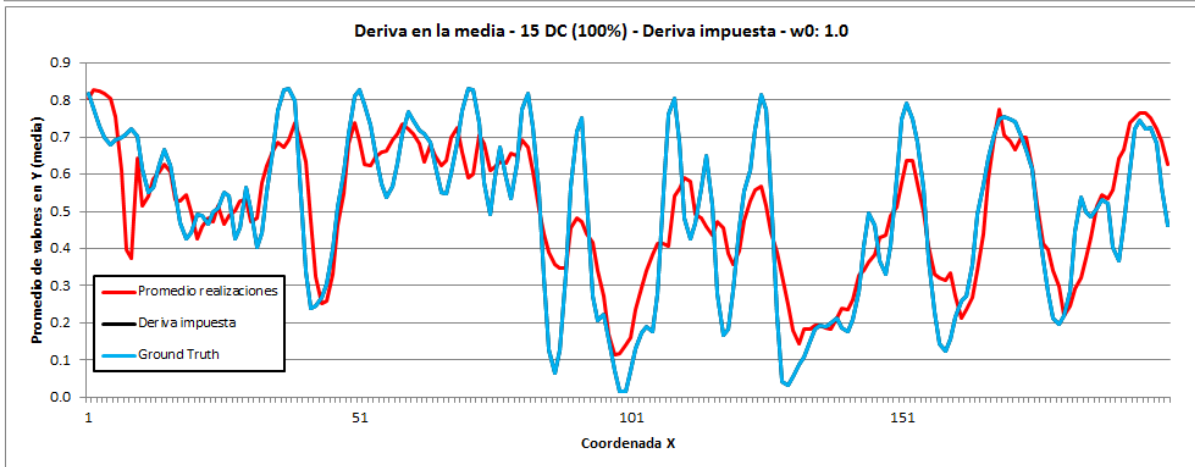
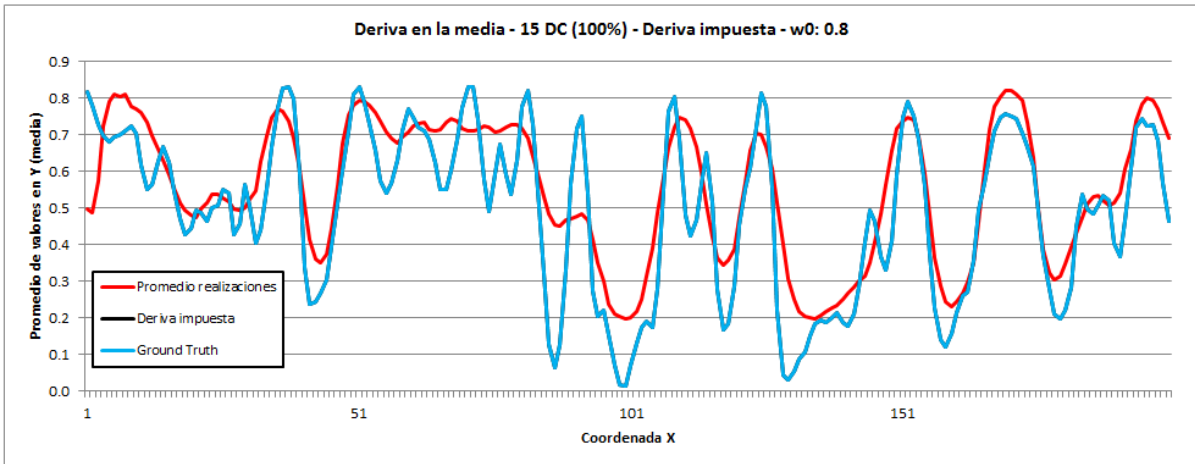
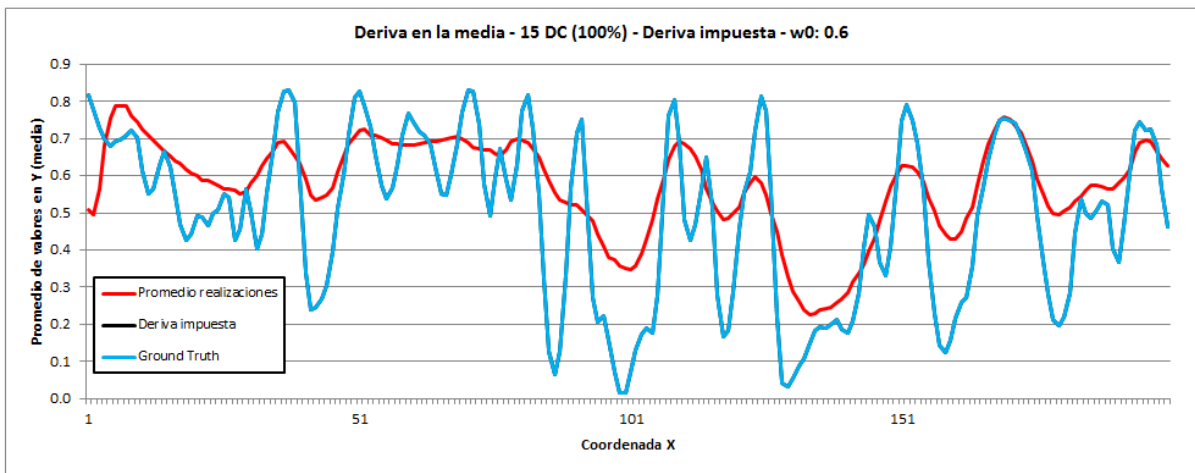
200 D.C. - 50 % D.C. - Media local Inferida desde los DC



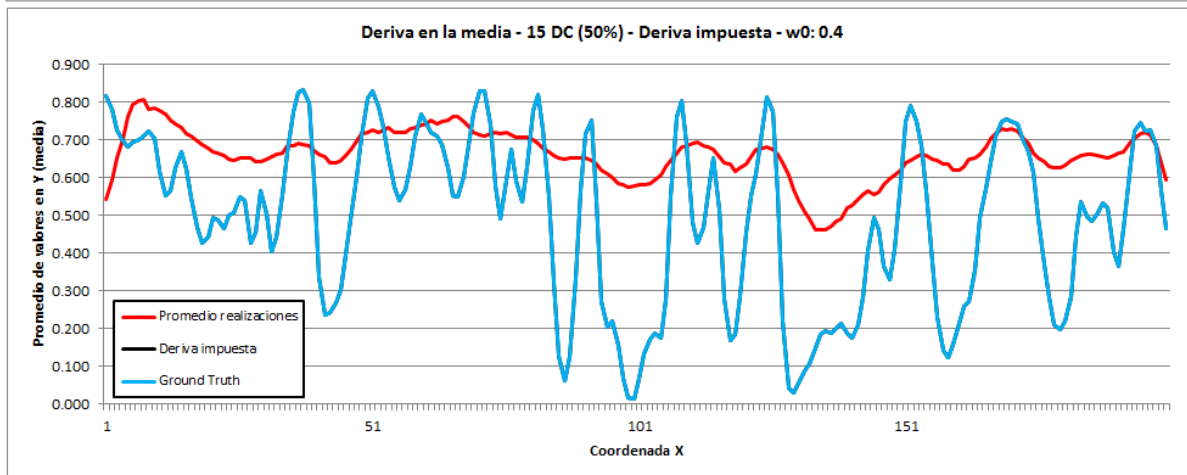
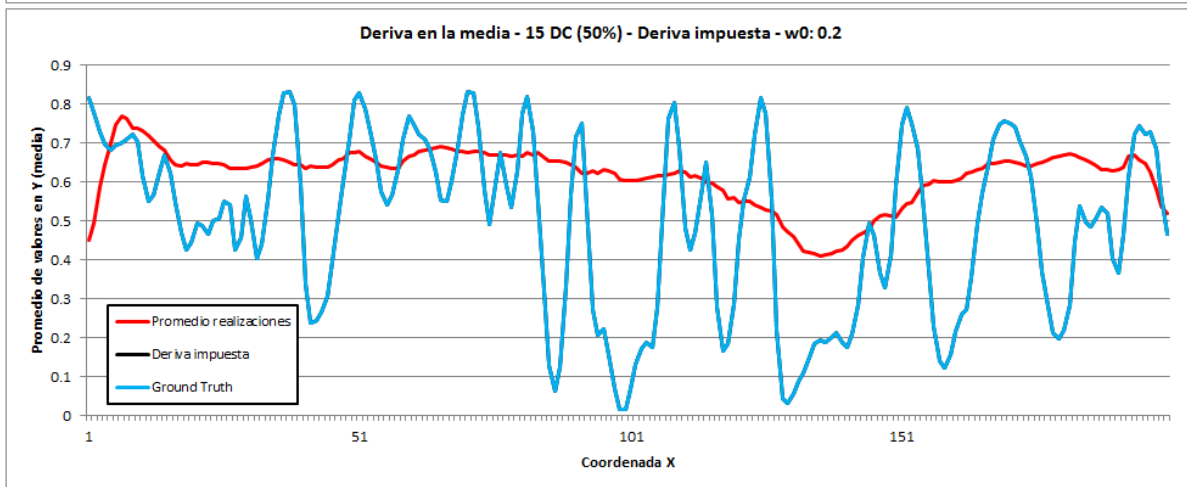
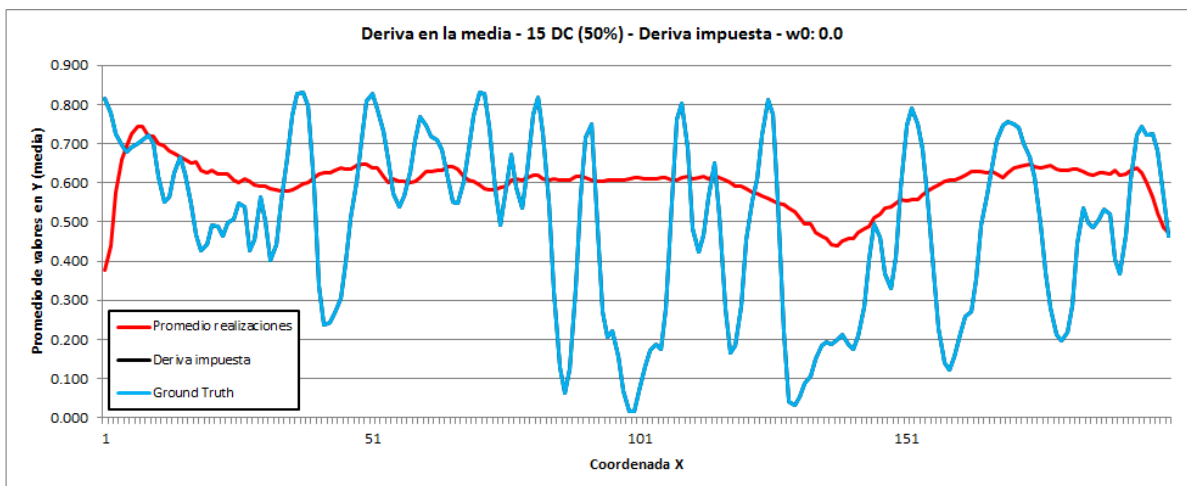


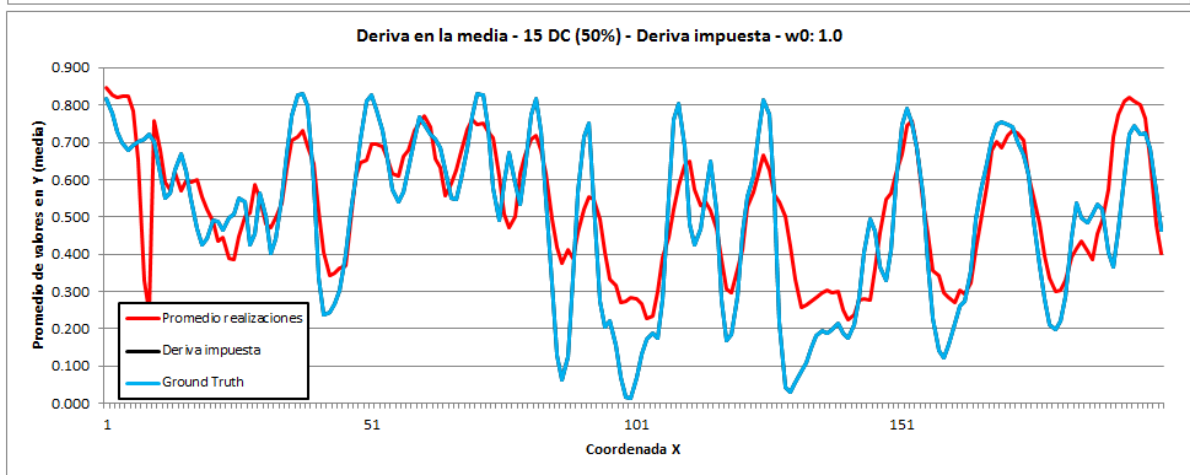
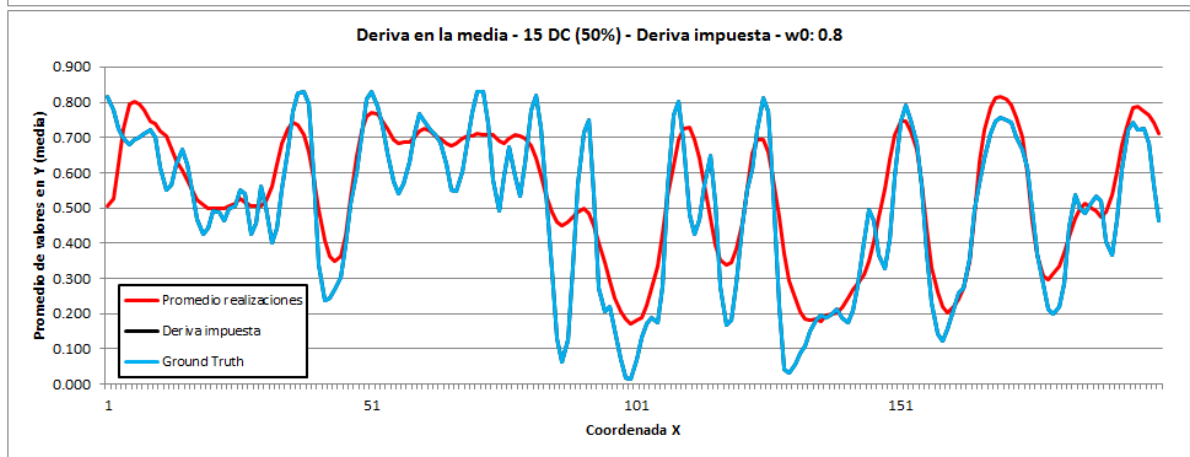
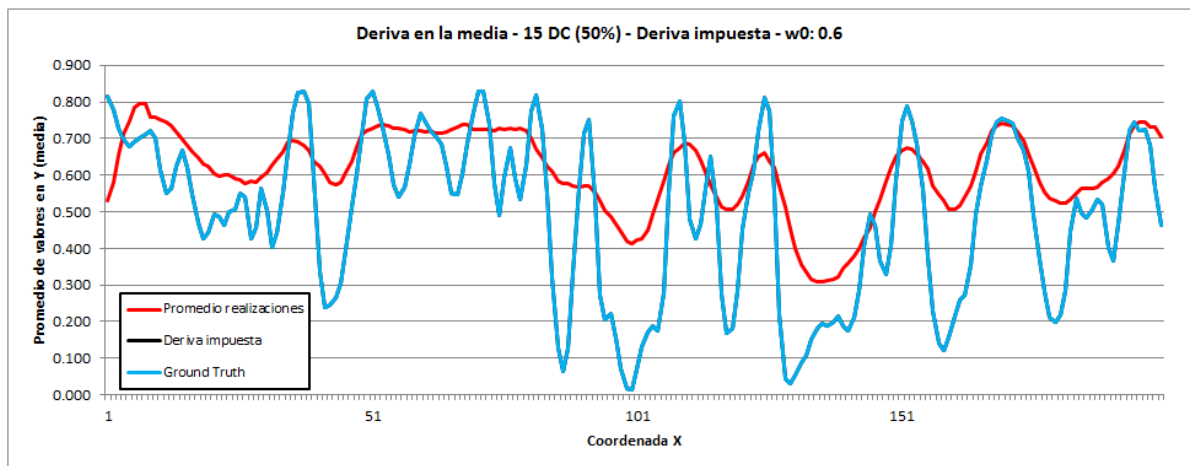
15 D.C. - 100 % D.C. - Media local Impuesta desde GT

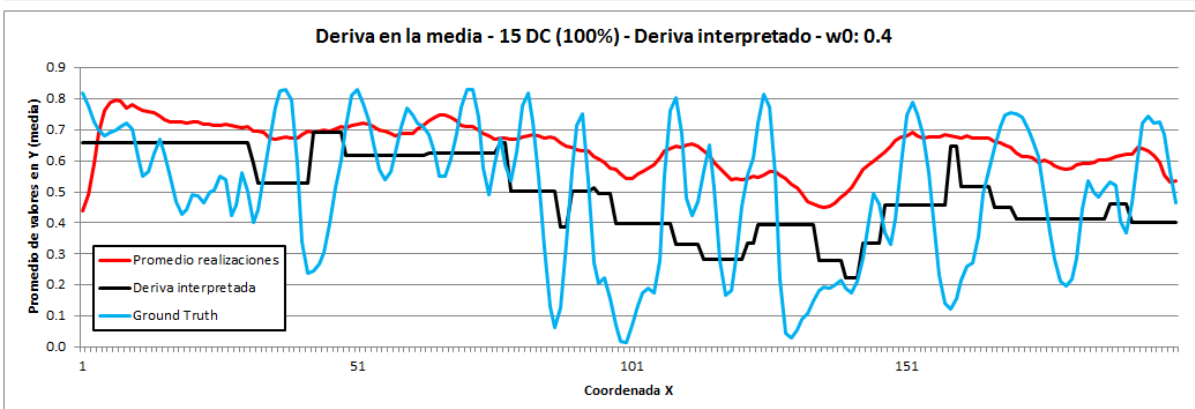
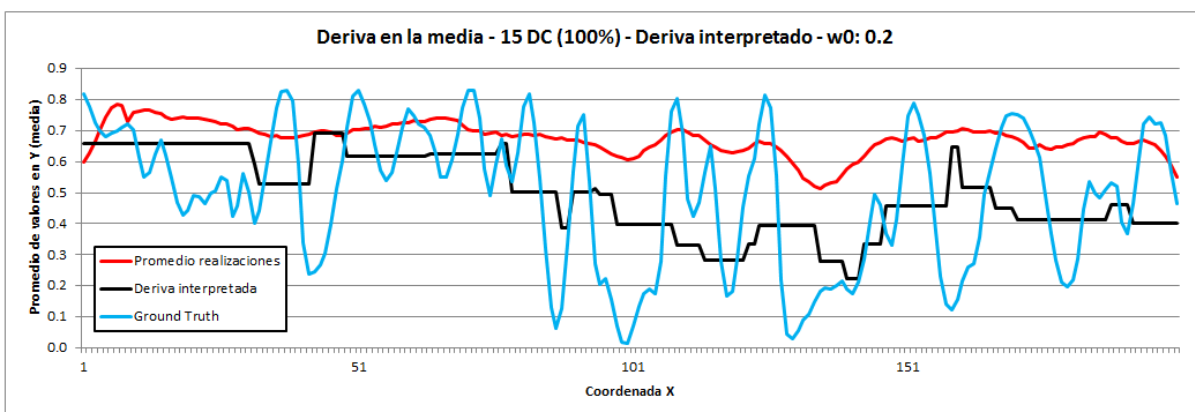
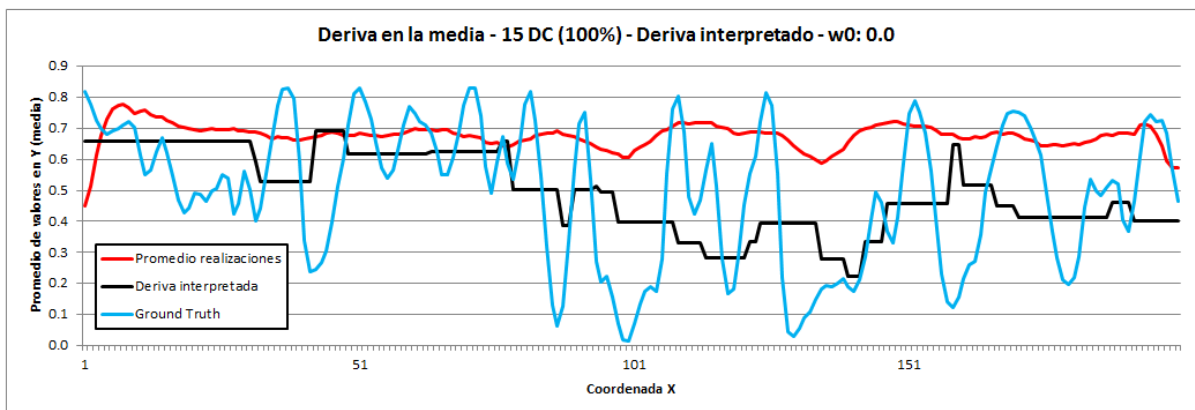


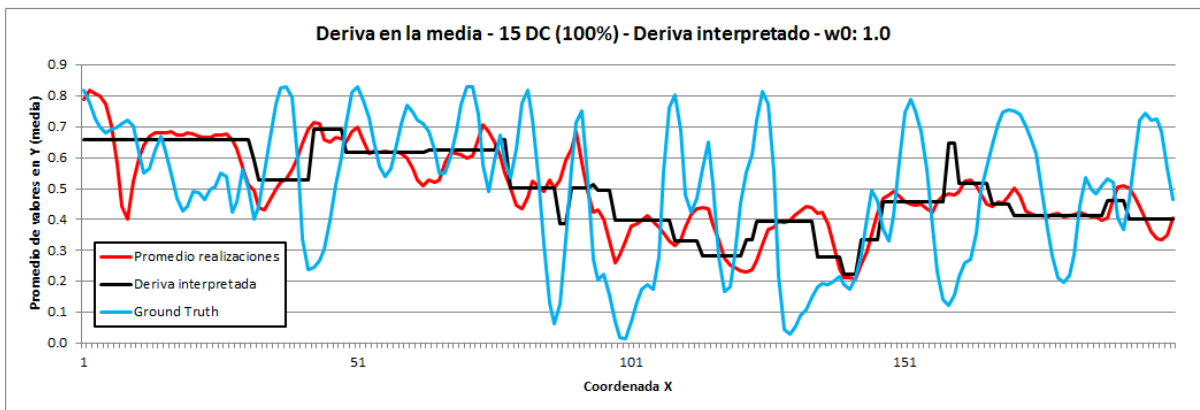
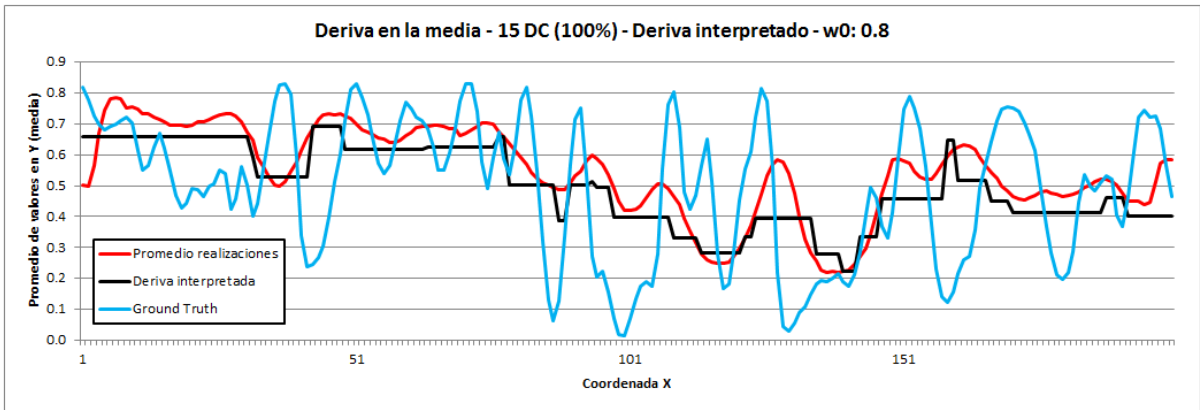
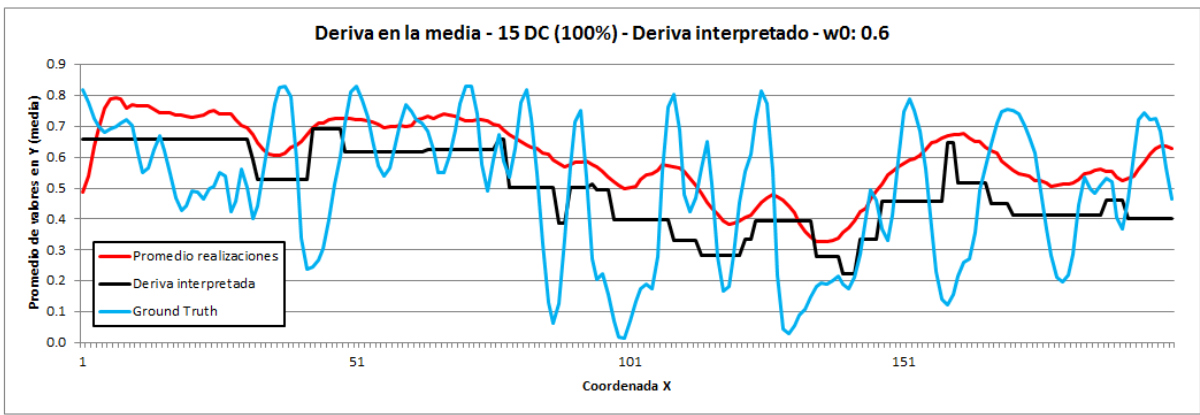


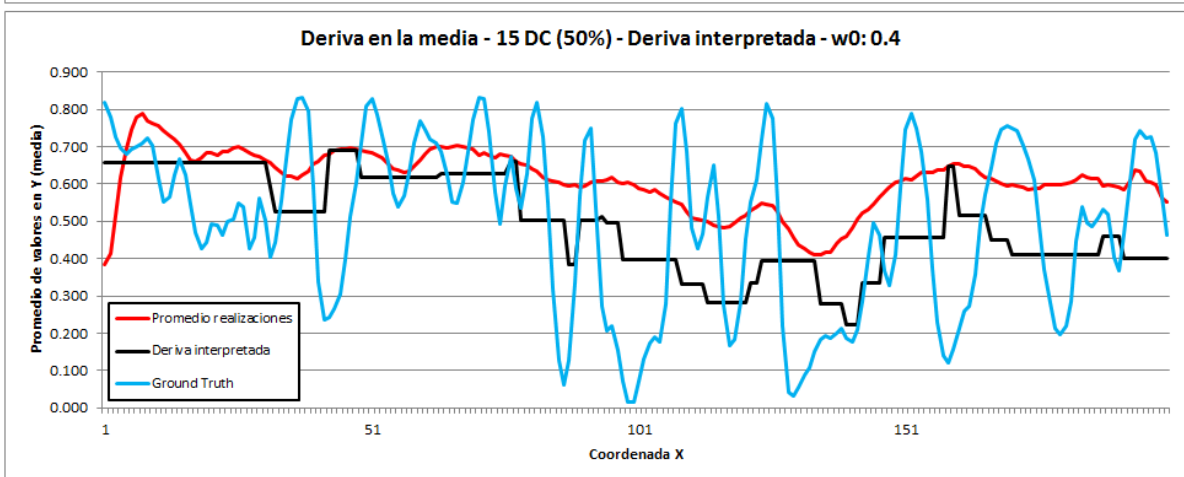
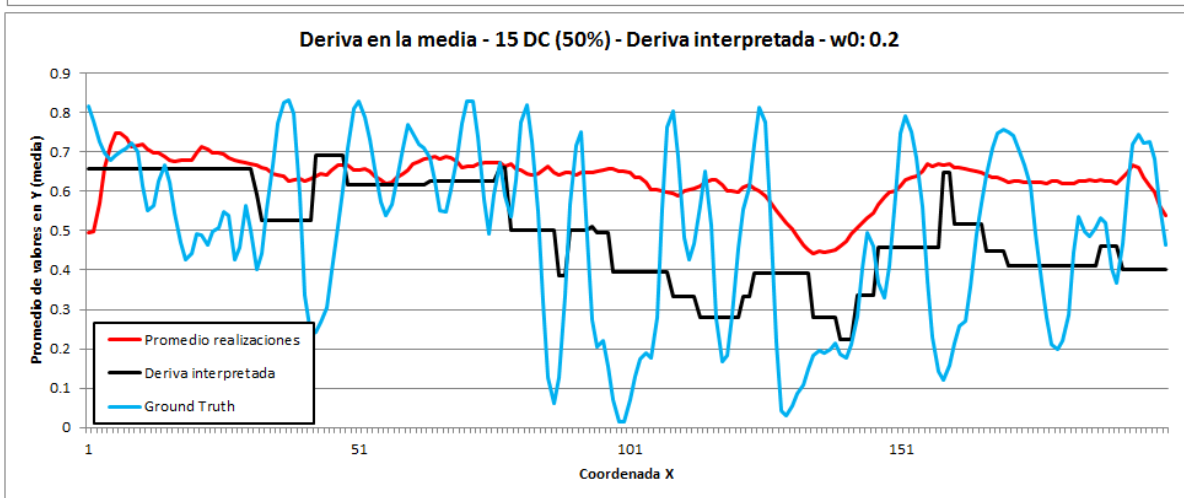
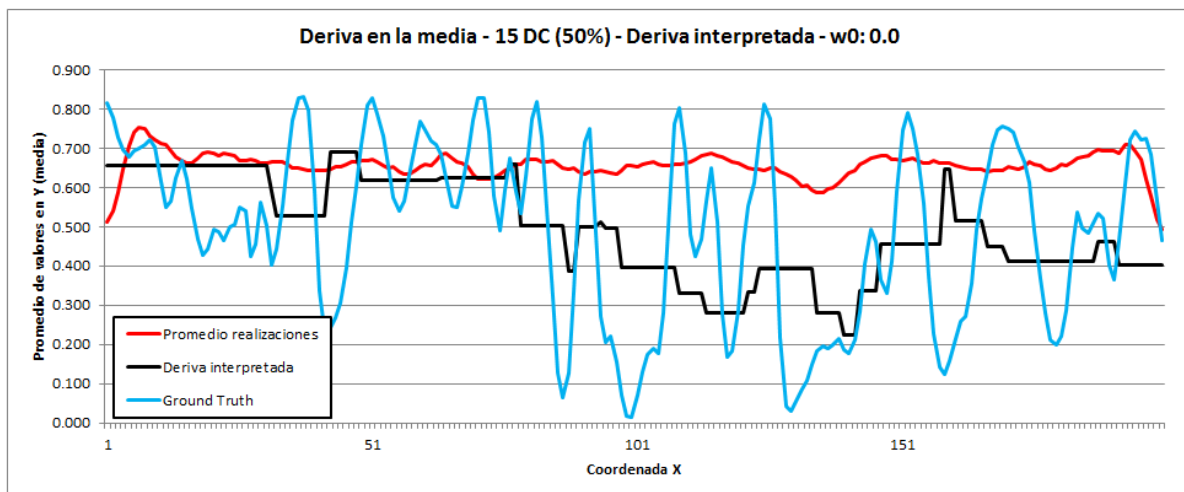
15 D.C. - 50 % D.C. - Media local Impuesta desde GT

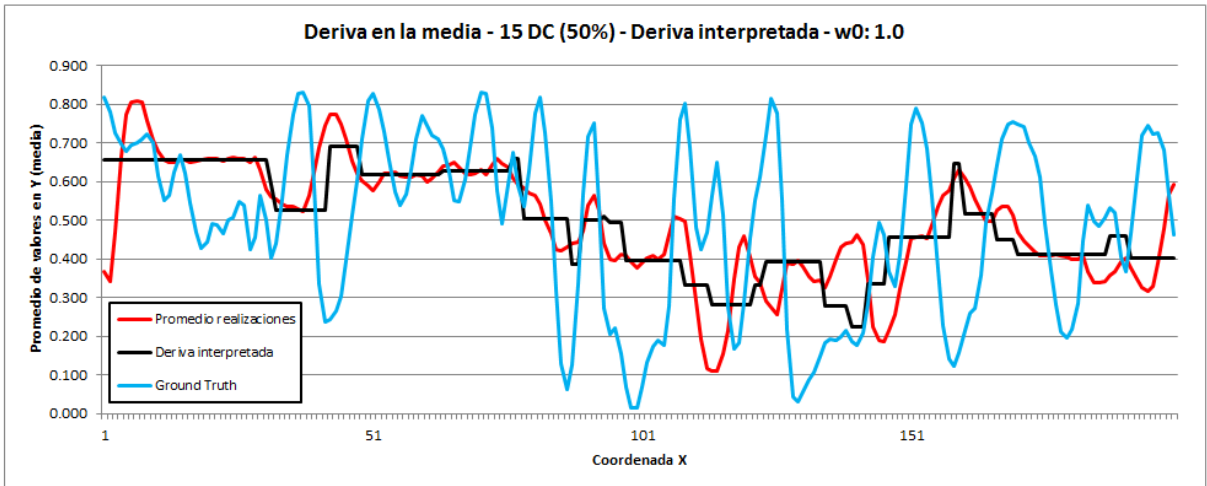
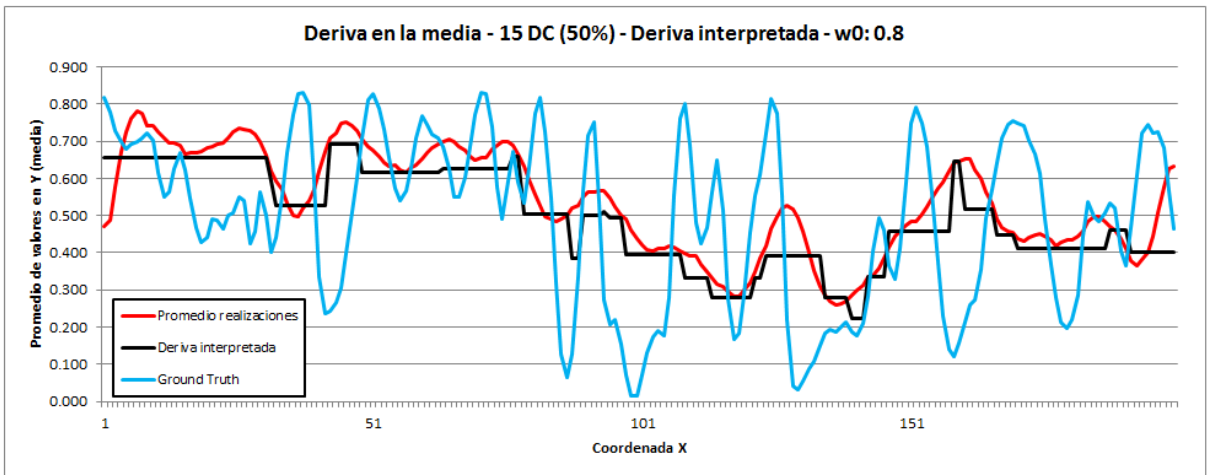
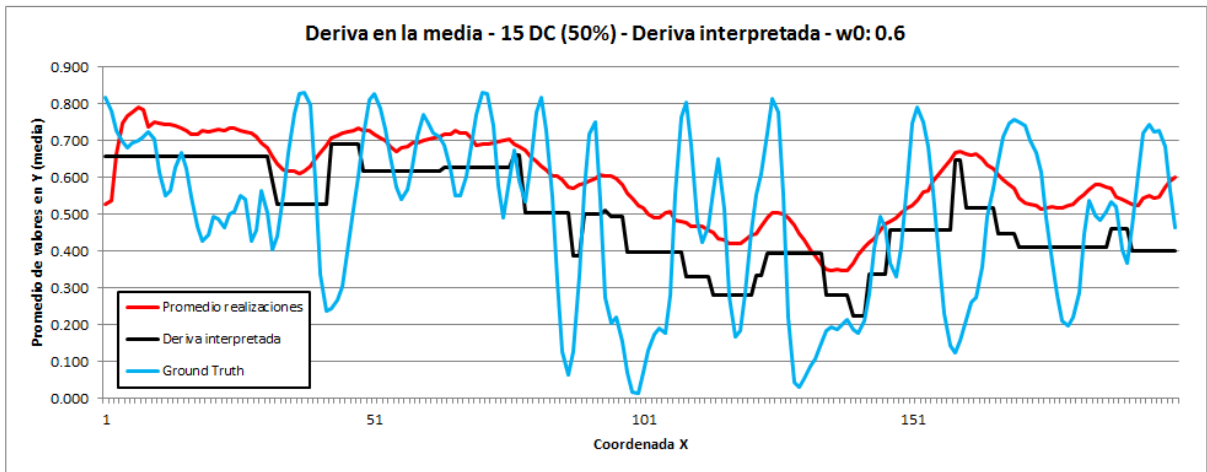










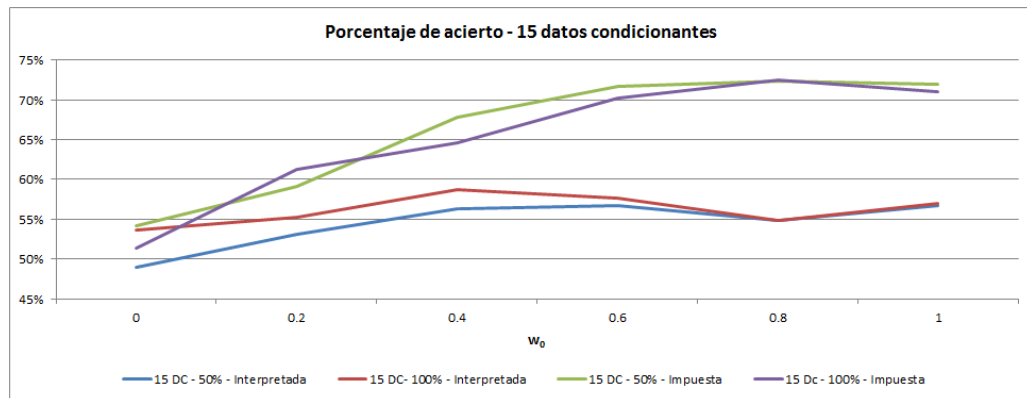


7.6. Anexo F

La presente sección de anexos presenta las matrices de acierto para los resultados expuestos en el caso de estudio - Roca sedimentaria. Se define previamente como umbral de corte el valor 0.63 indexando como 0 y 1 los valores bajo y sobre umbral respectivamente. Se denota por **Es** el valor 0 o 1 obtenido y por **D.S.** (debería ser) el valor 0 o 1 real aplicado sobre el Ground Truth. Por otro lado, se entiende por porcentaje de acierto a la suma de los porcentajes acertados al obtener 0 cuando debería ser 0 y 1 cuando debería ser 1.

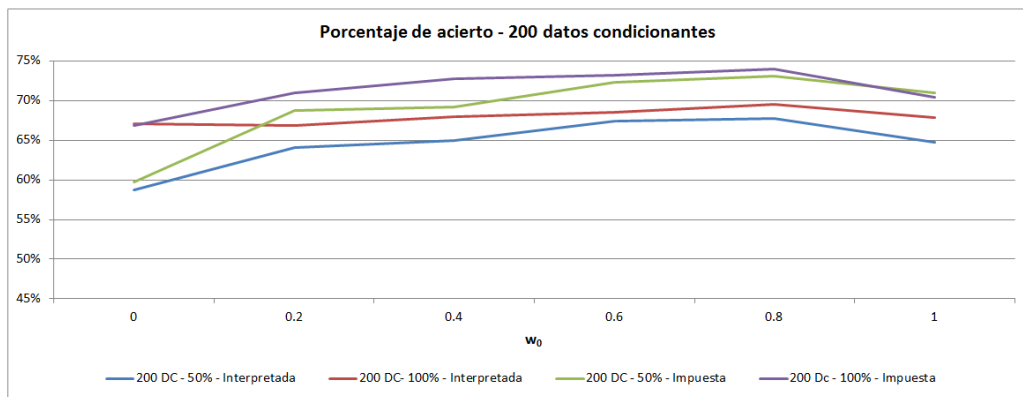
		15 Datos Condicionantes - Interpretado											
		50 por ciento						100 por ciento					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	5946	6411	6695	6759	6569	6738	6439	6670	7127	6852	6545	6872
0	1	5129	4664	4380	4316	4506	4337	4636	4405	3948	4223	4530	4203
1	0	5081	4730	4347	4337	4538	4328	4626	4561	4297	4246	4496	4412
1	1	3844	4195	4578	4588	4387	4597	4299	4364	4628	4679	4429	4513
% acierto		49.0	53.0	56.4	56.7	54.8	56.7	53.7	55.2	58.8	57.7	54.9	56.9

		15 Datos Condicionantes - Impuesto											
		50 por ciento						100 por ciento					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	6474	7279	7987	8344	8286	8369	6511	7213	7494	8117	8323	8179
0	1	4601	3796	3088	2731	2789	2706	4564	3862	3581	2958	2752	2896
1	0	4571	4378	3351	2927	2743	2894	5166	3892	3506	3007	2753	2908
1	1	4354	4547	5574	5998	6182	6031	3759	5033	5419	5918	6172	6017
% acierto		54.1	59.1	67.8	71.7	72.3	72.0	51.4	61.2	64.6	70.2	72.5	71.0



		200 Datos Condicionantes - Interpretado											
		50 por ciento						100 por ciento					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	7034	7478	7699	7769	7912	7517	7924	7688	7808	7953	8011	7841
0	1	4041	3597	3376	3306	3163	3558	3151	3387	3267	3122	3064	3234
1	0	4221	3585	3623	3202	3297	3503	3425	3241	3146	3162	3022	3190
1	1	4704	5340	5302	5723	5628	5422	5500	5684	5779	5763	5903	5735
% acierto		58.7	64.1	65.0	67.5	67.7	64.7	67.1	66.9	67.9	68.6	69.6	67.9

		200 Datos Condicionantes - Impuesto											
		50 por ciento						100 por ciento					
Es	D.S.	0	0.2	0.4	0.6	0.8	1	0	0.2	0.4	0.6	0.8	1
0	0	7174	7867	7955	8333	8369	8169	7729	8054	8448	8344	8483	8154
0	1	3901	3208	3120	2742	2706	2906	3346	3021	2627	2731	2592	2921
1	0	4151	3035	3044	2802	2674	2896	3282	2771	2814	2627	2619	2993
1	1	4774	5890	5881	6123	6251	6029	5643	6154	6111	6298	6306	5932
% acierto		59.7	68.8	69.2	72.3	73.1	71.0	66.9	71.0	72.8	73.2	73.9	70.4

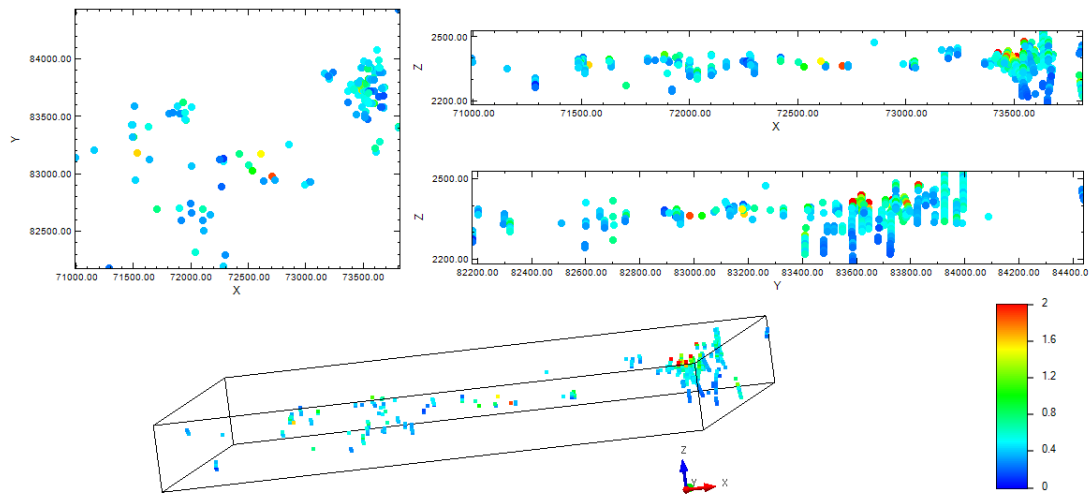


7.7. Anexo G

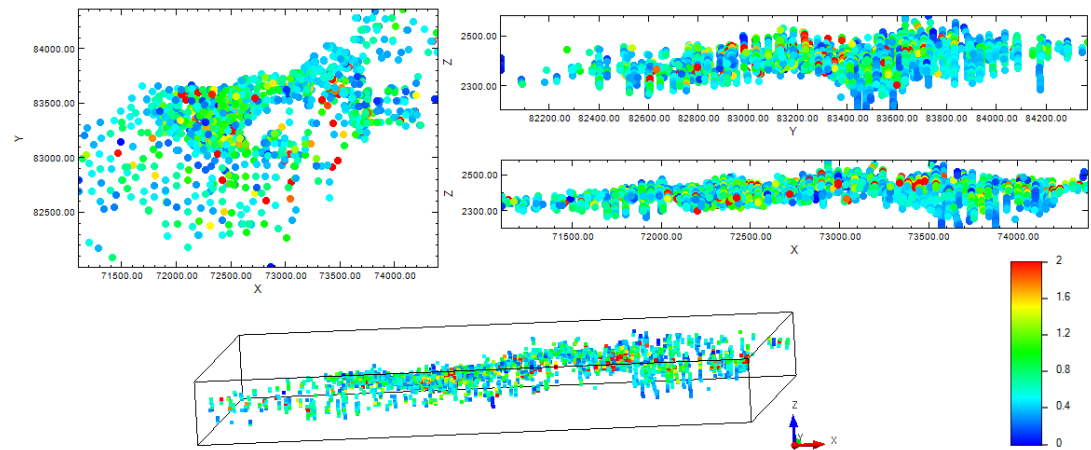
Se presentan a continuación las visualizaciones en planta, perfil e isométricos para las distintas unidades geológicas. La información proviene de los sondajes y pozos de tronadura en el caso de estudio Cerro Colorado.

Sondajes

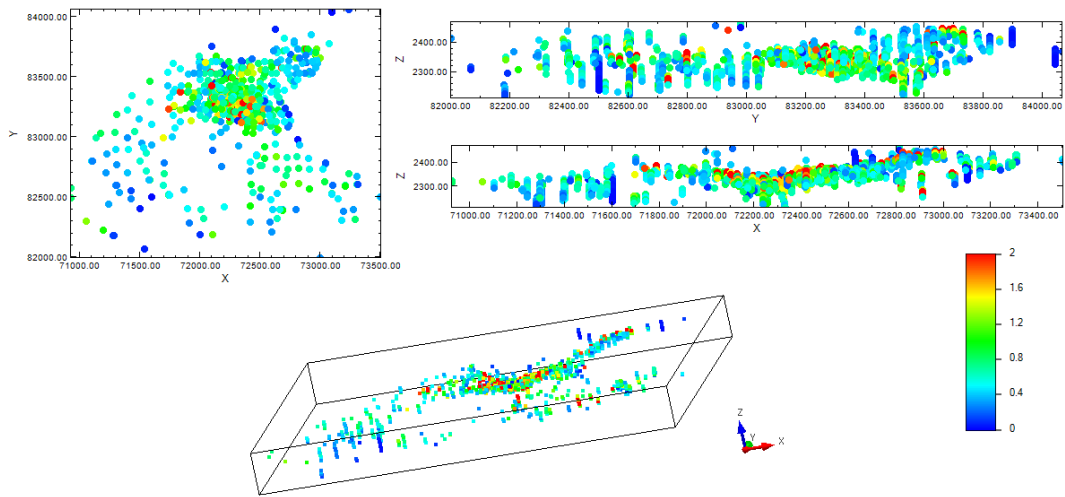
Vistas en planta, perfiles e isométrico UGCuT2.



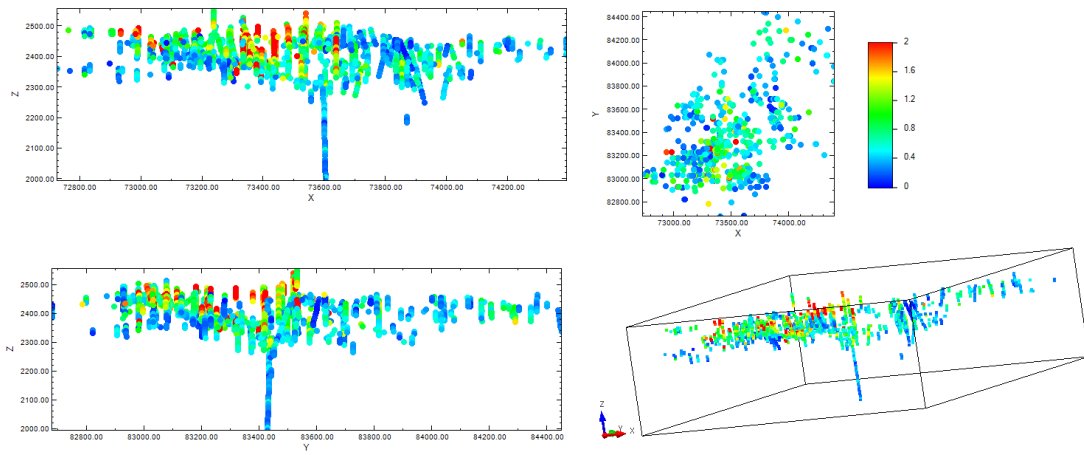
Vistas en planta, perfiles e isométrico UGCuT3.



Vistas en planta, perfiles e isométrico UGCuT4.

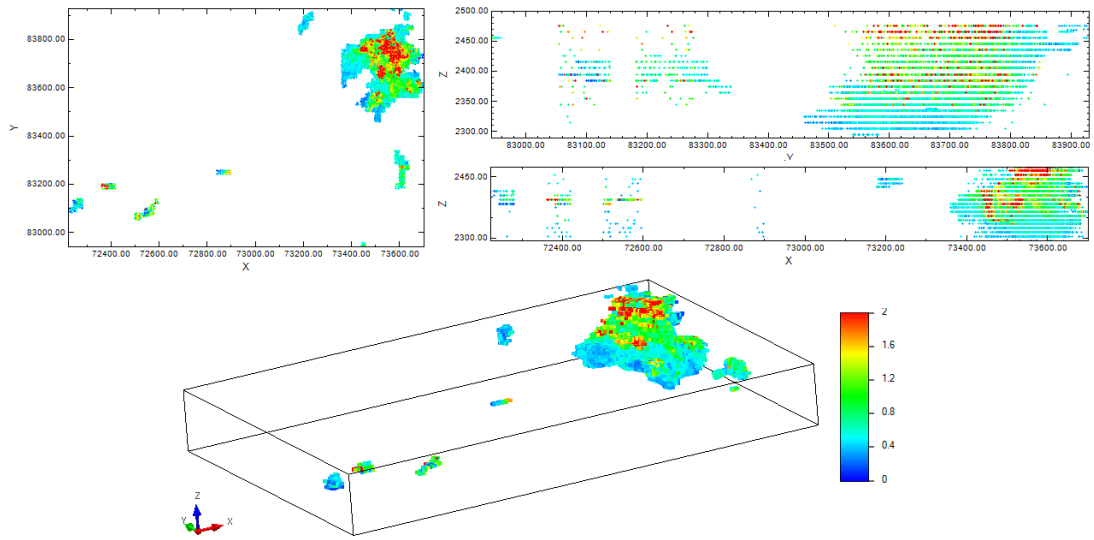


Vistas en planta, perfiles e isométrico UGCuT5.

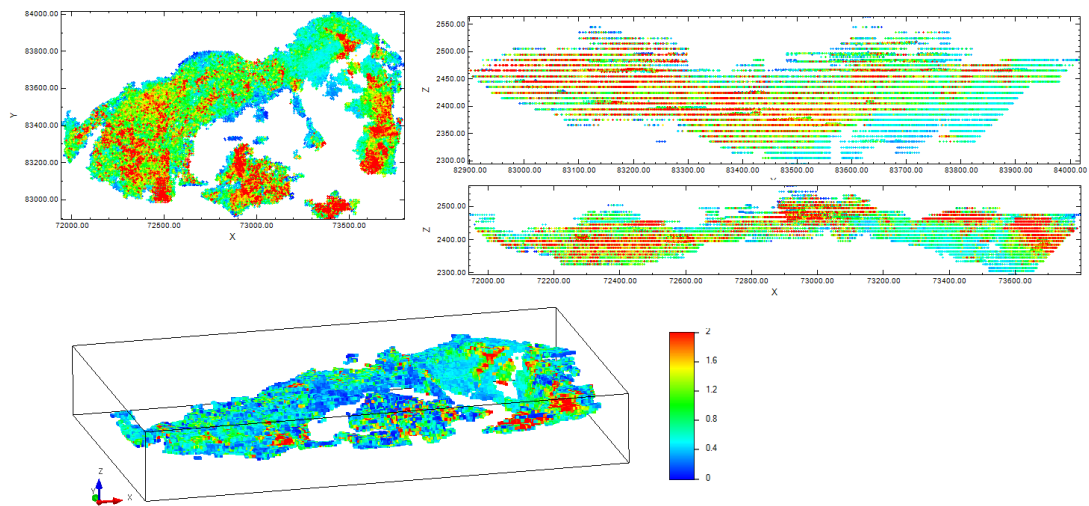


Pozos de tronadura

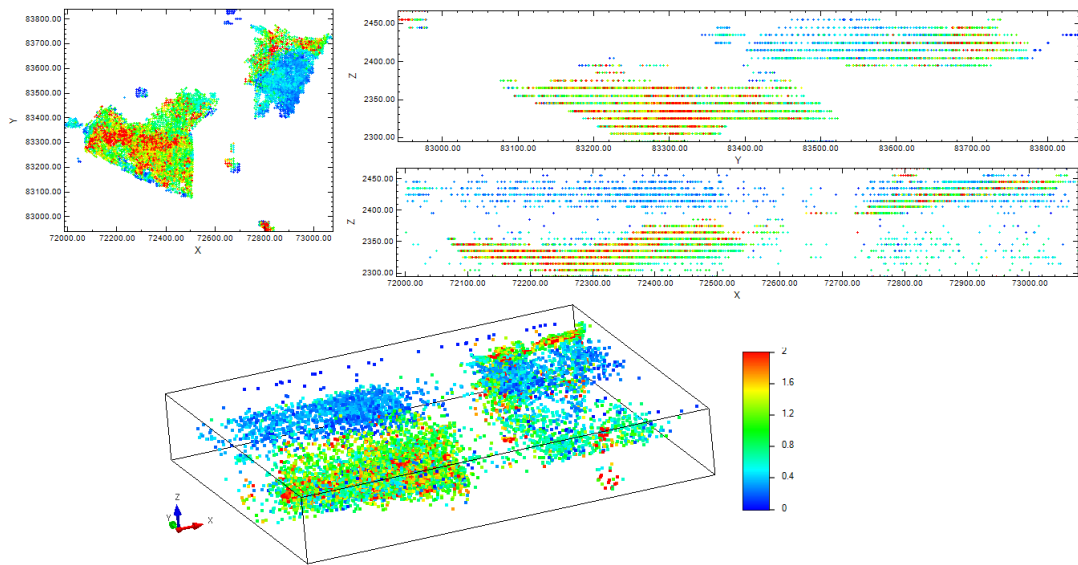
Vistas en planta, perfiles e isométrico UGCuT2.



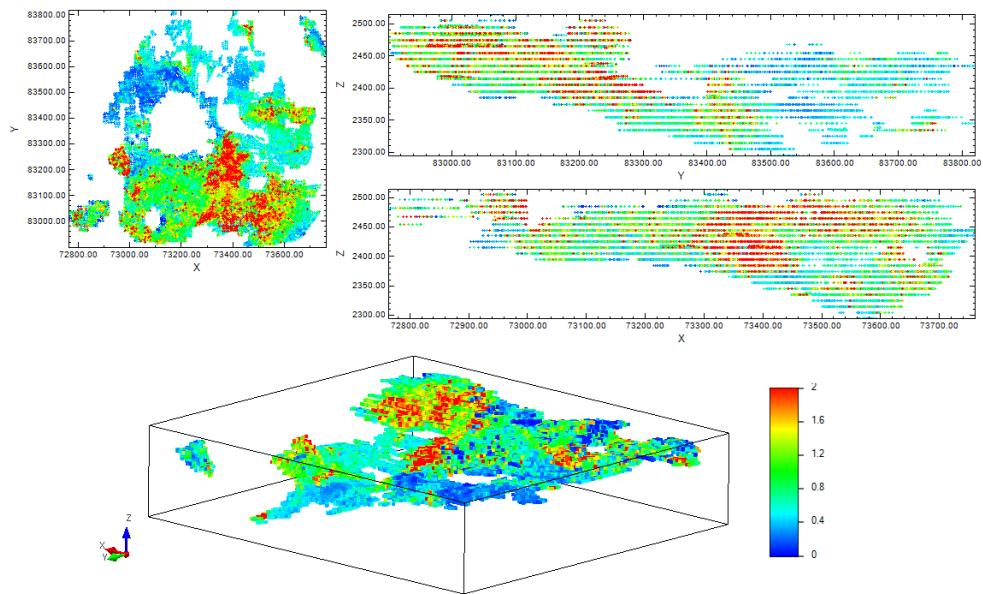
Vistas en planta, perfiles e isométrico UGCuT3.



Vistas en planta, perfiles e isométrico UGCuT4.



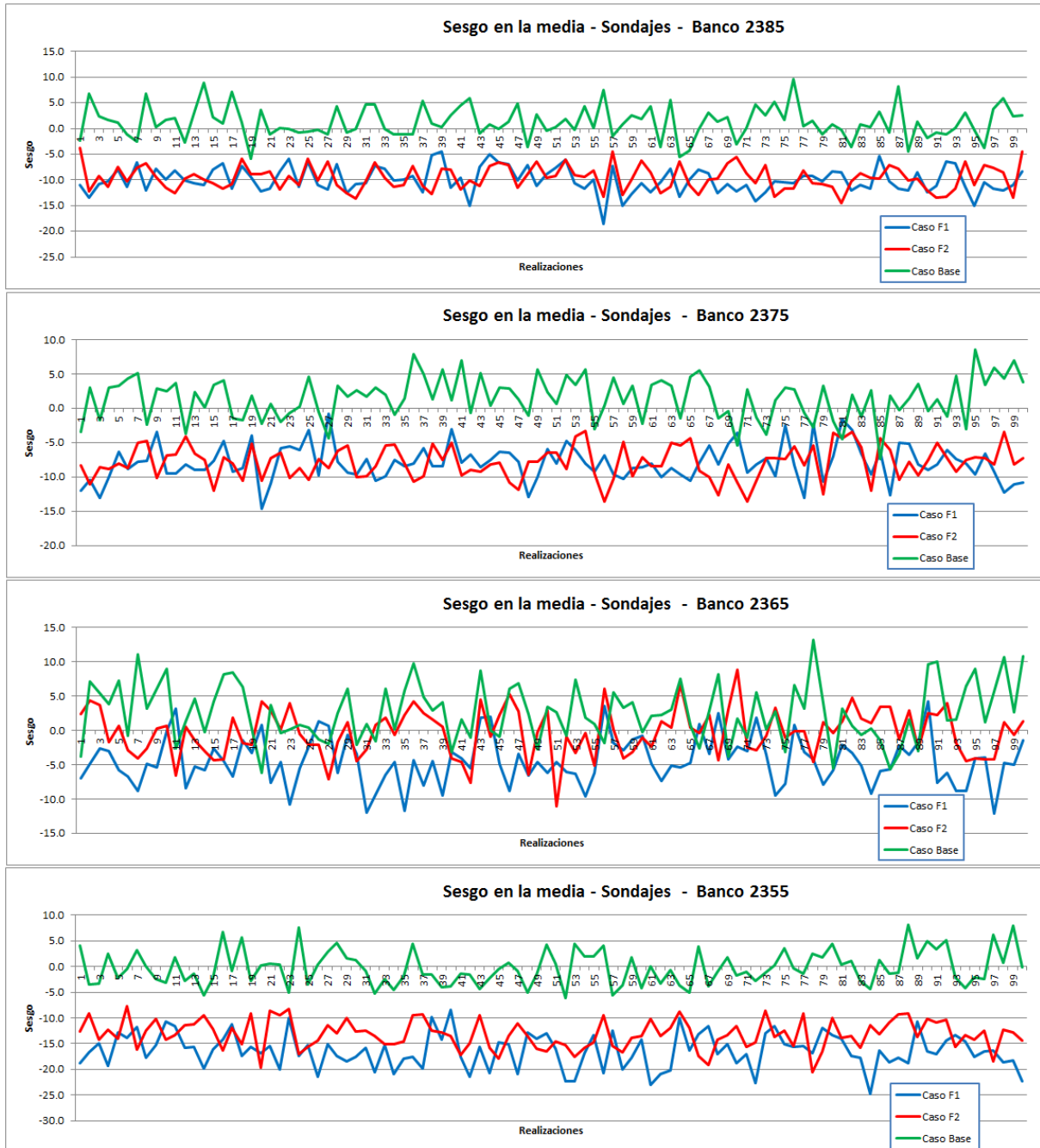
Vistas en planta, perfiles e isométrico UGCuT5.



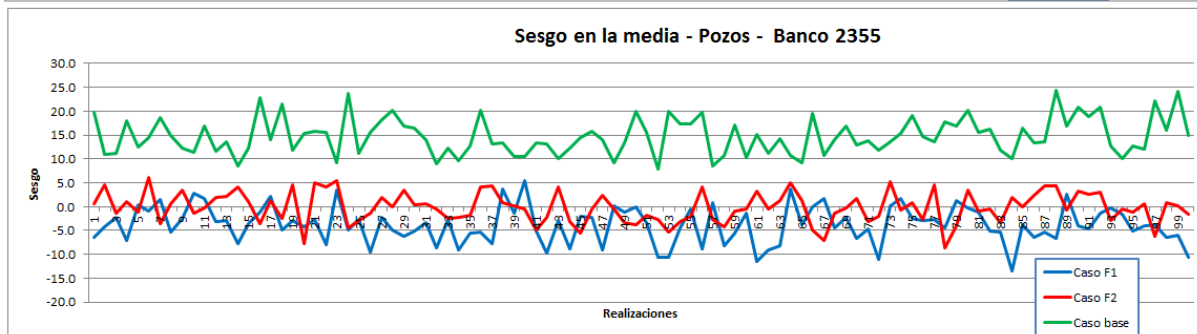
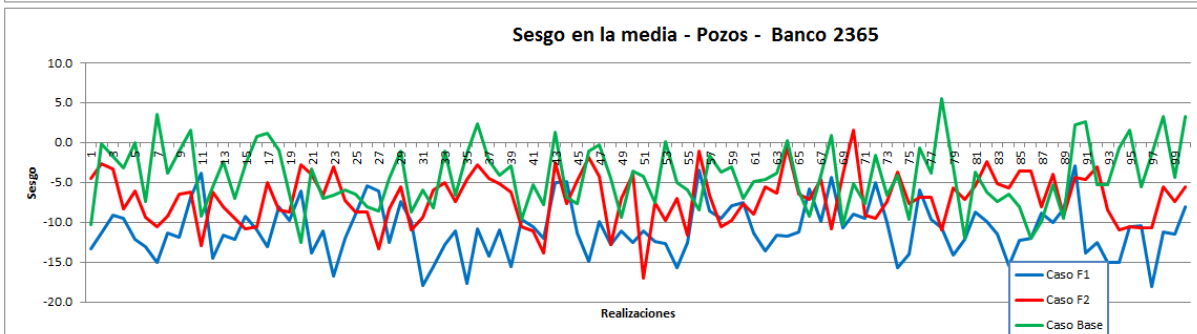
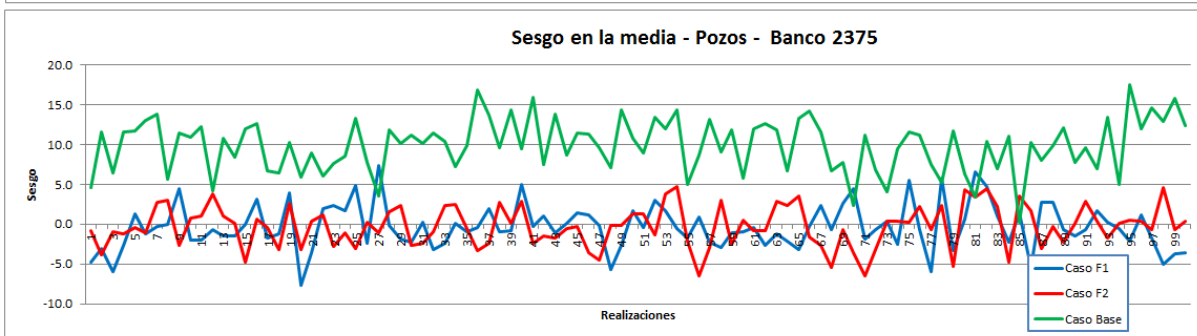
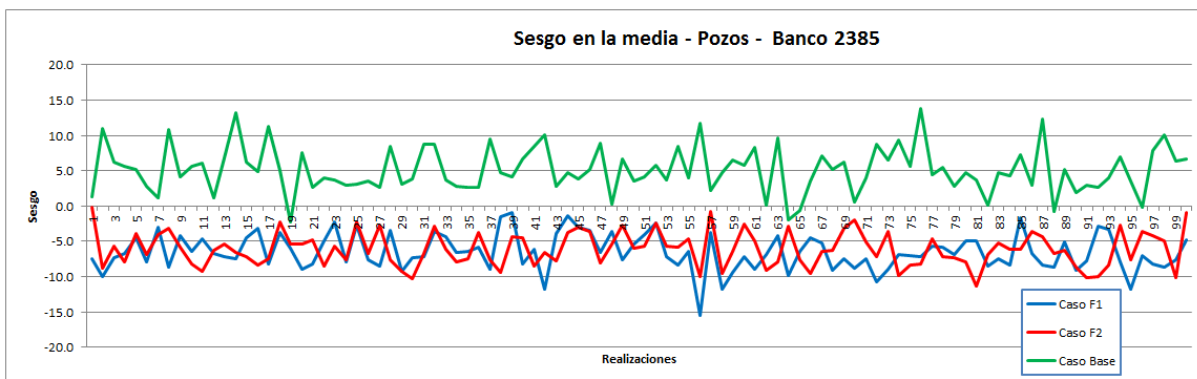
7.8. Anexo H

Se presenta a continuación los sesgos obtenidos en cada una de las 100 realizaciones para cada caso de simulación, banco a banco.

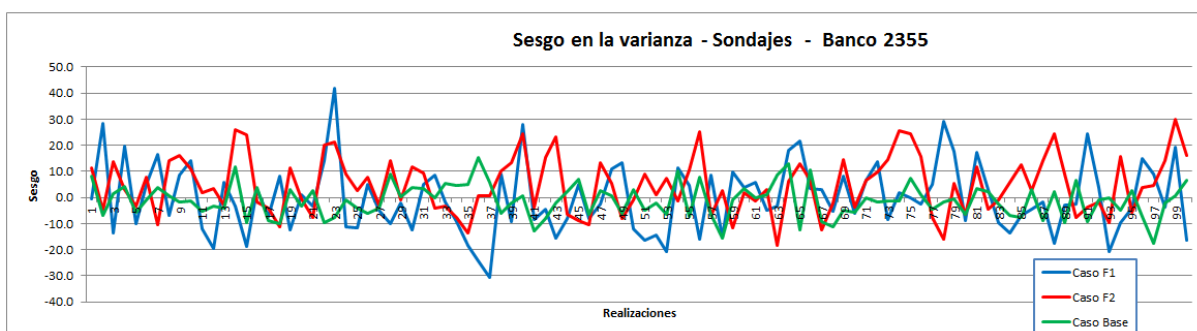
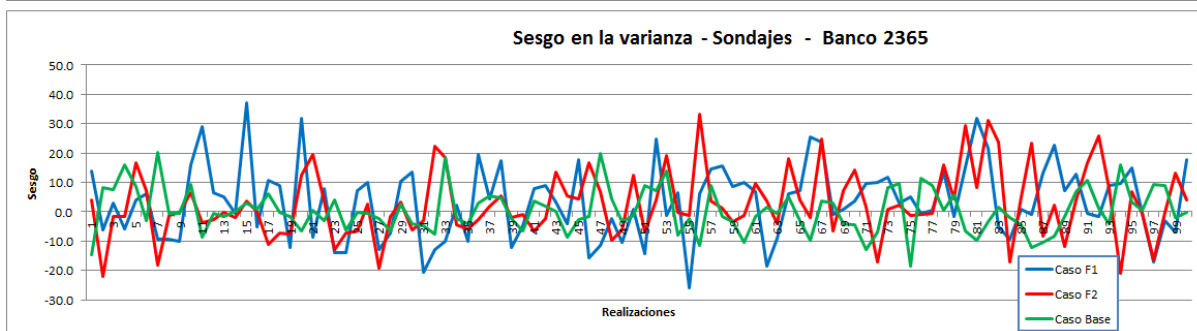
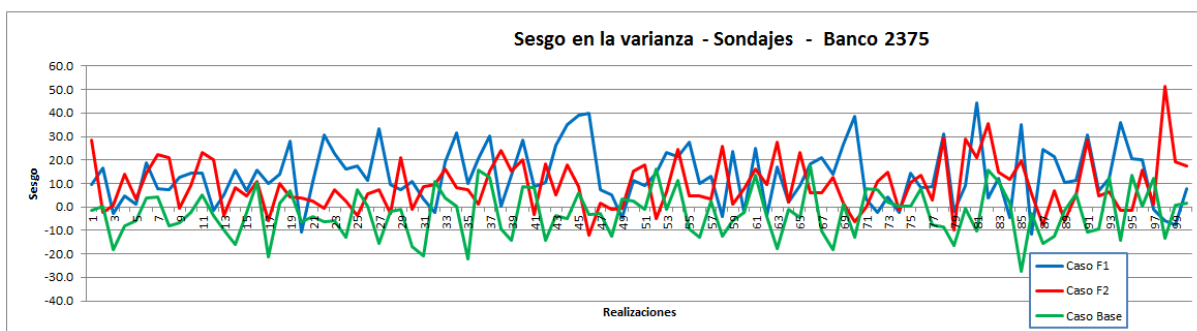
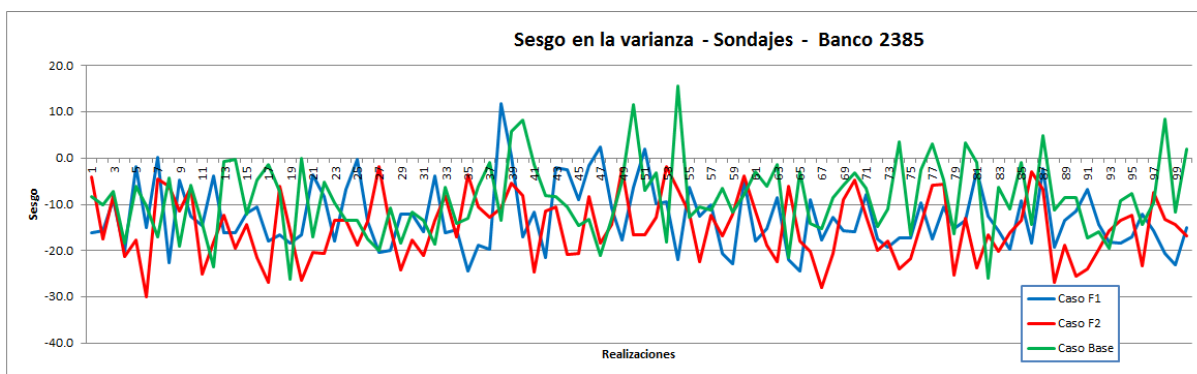
Sesgo en la media de las realizaciones en contraste con la media de los sondeos



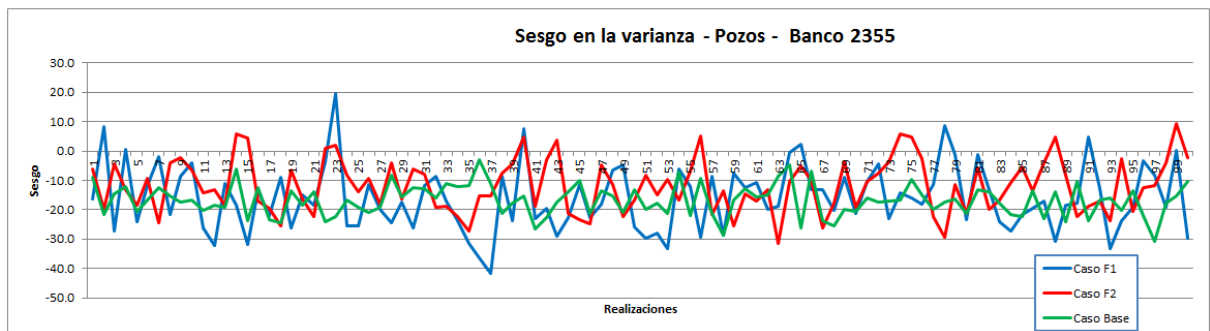
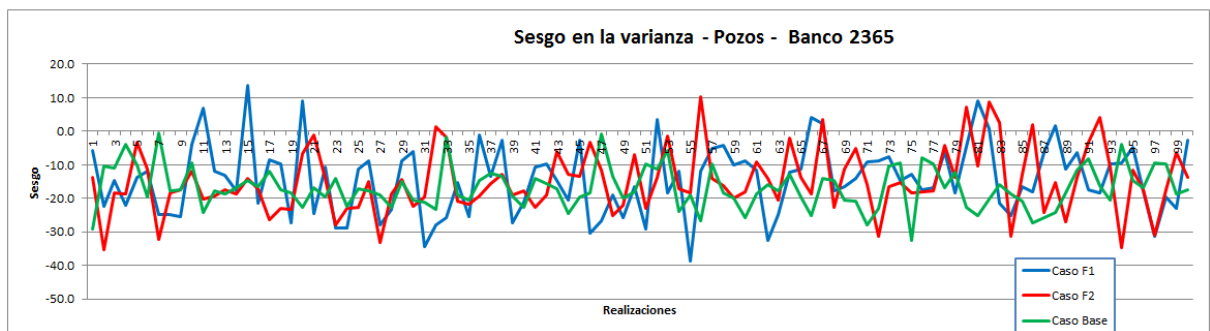
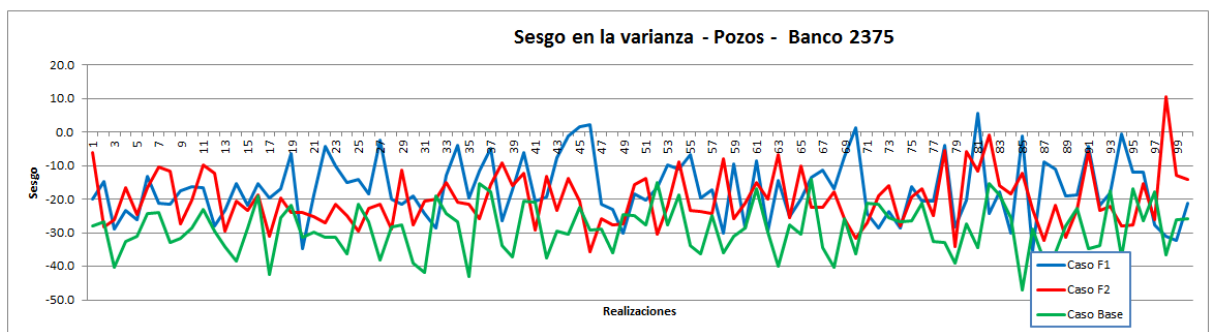
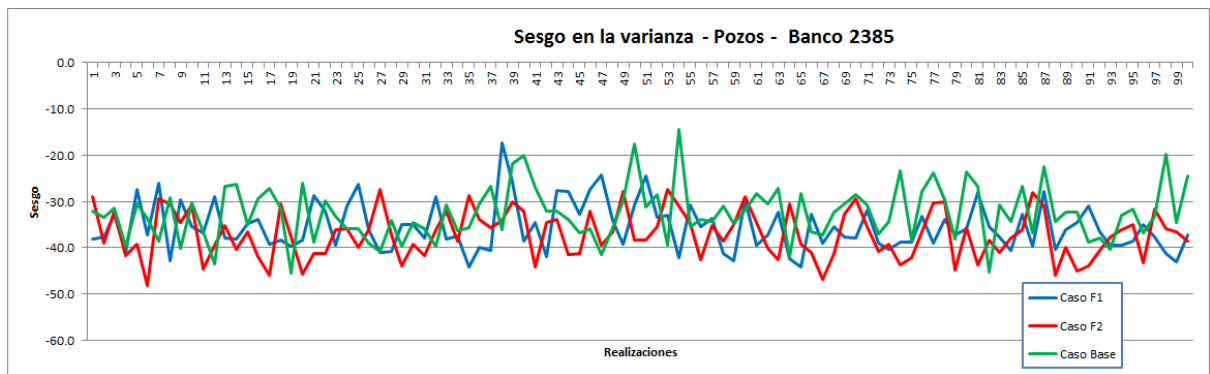
Sesgo en la media de las realizaciones en contraste con la media de los pozos



Sesgo en la varianza de las realizaciones en contraste con la varianza de los sondeos



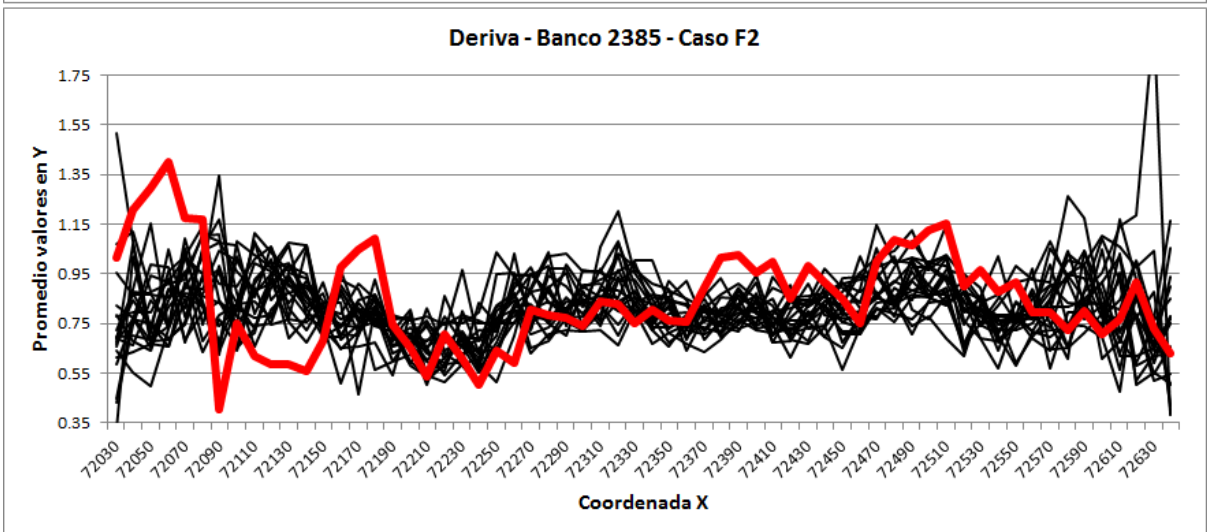
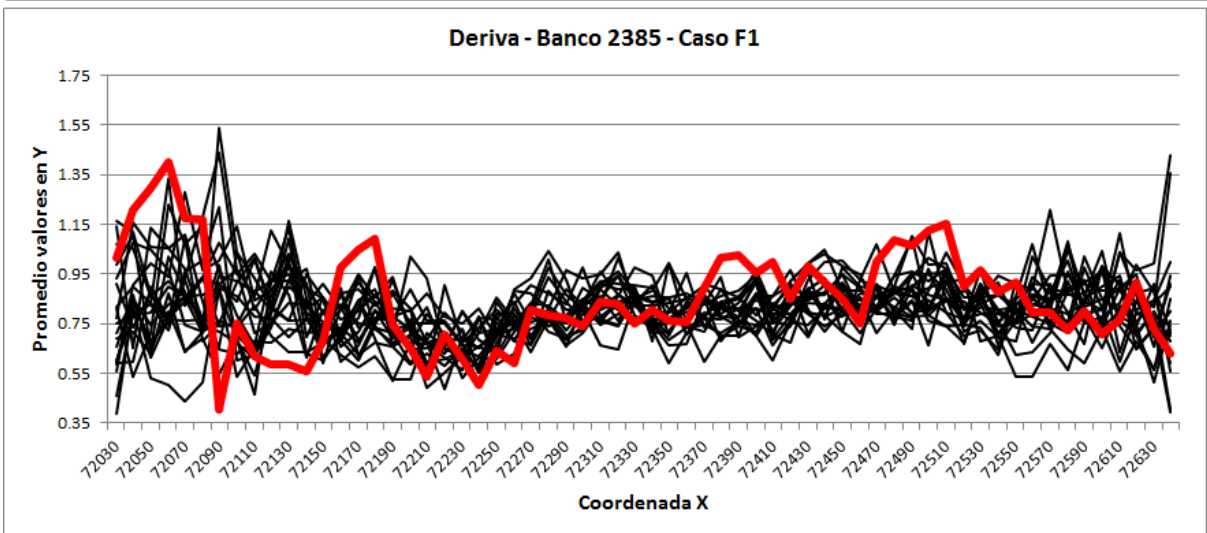
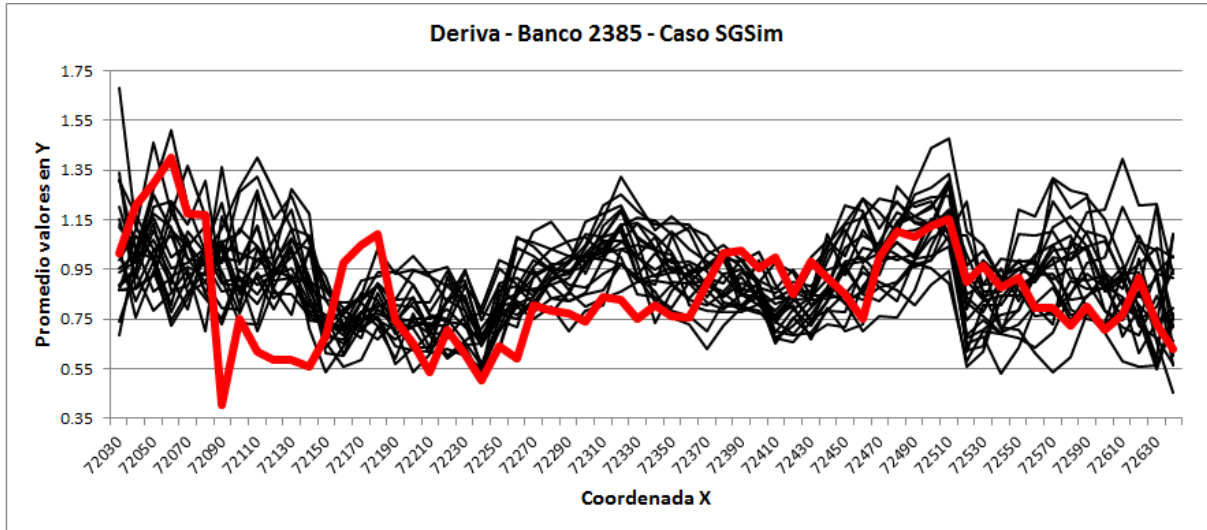
Sesgo en la varianza de las realizaciones en contraste con la varianza de los pozos

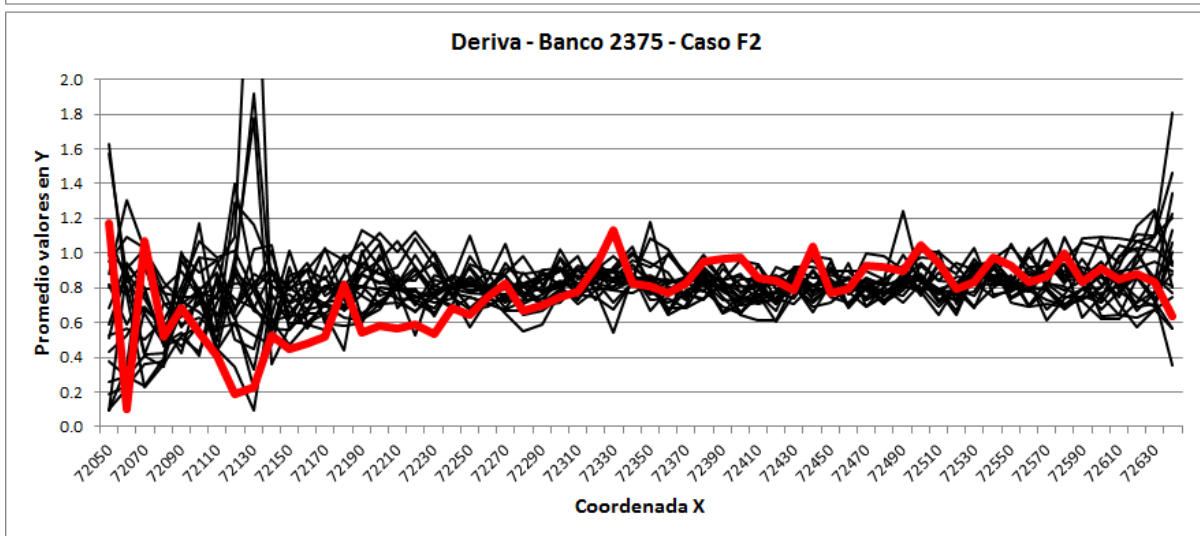
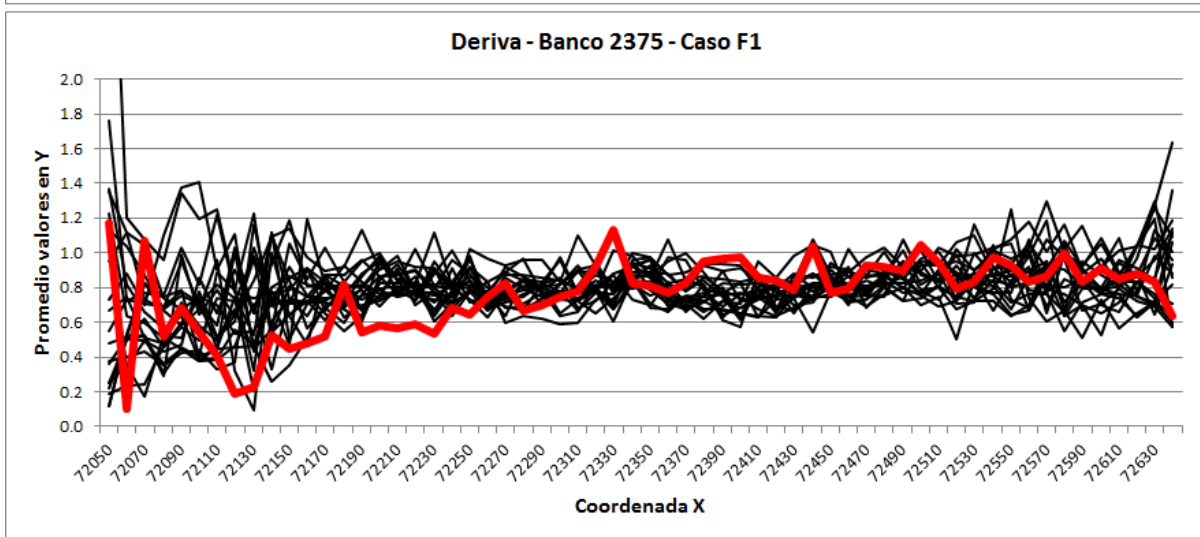
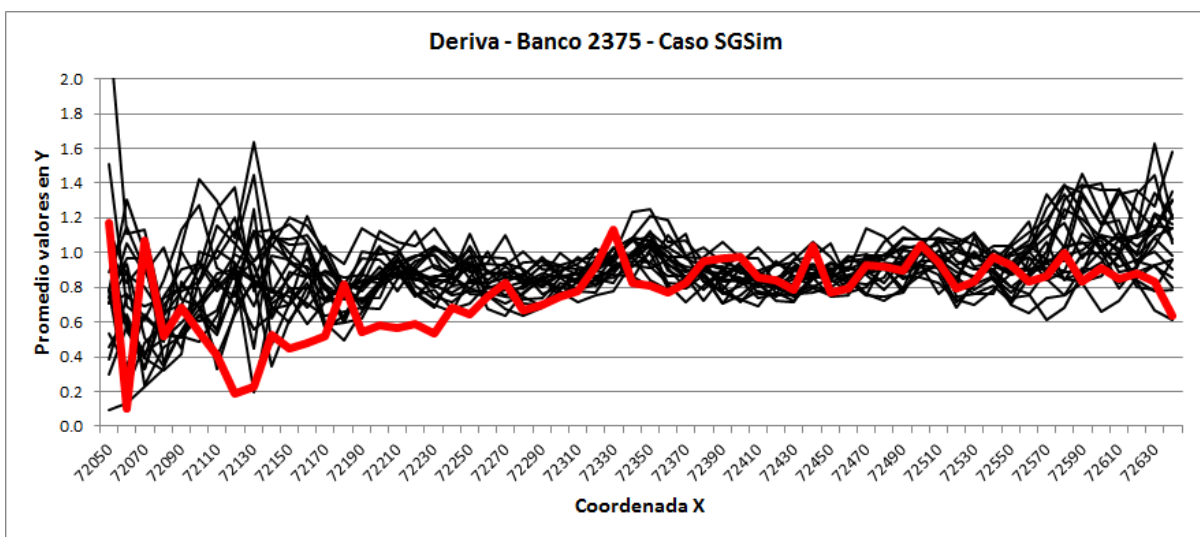


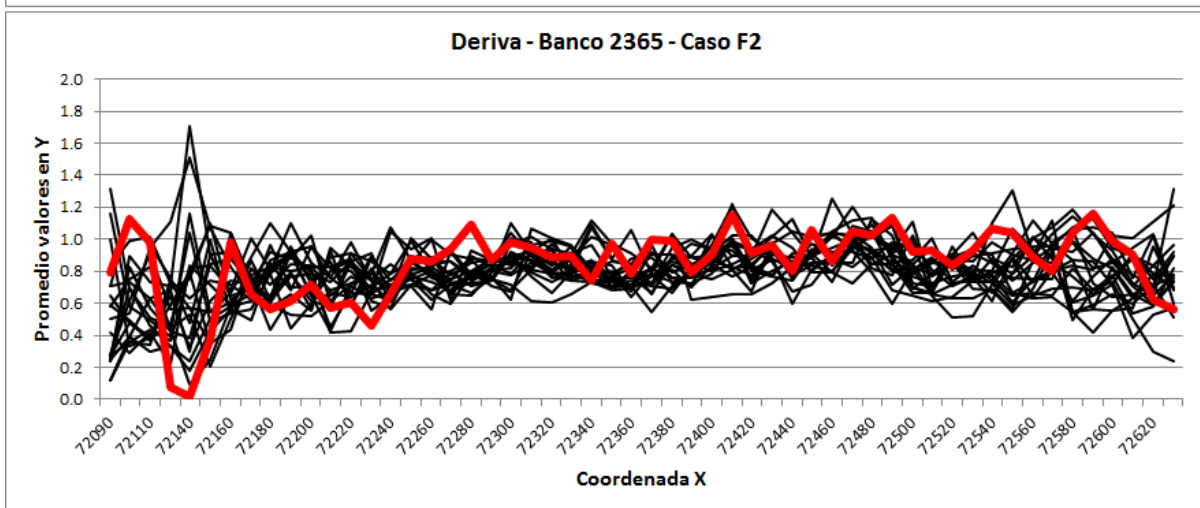
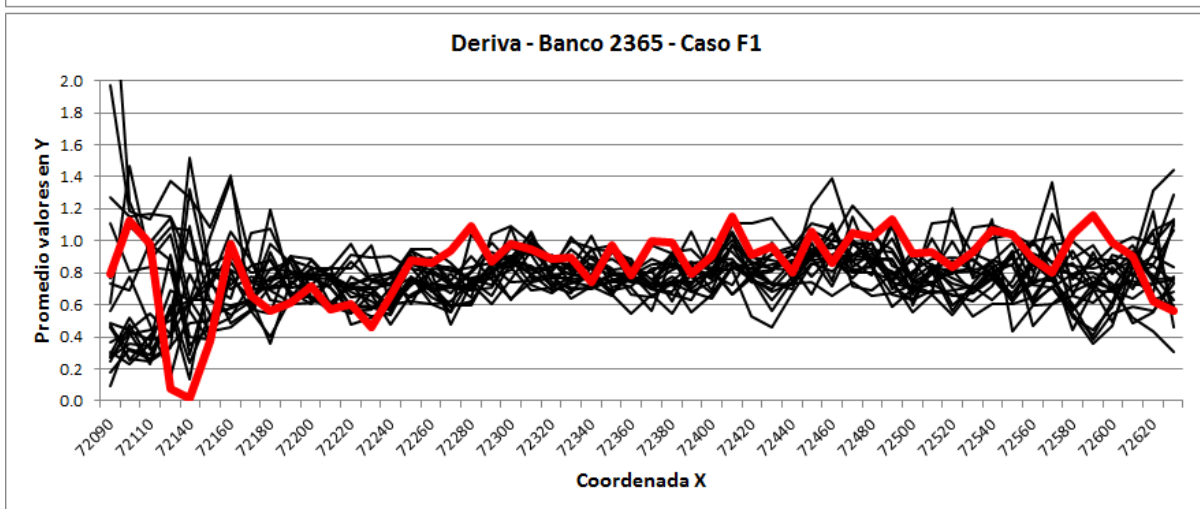
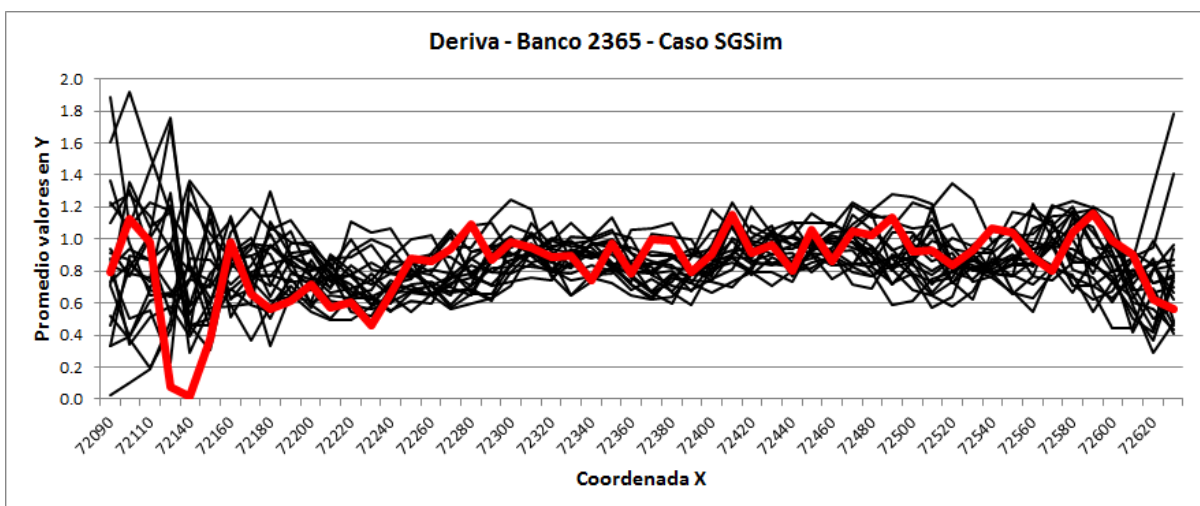
7.9. Anexo I

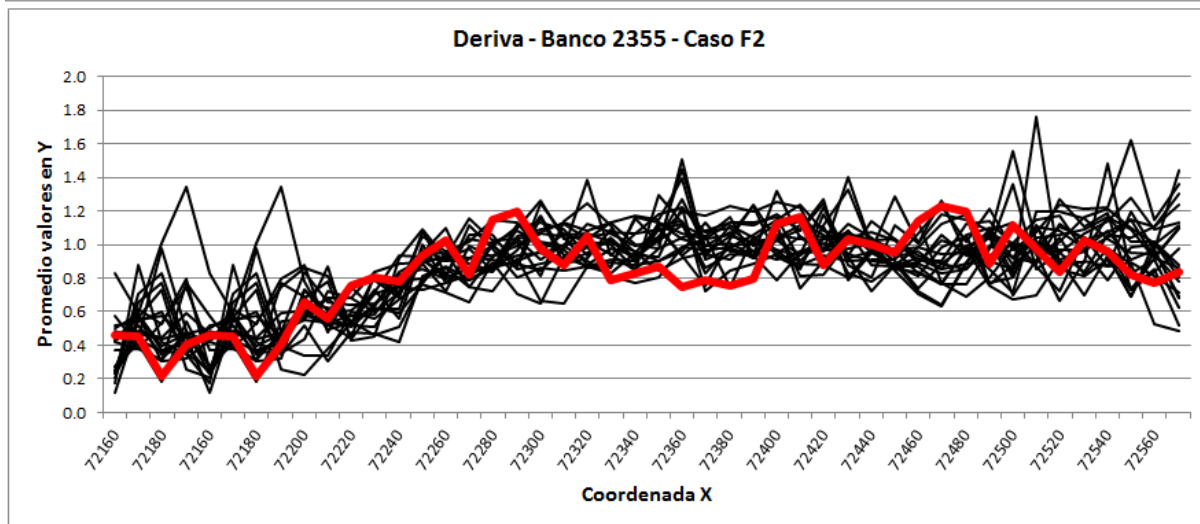
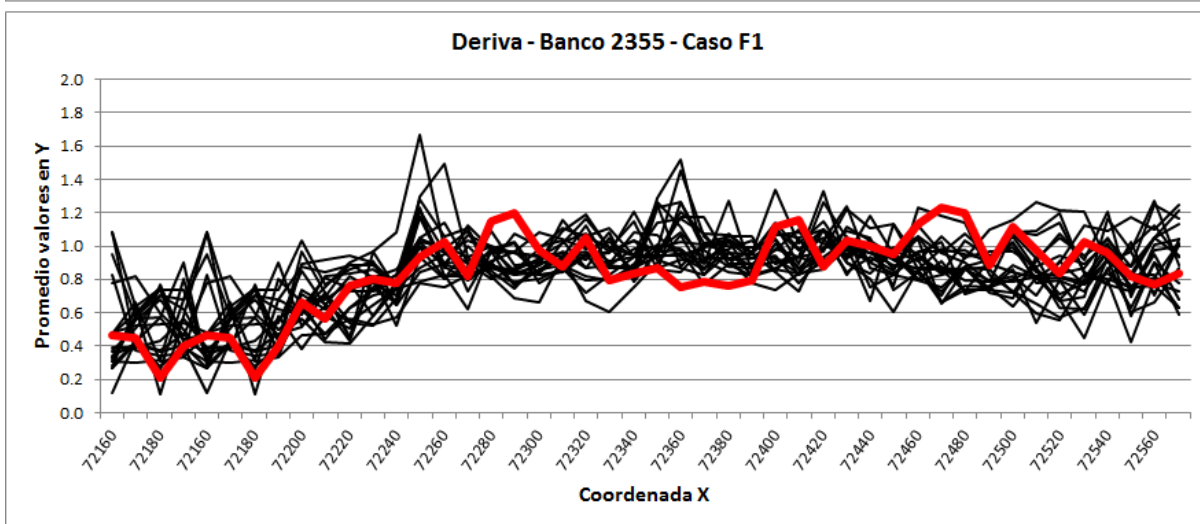
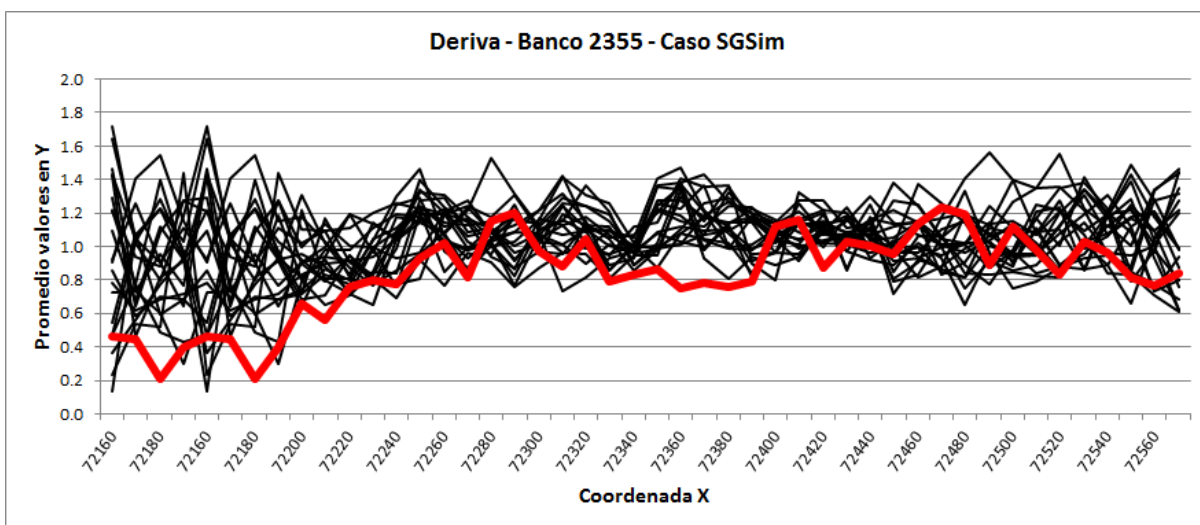
Se presenta a continuación las derivas en la media, a lo largo de la coordenada X, banco a banco.

Derivas en la media - Coordenada X - Banco 2385



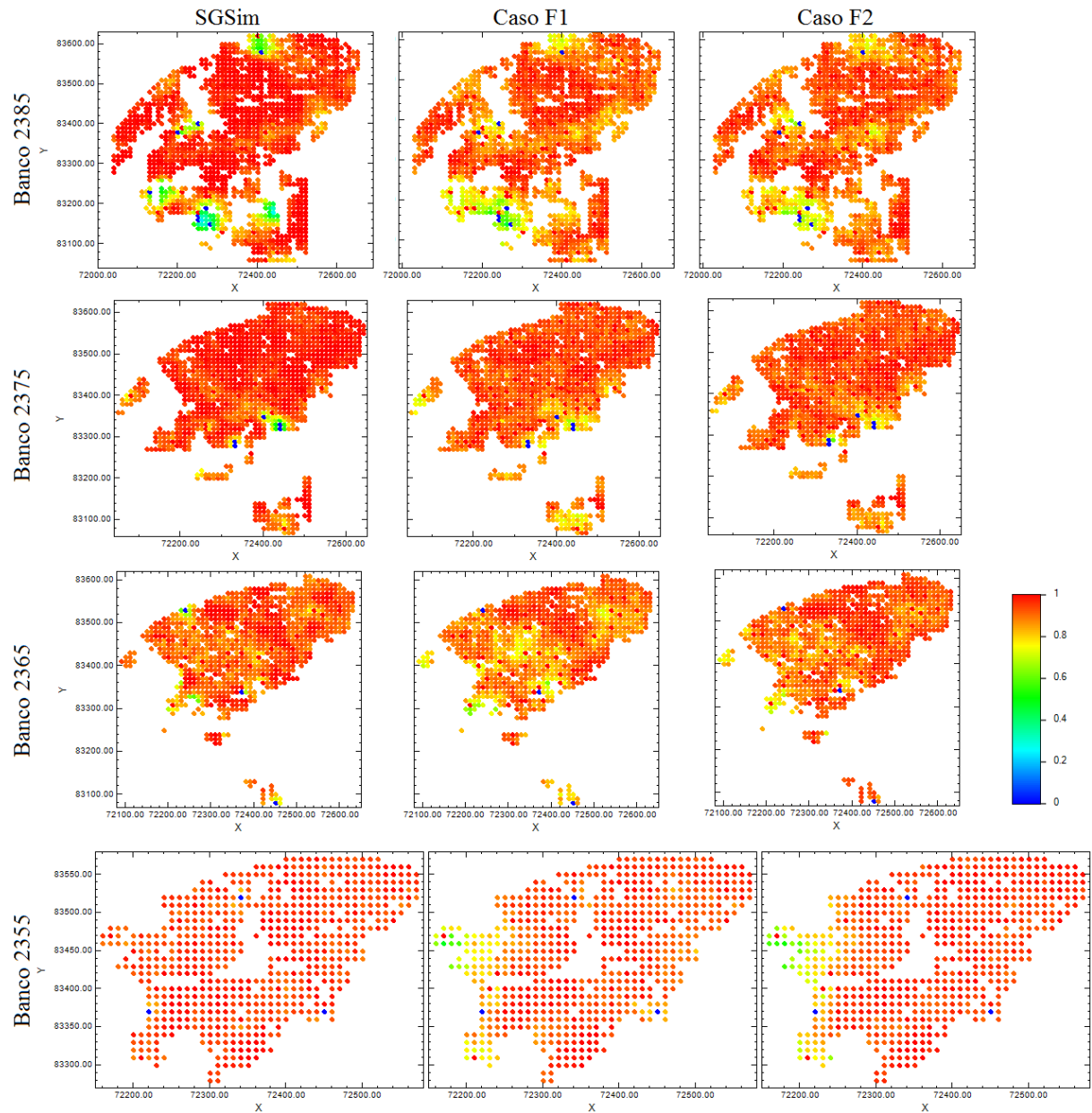


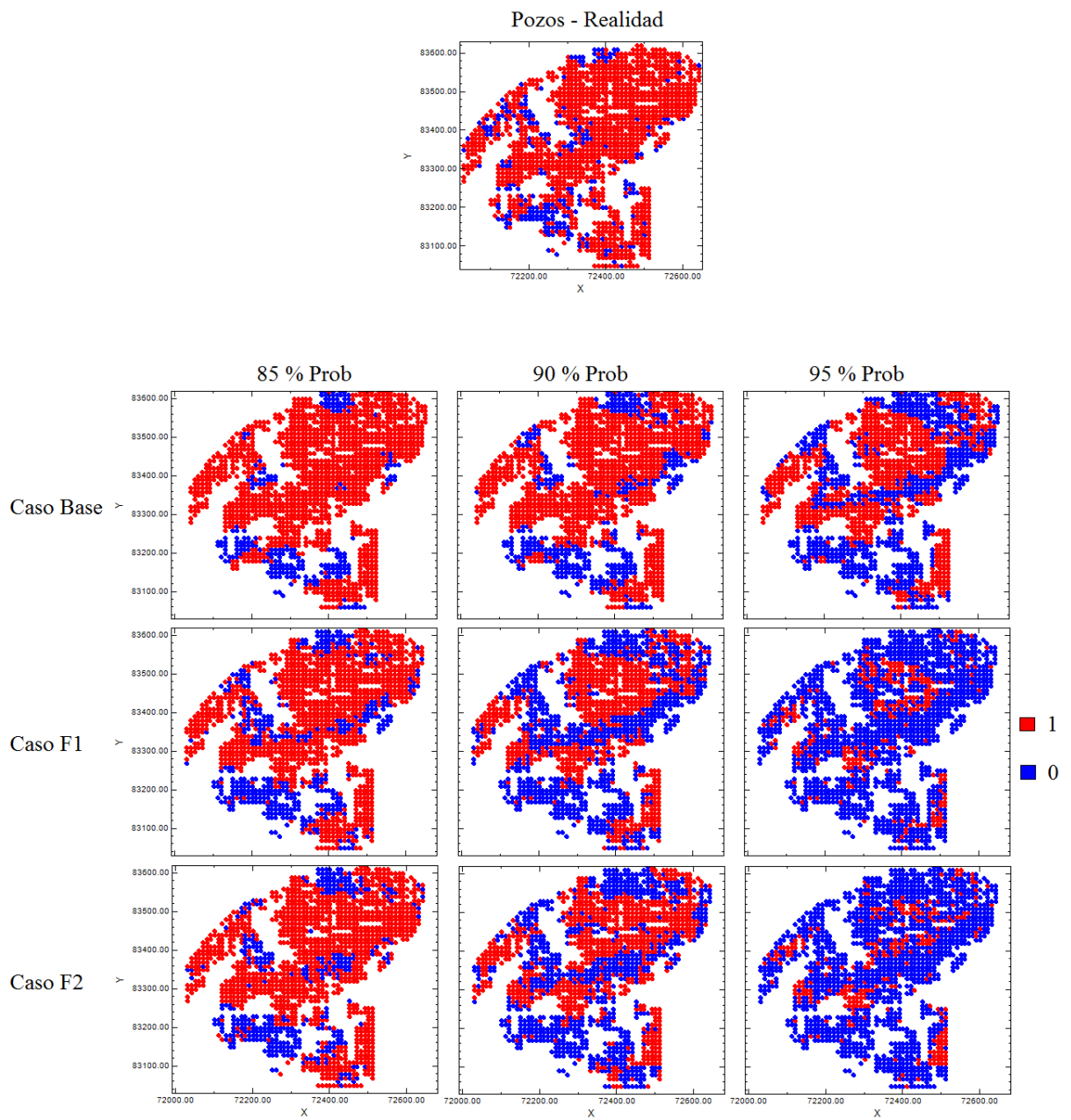




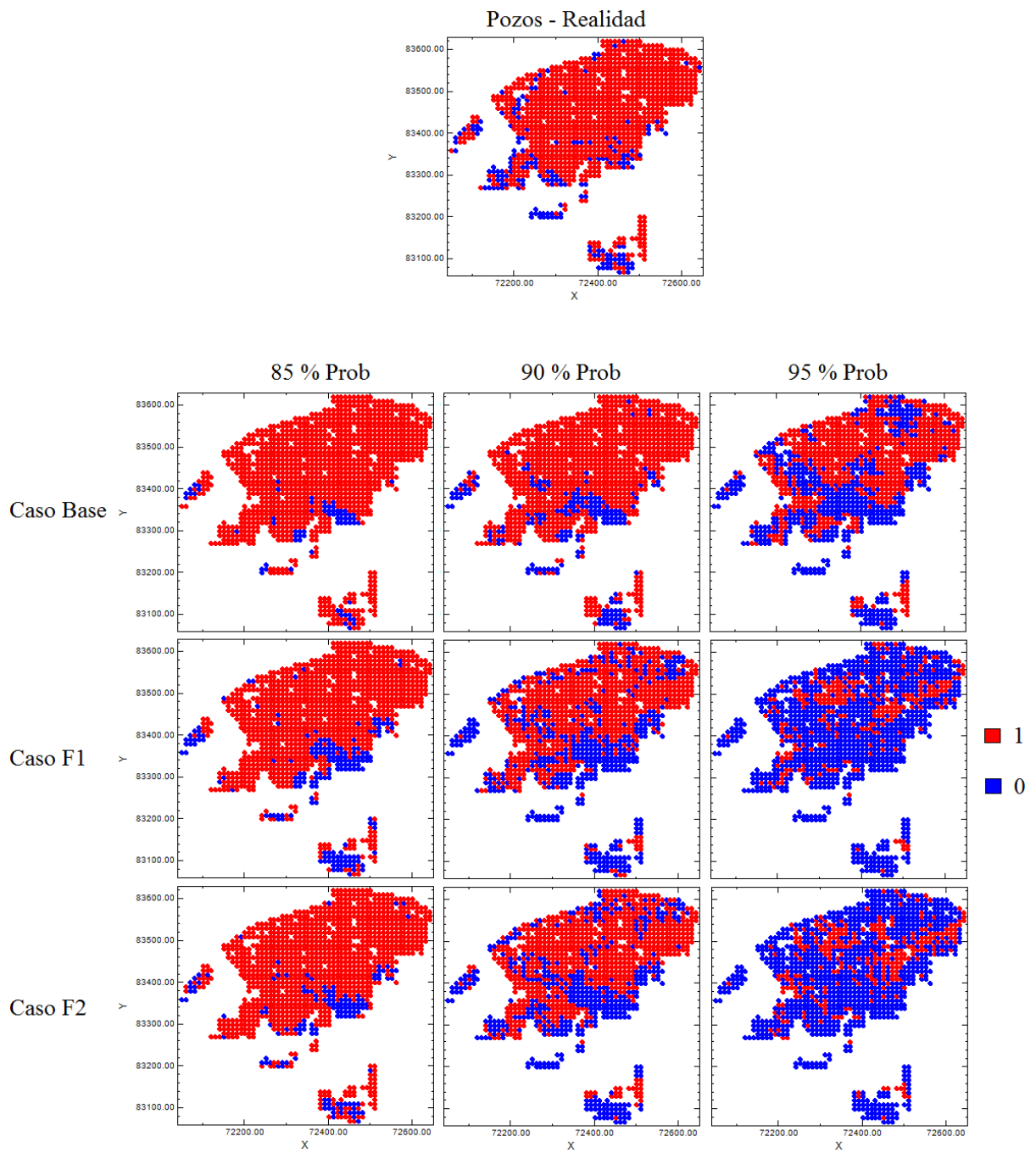
7.10. Anexo J

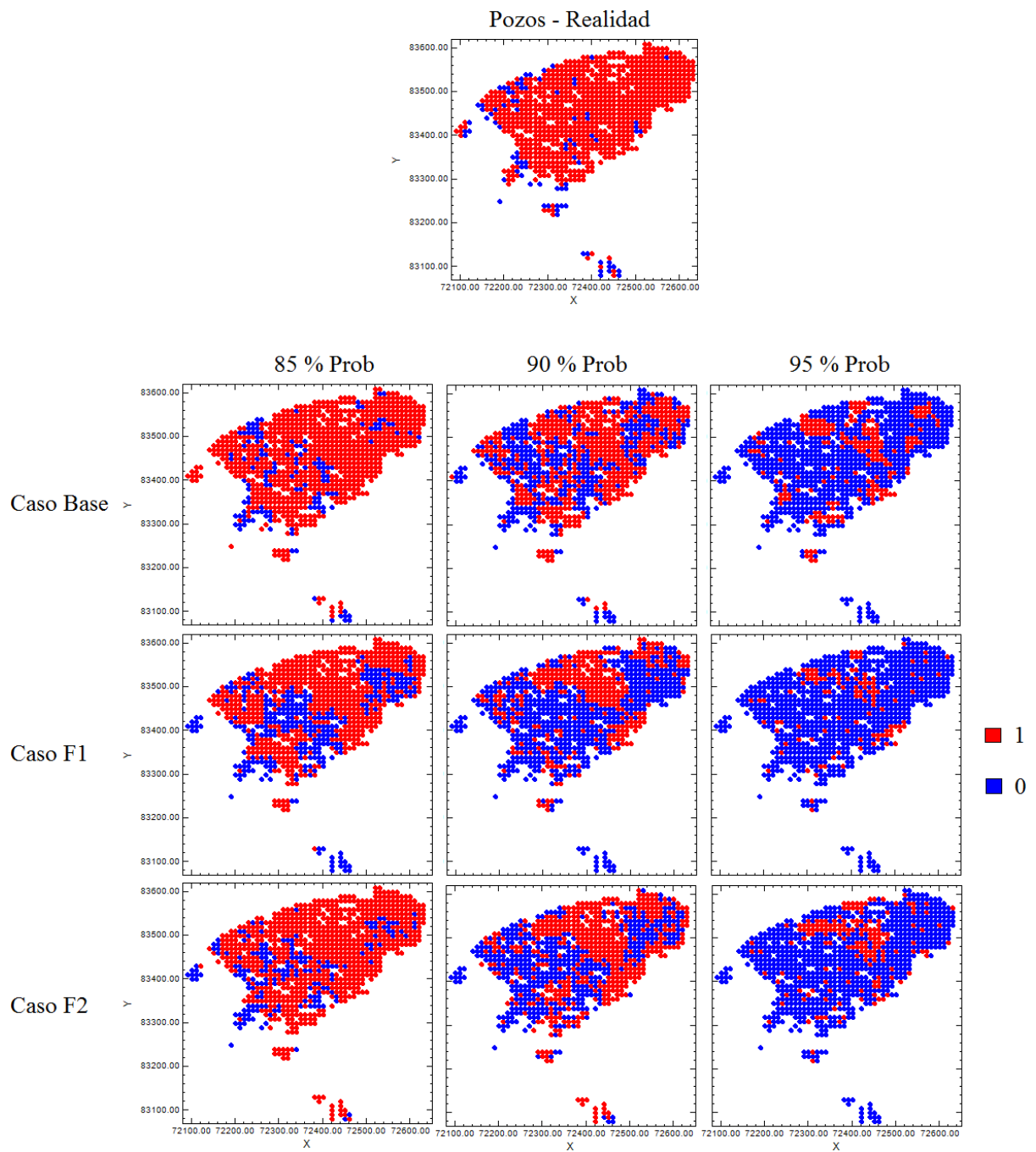
Se presenta a continuación los mapas de probabilidad banco a banco para cada simulación.



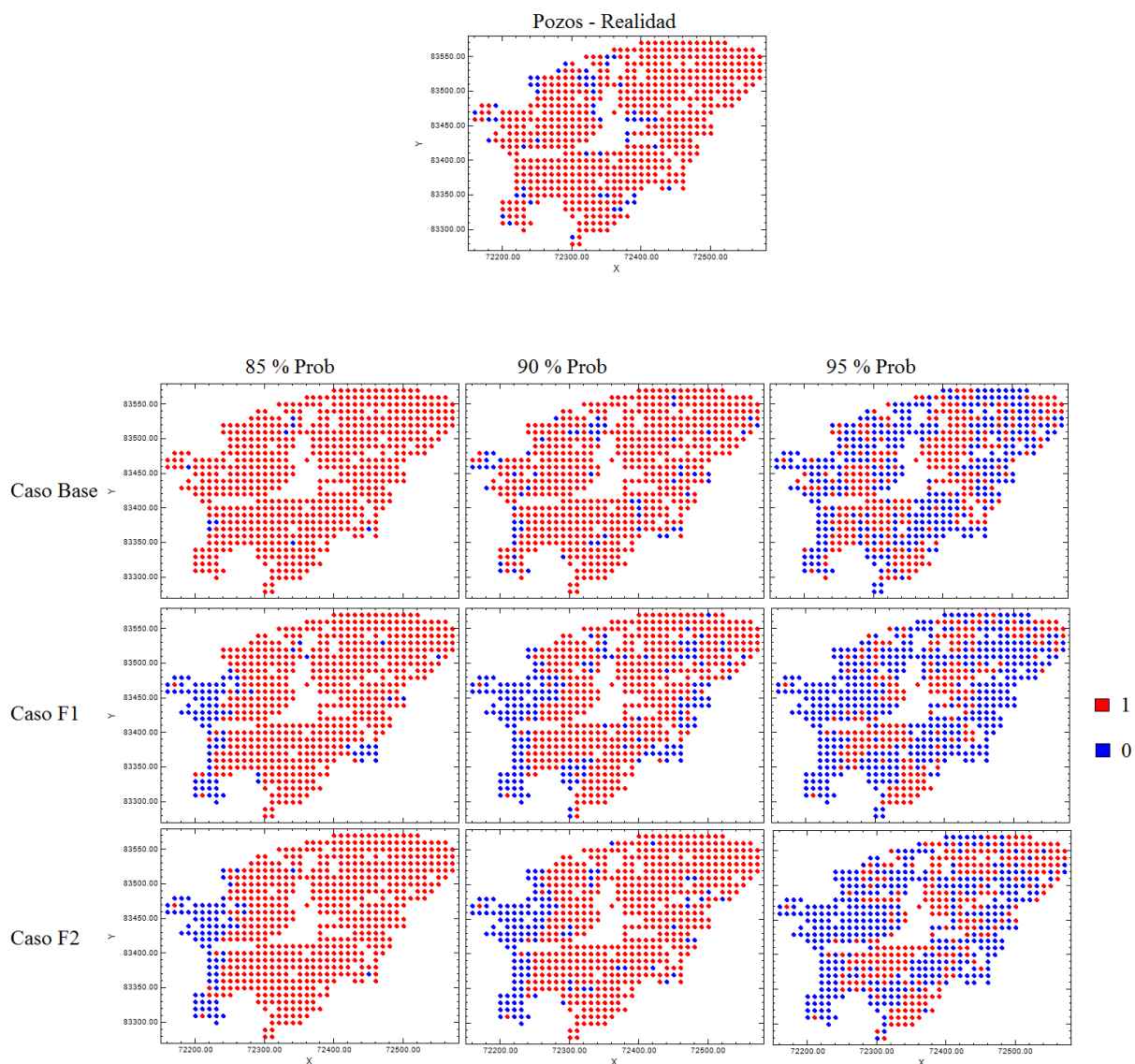


Categorización según umbral de probabilidad - Banco 2375





Categorización según umbral de probabilidad - Banco 2355



7.11. Anexo K

Se presenta a continuación las curvas Tonelaje-Ley para los 4 bancos en estudio. En color Rojo se presenta la realidad (pozos de tronadura) y en negro 20 de las 100 realizaciones.

Banco 2385

