



Trabajo final de master

Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia

Valencia, septiembre de 2017

TITULACIÓN: Master en Ingeniería de Caminos, Canales y Puertos

Curso: 2016/2017

AUTORA: Azcaray Fernández, Cristina

TUTOR: Rodrigo-Illarri, Javier

COTUTORA: Rodrigo-Clavero, M^a Elena

Índice

1.Introducción	9
1.1. Antecedentes	9
1.2. Aplicaciones potenciales de la geoestadística para la caracterización de emplazamientos de vertederos de residuos	10
1.3. Objetivos y estructura	11
2.Conceptos básicos de geoestadística.....	13
2.1. Introducción.....	13
2.2. Descripción univariada	14
2.2.1. Introducción	14
2.2.2. Métodos gráficos para la descripción univariada.....	14
2.2.3. Métodos numéricos para la descripción univariada	16
2.3. Descripción bivariada	19
2.3.1. Introducción	19
2.3.2. Métodos gráficos para la descripción bivariada.....	19
2.3.3. Métodos numéricos para la descripción bivariada	21
2.4. Descripción espacial	23
2.4.1. Introducción	23
2.4.2. Visualización espacial de datos	23
2.4.3. Continuidad espacial.....	25
2.4.4. Práctica de la descripción espacial.	34
2.5. El modelo de función aleatoria	42
2.5.1. Introducción	42
2.5.2. Tipos de modelos.....	43
2.5.3. El modelo de función aleatoria	45
2.6. Técnicas de estimación.....	50
2.6.1. Introducción	50
2.6.2. Técnicas de estimación de una sola variable	52
2.6.3. Técnicas de estimación de una variable incorporando información secundaria	68
2.6.4. Evaluación de la incertidumbre local.....	76
2.7. Las técnicas de simulación.....	92
2.7.1. Estimación versus simulación.....	93
2.7.2. Simulación secuencial.....	94
2.8. Ejercicio ejemplo de aplicación.....	100

2.8.1. Análisis de los datos.....	101
2.8.2. Análisis estructural	105
2.8.3. Krigeado	111
2.8.4. Cokrigeado	123
2.8.5. Simulación gaussiana secuencial	124
3. La elaboración de mapas de riesgo.....	129
4.Aplicación práctica a un vertedero real	131
4.1. Introducción.....	131
4.2. Estudio geológico de la zona de ampliación del vertedero	131
4.3. Estudio geoestadístico.....	135
4.3.1. Análisis Univariado	135
4.3.2. Análisis bivariado.....	138
4.3.3. Análisis estructural	140
4.3.4. Krigeado	145
4.3.5. Simulación gaussiana secuencial	148
5.Conclusiones.....	155
6.Líneas de futura investigación	157
7.Referencias bibliográficas.....	159
ANEJO I. Planos.....	161

Índice de figuras

Fig. 1. Histograma.....	14
Fig. 2. Histograma de frecuencias acumuladas	14
Fig. 3. Gráfico de probabilidad normal	15
Fig. 4. Gráfico de probabilidad lognormal	15
Fig. 5. Mediana del conjunto de datos.....	16
Fig. 6. Mínimo y máximo del conjunto de datos.....	17
Fig. 7. Cuartiles del conjunto de datos.....	17
Fig. 8. Diagrama de cuantiles variables u-v.	20
Fig. 9. Diagrama de dispersión variables u-v.....	20
Fig. 10. Opciones de resultado de la covarianza.	21
Fig. 11. Posibles valores a tomar por el coeficiente de correlación.....	22
Fig. 12. Sensibilidad del coeficiente de correlación a valores extremos.....	22
Fig. 13. Momento de inercia.....	23
Fig. 14. Mapa de localización de los datos (a) y mapa de isolíneas (b).....	24
Fig. 15. Mapa de símbolos.....	24
Fig. 16. Mapas indicadores.	25
Fig. 17. Ejemplo de continuidad espacial (a) y ruido blanco (b).....	26
Fig. 18. Nomenclatura vector.	27
Fig. 19. Diagrama de dispersión tipo h para distintos vectores de separación h.....	27
Fig. 20. Correlograma. Función de covarianza y variograma del ejemplo.	29
Fig. 21. Diagramas cruzados de dispersión tipo h para distintos vectores de separación h	31
Fig. 22. Correlograma. Función de covarianza y variograma del ejemplo.	32
Fig. 23. Tolerancias en la búsqueda de datos irregularmente espaciados.....	34
Fig. 24. Variograma experimental (a) y variograma experimental con modelo ajustado (b)	35
Fig. 25. Terminología del variograma experimental con el modelo ajustado.....	35
Fig. 26. Variograma omnidireccional (h=5).....	37
Fig. 27. Variograma omnidireccional (h=10).....	37
Fig. 28. Variogramas direccionales y sus correspondientes alcances.....	38
Fig. 29. Elipse de anisotropía.....	39
Fig. 30. Modelos posibles a ajustar en el variograma.....	40

Fig.31. Estructura imbricada.....	40
Fig.32. Modelo de anisotropía geométrica.....	41
Fig.33. Modelo de anisotropía zonal.....	41
Fig.34. Modelo de combinación de anisotropía geométrica y zonal.....	42
Fig.35. Ejemplo datos modelo función aleatoria.....	43
Fig.36. Ejemplo modelo determinístico I.....	44
Fig.37. Ejemplo modelo determinístico II.....	44
Fig.38. Ejemplo modelo probabilístico.....	44
Fig.39. Dominio de trabajo.....	45
Fig.40. Dominio de trabajo discretizado.....	45
Fig.41. Obtención de variables aleatorias a través de las funciones de distribución de probabilidad.....	46
Fig.42. Diferentes realizaciones de la misma función aleatoria.....	46
Fig.43. Datos disponibles cálculo de la covarianza.....	48
Fig.44. Datos disponibles tras decisión estacionariedad para cálculo de covarianza.....	49
Fig.45. Relación covarianza y variograma.....	50
Fig.46. Caso inicial del ejemplo de estimación de un punto.....	60
Fig.47. Efecto de aumentar la meseta.....	61
Fig.48. Efecto del aumento del efecto pepita.....	61
Fig.49. Efecto del efecto pepita puro.....	62
Fig.50. Efecto de disminuir el alcance.....	62
Fig. 51. Efecto de la forma.....	63
Fig. 52. Efecto de la anisotropía.	63
Fig. 53. Datos iniciales del ejemplo de la estimación de un campo completo.....	64
Fig. 54. Caso inicial del ejemplo.	64
Fig.55. Efecto del efecto pepita puro.....	64
Fig. 56. Efecto del alcance.	65
Fig. 57. Efecto de la anisotropía.	65
Fig. 58. Ejercicio ejemplo de estimación. Conjunto de datos.	66
Fig. 59. Ejercicio ejemplo de estimación. Conjunto de datos sobre malla regular.	66
Fig. 60. Ejercicio ejemplo de estimación. Datos disponibles de porosidad (%).....	72
Fig. 61. Datos ejemplo cokrigado ordinario.....	76
Fig.62. Valor estimado (a) y función de distribución de probabilidad(b)	76

Fig.63. Transformada de la función de distribución acumulada.....	79
Fig.64. Datos del ejemplo del Kriging Multigausiano. Información primaria (a) e información secundaria (b).....	81
Fig.65. Estimaciones por Kriging simple. Media.	82
Fig.66. Estimaciones por Kriging simple. Desviación estándar.	82
Fig.67. Cuantiles (0, 0.25, 0.5, 0.75, 1) obtenidos mediante KS.	83
Fig.68. Probabilidades de que el fenómeno se encuentre por debajo de un valor determinado obtenidos mediante KS.....	84
Fig.69. Estimaciones por coKriging colocalizado. Media.	84
Fig.70. Estimaciones por coKriging colocalizado. Desviación estándar.	84
Fig.71. Cuantiles obtenidos mediante cokriging colocalizado.	85
Fig. 72. Probabilidades de que el fenómeno se encuentre por debajo de un valor determinado obtenidos mediante cokriging colocalizado.	85
Fig.73. Datos iniciales ejemplo explicativo de estimación de proporción de valores por debajo en un umbral.....	86
Fig.74. Mapa de indicadores del ejemplo.	87
Fig.75. Valores medios de Indicadores para diferentes umbrales.....	87
Fig.76. Datos del ejemplo de kriging indicador.	88
Fig.77. Probabilidad de que la porosidad sea menor del 5%.....	89
Fig.78. Probabilidad de que la porosidad sea menor del 10%.....	89
Fig.79. Estimación versus simulación.....	93
Fig. 80. Estimación versus simulación. Suavizado de la estimación.....	94
Fig.81. Paso A. Selección localización.	95
Fig.82. Paso B. Construcción función local de distribución de probabilidades condicionada.	95
Fig.83. Paso C. Obtención del valor de la variable.....	96
Fig.84. Paso D. Asignar el valor al nodo simulado.	96
Fig.85. Mapa de probabilidades (a), mapa de cuantiles (b) y mapa de dispersión (varianza) (c).....	99
Fig.86. Ejemplo aplicación Datos porosidad disponibles.	101
Fig.87. Datos permeabilidad disponibles.	102
Fig.88. Histograma permeabilidad.	103
Fig.89. Histograma porosidad.....	103
Fig.90. Ejemplo aplicación Diagrama de dispersión.	104
Fig.91. Ejemplo aplicación Diagrama de cuantiles.....	105
Fig.92. Ejemplo aplicación Variograma omnidireccional.	107
Fig.93. Ejemplo aplicación Variograma omnidireccional ajustado.	107

Fig.94. Ejemplo aplicación Variogramas direccionales.....	109
Fig.95. Ejemplo aplicación. Elipsoide de anisotropía.	110
Fig.96. Ejemplo aplicación. Elipsoide de anisotropía. Variograma direccional máxima continuidad ajustada.....	110
Fig.97. Ejemplo aplicación. Variograma direccional mínima continuidad ajustada.....	111
Fig.98. Ejemplo aplicación. Campo permeabilidad mediante Krigeado ordinario.....	112
Fig.99. Ejemplo aplicación. Varianza campo de permeabilidad estimado mediante Krigeado ordinario.....	112
Fig.100. Ejemplo aplicación Efecto. pepita en la varianza del campo de permeabilidad estimado.....	114
Fig.101. Ejemplo aplicación. Efecto pepita puro en el campo de permeabilidades estimado.....	115
Fig.102. Ejemplo aplicación. Efecto pepita puro en la varianza del campo de permeabilidad estimado.....	115
Fig.103. Ejemplo aplicación. Efecto de doblar el variograma en la estimación del campo de permeabilidades.....	116
Fig.104. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al doblar el variograma.....	118
Fig.105. Ejemplo aplicación. Modelo esférico.....	118
Fig.106. Ejemplo aplicación. Modelo exponencial.	118
Fig.107. Ejemplo aplicación. Modelo gaussiano.	117
Fig.108. Ejemplo aplicación. Efecto de doblar el alcance en la estimación del campo de permeabilidades.....	119
Fig.109. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al doblar el alcance.....	120
Fig.110. Ejemplo aplicación. Efecto de igualar el alcance de máxima continuidad con el de mínima continuidad en la estimación del campo de permeabilidades.	121
Fig.111. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al igualar el alcance de máxima continuidad con el de mínima continuidad.	121
Fig.112. Ejemplo aplicación. Efecto de invertir los alcances de las direcciones de máxima y mínima continuidad en la estimación del campo de permeabilidades.	122
Fig.113. Ejemplo aplicación. Campo permeabilidades por krigeado ordinario mediante 16 datos permeabilidad....	123
Fig.114. Ejemplo aplicación. Campo permeabilidades por cokrigeado mediante 16 datos permeabilidad y 64 porosidad.	124
Fig.115. Ejemplo aplicación Simulaciones obtenidas del campo de permeabilidades (6/100).....	125
Fig.116. Ejemplo aplicación. Valor esperado estimaciones.	126
Fig.117. Ejemplo aplicación. Probabilidad permeabilidad tenga un valor menor de 45 (a), 50 (b), 55 (c).....	126
Fig.118. Ejemplo aplicación. Cuantiles correspondientes a las probabilidades 0.2(a), 0.46(b) y 0.95(c).....	127
Fig.119. Ejemplo aplicación. Varianzas condicionales.....	127
Fig.120. Ejemplo aplicación. Señalización de sondeos no aptos según normativa y según las recomendaciones.....	134
Fig. 121. Ejercicio Aplicación. Datos espesores disponibles.	134
Fig. 122. Ejercicio Aplicación. Datos permeabilidad disponibles.	135
Fig. 123. Ejercicio Aplicación. Histograma variable espesor.....	136

Fig.124. Ejercicio Aplicación. Histograma variable espesor. Diagrama normal.....	137
Fig. 125. Ejercicio Aplicación. Histograma variable espesor. Diagrama lognormal.....	137
Fig. 126. Ejercicio Aplicación. Diagrama de dispersión permeabilidad-espesor. Valores permeabilidad (a)(10-8-10-9), (b)(10-10-10-11) y (10-11-10-12).	139
Fig. 127. Ejercicio Aplicación. Diagrama de cuantiles.	140
Fig.128. Ejercicio Aplicación. Variograma omnidireccional variable espesor.....	141
Fig.129. Ejercicio Aplicación. Variograma omnidireccional ajustado.	142
Fig. 130. Ejercicio Aplicación. Variogramas direccionales.	143
Fig. 131. Ejercicio Aplicación. Elipsoide de anisotropía.....	144
Fig. 132. Ejercicio Aplicación. Variograma direccional mínima continuidad ajustada.	144
Fig. 133.Ejercicio Aplicación. Variograma direccional máxima continuidad ajustada.....	145
Fig. 134.Ejercicio Aplicación. Campo permeabilidad estimado mediante Krigeado ordinario.....	146
Fig. 135. Ejercicio Aplicación. Varianza campo de permeabilidad estimado mediante Krigeado ordinario.....	146
Fig. 136. Ejercicio Aplicación. Probabilidades estimadas campo espesor sea superior a 1 m.....	147
Fig. 137. Ejercicio Aplicación. Probabilidades estimadas campo espesor sea superior a 10 m.....	147
Fig. 138. Ejercicio Aplicación. Mapas de riesgo de superación de 1 y 10m respectivamente.....	148
Fig. 139. Ejercicio Aplicación. Función de probabilidad acumulada de la variable espesor.....	149
Fig.140. Ejercicio Aplicación. Simulaciones campo de espesores (6/200).....	149
Fig.141. Ejercicio Aplicación. Probabilidad espesores estén por encima del umbral 1m.....	150
Fig.142.Ejercicio Aplicación. Probabilidad espesores estén por encima del umbral 10m.....	151
Fig.143. Propuesta delimitación de ubicación del vertedero (Sobre mapa riesgo espesor>10m.). Zona de probabilidad menor (círculo verde).	152
Fig.144. Propuesta final delimitación ubicación del vertedero.....	152
Fig.145. Ejercicio Aplicación. Varinzas condicionales.....	153
Fig. 146.Cuantiles para una probabilidad acumulada de 0.1 0.25 0.5 0.75 0.85 0.99.....	153

Índice de tablas

<i>Tabla. 1. Tabla de frecuencias.....</i>	<i>14</i>
<i>Tabla.2. Tabla de frecuencias acumuladas.....</i>	<i>14</i>
<i>Tabla 3. Coeficiente de correlación, covarianza y momento de inercia obtenido del ejemplo.....</i>	<i>28</i>
<i>Tabla 4. Coeficiente de correlación, covarianza y momento de inercia obtenido del ejemplo.....</i>	<i>32</i>
<i>Tabla 5. Número de parejas, distancia separación y variograma.....</i>	<i>37</i>
<i>Tabla 6. Datos del ejemplo para la obtención de la transformada normal.....</i>	<i>79</i>
<i>Tabla 7. Datos del ejemplo krigeado multigaussiano (Mínimo, media, desviación estándar y máximo).....</i>	<i>81</i>
<i>Tabla 8. Numero de intervalos, vector separación y tolerancia óptimos variograma omnidireccional.....</i>	<i>106</i>
<i>Tabla 9. Direcciones tomadas y alcances obtenidos.</i>	<i>108</i>
<i>Tabla.10. Ejercicio aplicación. Datos de los sondeos.</i>	<i>133</i>
<i>Tabla 11. Ejercicio Aplicación. Valores óptimos variograma omnidireccional.....</i>	<i>141</i>
<i>Tabla.12. Ejercicio Aplicación. Direcciones y alcances obtenidos variogramas direccionales.....</i>	<i>142</i>

1

Introducción

Edward H. Isaaks (1989) y R. Mohan Srivastava (1989) definen que las herramientas estadísticas son útiles para obtener datos cualitativos de una amplia variedad de fenómenos naturales y para obtener respuestas cuantitativas de preguntas específicas. Desgraciadamente, la estadística no hace uso de la información espacial de la tierra. Es por ello que la geoestadística nos proveerá de una serie de herramientas para modelar sobre fenómenos que se suelen caracterizar a partir de la medida de parámetros que se encuentran espacialmente distribuidos, es decir, que además de tener en cuenta el valor de la variable registrada, tiene en cuenta la posición que tiene esa variable en el espacio.

1.1. Antecedentes

La geoestadística nació en la industria minera, concretamente en Sudáfrica, país con de multitud de minas de oro y diamantes. Fue Daniel Krige (1951) el que se encargaba de estudiar la posibilidad de explotar las minas de estos minerales a partir de la estimación de su concentración media, disponiendo únicamente del registro de una pequeña cantidad de puntos. Con ello tenía que ser capaz de obtener un valor de dicha concentración que fuese representativo de grandes volúmenes de roca, de manera que se comparase ese valor con unos umbrales establecidos que indicaban si la explotación de la mina sería rentable o no. Krige estableció que la medida que debía obtener debía estar basada en esas medidas parciales, las cuales no estaban aisladas entre ellas, pero no terminaba de ver dicha relación como para poder construir ese estimador sobre un gran volumen de roca. En 1962, Georges Matheron, creador de los pilares teóricos de la geoestadística, consiguió establecer dicha relación.

Como ya se ha indicado, fue la industria minera el primer ámbito donde se desarrollaron las primeras herramientas de la geoestadística. Una vez que estas técnicas empezaron a funcionar adecuadamente, se aplicaron en el ámbito de la ingeniería del petróleo, en el cual se necesita caracterizar la geología para describir la trampa en la que se encuentra el recurso. Por tanto, interesa conocer las profundidades y espesores de los estratos geológicos, y para ello, se perfora el terreno, se obtienen medidas de los materiales y se reconstruye el sólido en el que se encuentra atrapado el petróleo, para poder así explotarlo adecuadamente.

Estos dos ámbitos, la minería y la industria del petróleo, fueron los pilares de la historia de la geoestadística. Más adelante la hidrogeología recurrió a estas técnicas, trabajando no tanto con espesores y profundidades, sino con los parámetros que caracterizan los acuíferos: la

conductividad hidráulica, la transmisividad, el coeficiente de almacenamiento, la porosidad, la dispersión de un contaminante, etc., es decir, aquellos parámetros que describen el flujo de agua y el transporte de masa en un medio poroso saturado o no saturado. Por ello, se crearon herramientas nuevas para atender ese tipo de problemas.

Actualmente, la geoestadística se asienta sobre estas tres disciplinas. Pero como se trata de una caja de herramientas, que no está ligada a una clase de problemas concretos, se comenzó a aplicar en otros ámbitos, como por ejemplo para cartografiar la precipitación, la migración de isótopos radiactivos,....

En resumen, los problemas en los que influyen variables distribuidas en el espacio son susceptibles de ser modelados con estas herramientas geoestadísticas.

1.2. Aplicaciones potenciales de la geoestadística para la caracterización de emplazamientos de vertederos de residuos

El Real Decreto 1481/2001, de 27 de diciembre, por el que se regula la eliminación de residuos mediante depósito en vertedero, establece en el Anexo I “Requisitos generales para todas las clases de vertederos”, que todo vertedero deberá estar ubicado y diseñado de forma que cumpla las condiciones necesarias para impedir la contaminación del suelo, de las aguas subterráneas o de las aguas superficiales y garantizar la recogida eficaz de los lixiviados.

La protección del suelo, de las aguas subterráneas y de las aguas de superficie se realizará mediante la combinación de una barrera geológica y un revestimiento inferior durante la fase activa o de explotación, y mediante la combinación de una barrera geológica y un revestimiento superior durante la fase pasiva o posterior a la clausura.

Según el Real Decreto, existe una barrera geológica cuando las condiciones geológicas e hidrogeológicas subyacentes y en las inmediaciones de un vertedero tienen la capacidad de atenuación para impedir un riesgo potencial para el suelo y las aguas subterráneas.

Por tanto, la base y los lados del vertedero deberán disponer de una capa mineral (material arcilloso) que cumpla unos requisitos de permeabilidad y espesor cuyo efecto combinado en materia de protección del suelo, de las aguas subterráneas y de las aguas superficiales sea por lo menos equivalente al derivado de los requisitos siguientes:

- Vertederos para residuos peligrosos: $K \leq 1,0 * 10^{-9}$ m/s; espesor ≥ 5 m
- Vertederos para residuos no peligrosos: $K \leq 1,0 * 10^{-9}$ m/s; espesor ≥ 1 m
- Vertederos para residuos inertes: $K \leq 1,0 * 10^{-7}$ m/s; espesor ≥ 1 m

Es en este punto donde la geoestadística entra en juego. A partir de un número determinado de sondeos, que proporcionan datos puntuales de permeabilidad y espesores de los estratos, aplicando distintas técnicas geoestadísticas y con ayuda del programa informático SGeMS se consigue:

- Determinar el espesor de estratos geológicos de interés
- Interpretar los resultados de campañas de prospección geofísica para caracterizar vertederos existentes

Con los resultados obtenidos se podrán realizar mapas de probabilidad donde el valor del espesor de la capa de arcilla no sea inferior al valor mínimo y así poder delimitar una zona donde dichas condiciones sean óptimas para la ubicación de un vertedero determinado.

1.3. Objetivos y estructura

El objetivo de este Trabajo Fin de Máster es la elaboración de mapas de riesgo a partir de técnicas geoestadísticas que permitan evaluar y delimitar la idoneidad del emplazamiento de un vertedero en una zona determinada.

El presente trabajo se estructura en los siguientes apartados:

- En el capítulo 1 contiene la introducción, con los antecedentes y las aplicaciones potenciales de la geoestadística para la caracterización de emplazamientos de vertederos de residuos.
- En el capítulo 2 se analizan los conceptos básicos de geoestadística, desde la estadística descriptiva hasta los distintos procesos de estimación y simulación, finalizando con un caso práctico donde se puede contemplar todo lo visto a lo largo del capítulo.
- En el capítulo 3 se presenta la elaboración de mapas de riesgo.
- En el capítulo 4 se analiza el caso práctico de aplicación a un vertedero sintético localizado en la provincia de Valencia.
- En el capítulo 5 se recogen las conclusiones fundamentales.
- En el capítulo 6 se comentan las líneas de futura investigación.
- En el capítulo 7 se incluyen las referencias bibliográficas utilizadas.

2

Conceptos básicos de geoestadística

2.1. Introducción

Según E. Cassiraga (2015), la estadística es la parte de las matemáticas que sirve para recoger datos, agruparlos, ordenarlos, realizar una serie de cálculos, representarlos gráficamente y establecer una serie de conclusiones o realizar posibles predicciones.

La estadística, además, está dividida en dos grandes áreas: la estadística descriptiva y la inferencial.

- La estadística descriptiva se dedica a la descripción, visualización y resumen de datos originados a partir de los fenómenos de estudio.
- La estadística inferencial, sin embargo, se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para modelo patrones en los datos y extraer inferencias acerca de la población bajo estudio.

En este estudio, E. Cassiraga (2015) utiliza ante todo la estadística descriptiva para poder realizar la descripción espacial de las variables a estudiar.

Dentro de la estadística descriptiva se pueden encontrar distintas tipologías de descripciones, que se pueden agrupar según:

- El número de variables a describir:
 - Descripción univariada
 - Descripción bivariada
 - ...
 - Descripción multivariada
- La descripción espacial

Además, para la realización de la descripción de dichas variables se utilizan distintas técnicas, ya sean métodos gráficos y/o métodos numéricos.

A continuación, se detalla tanto la descripción univariada como bivariada y sus diferentes herramientas de análisis.

2.2. Descripción univariada

2.2.1. Introducción

La descripción univariada, según Spiegel, Murray R. y Stephens, Larry J. (2012), incluye todas las técnicas que hacen referencia a la descripción de una sola variable.

2.2.2. Métodos gráficos para la descripción univariada

2.2.2.1. Tablas de frecuencia e histogramas

La manera más fácil y útil de representar el conjunto de datos es a través de la tabla de frecuencia (Tabla 1), que muestra cómo de frecuente caen ciertos valores en un intervalo o clase, y su correspondiente histograma (Fig. 1)

Clase		Número de datos
0 ≤ V	< 10	1
10 ≤ V	< 20	1
20 ≤ V	< 30	0
30 ≤ V	< 40	0
40 ≤ V	< 50	3
50 ≤ V	< 60	2
60 ≤ V	< 70	2
70 ≤ V	< 80	13
80 ≤ V	< 90	16
90 ≤ V	< 100	11
100 ≤ V	< 110	13
110 ≤ V	< 120	17
120 ≤ V	< 130	13
130 ≤ V	< 140	4
140 ≤ V	< ∞	4

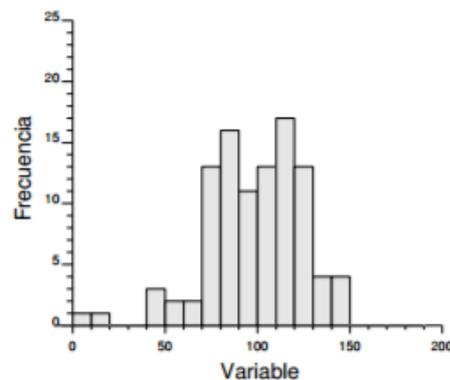


Tabla 1. Tabla de frecuencias
(Fuente: Isaaks.E, Srivastaba.R, 1989)

Fig. 1. Histograma
(Fuente: Isaaks.E, Srivastaba.R, 1989)

2.2.2.2. Tablas de frecuencias acumuladas e histogramas

Sin embargo, la mayoría de los textos estadísticos organizan los datos de forma ascendente con frecuencia acumulada (Tabla 2 y Fig.2, respectivamente).

Clase	Número de datos	Número de datos acumulados
V < 10	1	1
V < 20	2	2
V < 30	2	2
V < 40	2	2
V < 50	5	5
V < 60	7	7
V < 70	9	9
V < 80	22	22
V < 90	38	38
V < 100	49	49
V < 110	62	62
V < 120	79	79
V < 130	92	92
V < 140	96	96
V < ∞	100	100

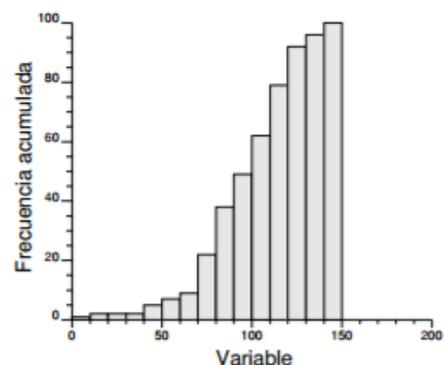


Tabla 2. Tabla de frecuencias acumuladas
(Fuente: Isaaks.E, Srivastaba.R, 1989)

Fig. 2. Histograma de frecuencias acumuladas
(Fuente: Isaaks.E, Srivastaba.R, 1989)

2.2.2.3. Gráficos de probabilidad normal y lognormal

Aunque los métodos anteriores son ampliamente usados y válidos, la mayoría de los datos responden mejor si la distribución de los valores está próxima a una distribución normal.

Un gráfico de probabilidad normal (Fig. 3) permite comparar la distribución empírica de una muestra de datos con la distribución normal.

Los n valores son ordenados de menor a mayor y cada valor es representado gráficamente frente a la proporción de datos que son menores que él. La idea básica consiste en comparar dicha representación gráfica con la que se obtendría en una distribución normal, donde los valores seguirían una línea recta. Si la distribución de valores queda cerca de una línea recta, éstos siguen una distribución normal.

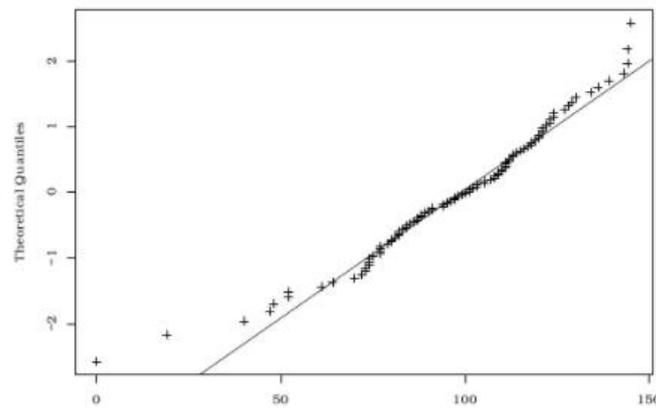


Fig. 3. Gráfico de probabilidad normal
(Fuente: Sanz, 2005)

En ocasiones, esta distribución no es válida porque los datos son muy dispares y por lo tanto no siguen una disposición rectilínea. Para estos casos, se suele comprobar si los datos siguen una distribución lognormal.

Se considera que una variable está lognormalmente distribuida si la distribución de sus logaritmos es normal. En este caso, el eje x está en escala logarítmica, de tal modo que, si la variable está lognormalmente distribuida, las frecuencias acumuladas caen sobre una línea recta (Fig. 4).

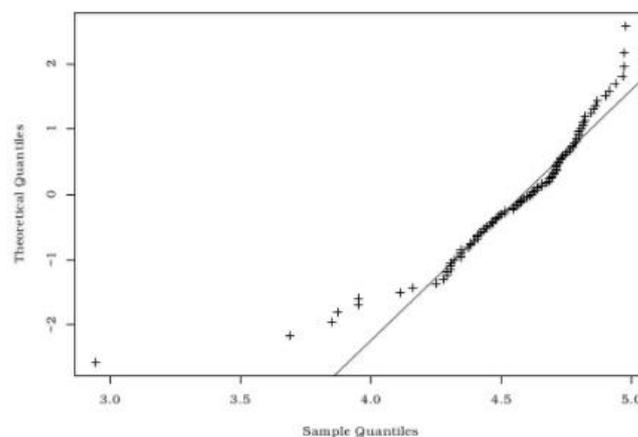


Fig. 4. Gráfico de probabilidad lognormal.
(Fuente: Sanz, 2005)

2.2.3. Métodos numéricos para la descripción univariada

2.2.3.1. Medidas de localización

Spiegel, Murray R. y Stephens, Larry J. (2012) indican que las medidas de localización son aquellas que informan acerca de donde están localizadas las distintas partes de una distribución, como puede ser el centro.

Como medidas de localización del centro se tiene:

- Media

La media aritmética o media m_z de un conjunto de n datos z_1, z_2, \dots, z_n es:

$$m_z = \frac{1}{n} \sum_{i=1}^n z_i$$

Es la medida más utilizada tradicionalmente en las ciencias porque es muy eficiente.

Capta la variación de cada una de las puntuaciones. No obstante, la media no es un buen estadístico de tendencia central cuando en la distribución hay valores extremos o cuando es muy asimétrica.

- Mediana

La mediana M de un conjunto de datos z_1, z_2, \dots, z_n es la puntuación central de una serie de datos ordenados (Fig. 5). Es la puntuación que deja por debajo el 50% de los datos y también por arriba. Es el valor central si n es impar y es la media de los dos valores centrales si n es par, esto es:

$$M = \begin{cases} \frac{z_{n+1}}{2}, & \text{si } n \text{ es impar} \\ \frac{\left(\frac{z_n}{2} + \frac{z_{n+1}}{2}\right)}{2}, & \text{si } n \text{ es par} \end{cases}$$

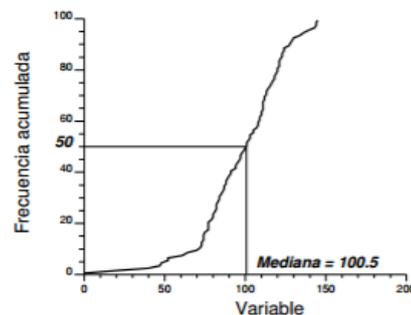


Fig. 5. Mediana del conjunto de datos.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

Es una buena medida de tendencia central cuando tenemos puntuaciones extremas. Además, es un buen estadístico cuando la distribución es muy asimétrica porque no se ve tan afectada como la media por valores extremos.

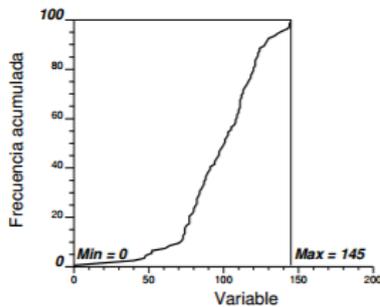
- Moda

La moda de un conjunto de datos es el valor que ocurre más frecuentemente, es decir, el valor más común. Esta puede o no existir, incluso puede no ser única.

Si la distribución de frecuencias de un conjunto de datos es simétrica, la media, la mediana y la moda coinciden.

Además de las medidas de localización del centro, hay otras que cabe destacar, como son:

- Mínimo y máximo



El mínimo es el valor de la variable para una frecuencia acumulada igual a 0, mientras que el máximo, es el valor de la variable para una frecuencia acumulada igual a 1 o 100% (Fig. 6)

Fig. 6. Mínimo y máximo del conjunto de datos.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

- Cuartiles

Son las puntuaciones que dividen a una distribución en cuatro partes iguales (Fig. 7):

Q1 = Es el que deja por debajo de sí una cuarta parte de las puntuaciones, y por encima las tres cuartas partes restantes. Es decir, deja por debajo el 25% y por encima el 75% de las observaciones.

Q2 = Se encuentra en el punto medio, dejando por debajo el 50% de los datos y por encima el otro 50%.

Q3 = Deja el 75% de las observaciones por debajo, y el 25% por encima.

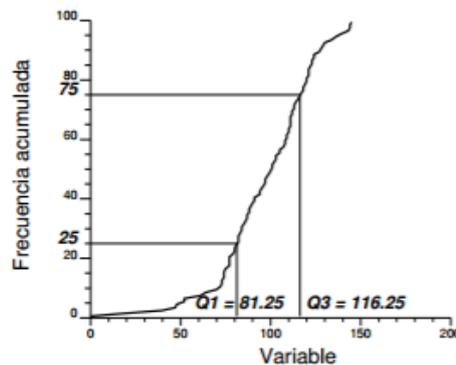


Fig. 7. Cuartiles del conjunto de datos.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

- Cuantiles, déciles y percentiles

Un cuantil es el valor q de la variable para la cual la frecuencia acumulada es igual a p .

Si p se incrementa décimos, hablamos de déciles, mientras que, si se incrementa centésimos, hablamos de percentiles.

2.2.3.2. Medidas de dispersión

Las medidas de dispersión describen la variabilidad de los valores de los datos, generalmente con respecto al centro de la distribución.

Las principales medidas de dispersión son:

- El rango

El rango de un conjunto de números es la diferencia entre el número mayor y número menor del conjunto

- Rango intercuartil

Es la diferencia entre el cuartil superior e inferior, entre el primer y el tercer cuartil.

Al contrario que la varianza y la desviación estándar, que a continuación se muestra, no usa la media como el centro de distribución.

- Varianza

La varianza de un conjunto n datos z_1, z_2, \dots, z_n es la media de la diferencia al cuadrado de los valores observados respecto a su media.

$$\sigma_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - m_z)^2$$

- Desviación típica

La desviación típica es la raíz cuadrada de la varianza

$$\sigma_z = \left(\frac{1}{n} \sum_{i=1}^n (z_i - m_z)^2 \right)^{1/2}$$

2.2.3.3. Medidas de forma

Las medidas de forma, sin embargo, proveen información acerca de la forma de una distribución (simetría, longitud de colas, picos)

Entre las medidas de forma cabe destacar:

- Coficiente del sesgo

Valora la asimetría de una distribución y se define como:

$$CS = \frac{\frac{1}{n} \sum_{i=1}^n (z_i - m_z)^3}{\sigma^3}$$

- Si $CS > 0 \rightarrow$ cola hacia la derecha y se dice que es sesgada positiva
- Si $CS < 0 \rightarrow$ cola hacia la izquierda y se dice que es sesgada negativa

Es muy sensible a los cambios de valores ya que la diferencia de cada valor y su media están elevadas al cubo.

- Coeficiente de curtosis

Valora que tan puntiaguda es una distribución y se define como:

$$CC = \frac{1}{n} \sum_{i=1}^n (z_i - m_z)^4$$

Cuando los valores de una distribución están concentrados cerca de la media, ésta presenta un pico grande, mientras que si la forma es plana significa que los valores se dispersan más alrededor de la media.

Por lo general, se compara una distribución con la curva normal que tiene una curtosis igual a 3 de manera que:

Si $CC < 3$ la distribución presenta un pico menos agudo que la distribución normal.

Si $CC > 3$ la distribución presenta un pico más agudo que la distribución normal.

2.3. Descripción bivariada

2.3.1. Introducción

Spiegel, Murray R. y Stephens, Larry J. (2012) indican que las herramientas para la descripción univariada describen la distribución de variables de forma individual, pero si se dispone de un conjunto con más de una variable la información obtenida es limitada. Alguna de las características más interesantes de las ciencias de la tierra son relaciones entre varios atributos.

2.3.2. Métodos gráficos para la descripción bivariada

2.3.2.1. Tablas de frecuencia e histogramas

En el análisis del conjunto de datos de las ciencias de la tierra frecuentemente se quieren comparar dos distribuciones. La presentación de sus histogramas junto con sus tablas de frecuencia revelará las diferencias entre ellas. Desafortunadamente, si las dos distribuciones son muy similares, este método de comparación no nos ayudará a destapar las sutiles diferencias.

2.3.2.2. Diagrama de cuantiles

Para una buena comparación visual de dos distribuciones se usan los diagramas de cuantiles (Fig. 8). Este gráfico es comúnmente usado cuando se espera que las distribuciones sean similares.

Un diagrama de cuantiles es un gráfico x-y en el cual los cuantiles de dos distribuciones son representados uno frente al otro.

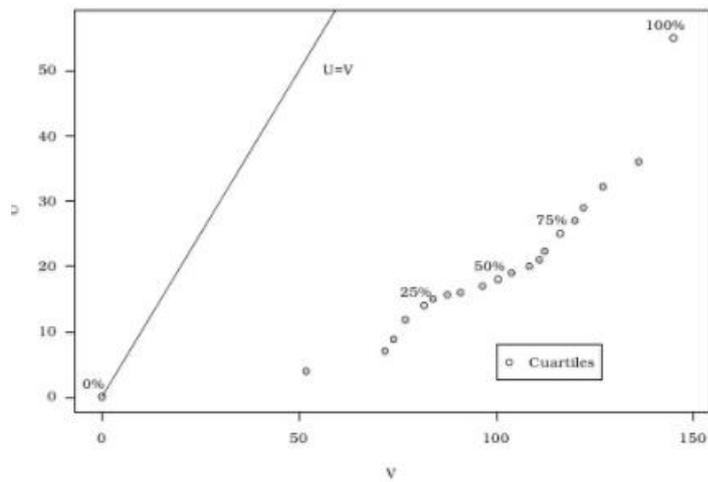


Fig. 8. Diagrama de cuantiles variables u-v. (Fuente: Sanz, 2005)

Si las dos distribuciones tienen el mismo número de datos, el cálculo de los cuantiles para cada distribución no es necesario, basta con ordenar en orden ascendente los correspondientes pares de valores y dibujarlos.

Si las dos distribuciones son idénticas, saldrá como diagrama de cuantiles una línea recta. Sin embargo, si las distribuciones son similares, las pequeñas desviaciones respecto a la línea $x=y$ revelan donde están las diferencias.

2.3.2.3. Diagrama de probabilidades acumuladas

Un diagrama de probabilidades acumuladas es un gráfico x-y en el cual las probabilidades acumuladas de dos distribuciones son representadas unas versus las otras.

2.3.2.4. Diagrama de dispersión

Un diagrama de dispersión (Fig. 9) es un gráfico x-y en el cual la coordenada x corresponde a los valores de una variable y la coordenada y a los de otra.

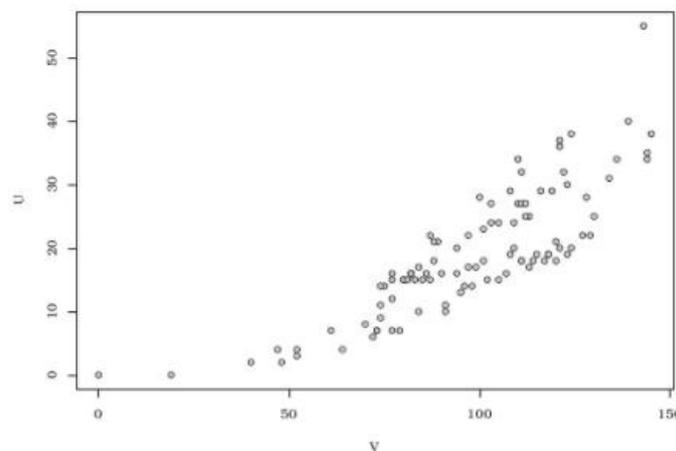


Fig.9. Diagrama de dispersión variables u-v. (Fuente: Sanz, 2005)

Un diagrama de dispersión es un buen indicador visual de la relación que existe entre dos variables y a la vez permite detectar valores que se encuentran fuera del rango de los datos, es decir, ayuda en la validación de los datos iniciales.

En el más amplio sentido, hay tres patrones que se pueden observar en un diagrama de dispersión

- Dos variables están positivamente correlacionadas si los mayores valores de una de las variables están asociadas a los mayores valores de la otra variable e igual con los menores.
Un ejemplo se aprecia en las rocas porosas donde la permeabilidad y la porosidad están positivamente correlacionadas
- Se dice que están negativamente correlacionadas si los valores mayores de una de las variables están asociados a los menores de la otra variable.
Por ejemplo, en los datos geológicos, la concentración de dos elementos. En una roca tipo caliza el incremento de la cantidad de calcio resulta en el decremento en la cantidad de magnesio.
- Las dos variables no están correlacionadas. Un incremento en una de las variables no afecta aparentemente en la otra.

2.3.3. Métodos numéricos para la descripción bivariada

2.3.3.1. Covarianza

La covarianza, según E. Cassiraga (2015), se define como:

$$C_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)$$

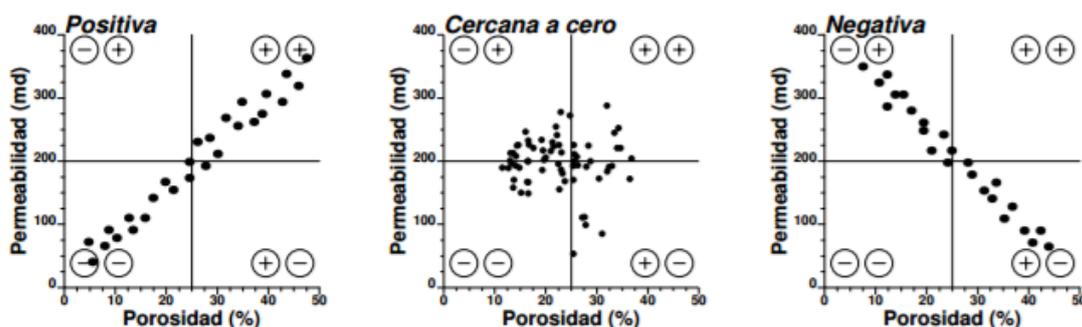


Fig. 10. Opciones de resultado de la covarianza. (Fuente: Cassiraga, 2015)

Generalmente, este estadístico se usa como resumen estadístico de un diagrama de dispersión (Fig. 10)

Además, la covarianza entre dos variables depende de la magnitud del valor de los datos. Por ejemplo, si multiplicamos todos los valores de los datos por 100, la covarianza será 100 veces mayor. Sin embargo, el diagrama de dispersión mantendría la misma apariencia.

2.2.3.2. Coeficiente de correlación

El coeficiente de correlación o de Pearson es el estadístico más comúnmente utilizado para analizar la correlación (lineal) entre dos variables y se define como:

$$\rho_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sigma_x \sigma_y}$$

Donde:

- n: número de datos
- x_1, \dots, x_n : valores de los datos de la primera variable
- y_1, \dots, y_n : valores de los datos de la segunda variable
- m_x : valor medio
- σ_x : desviación estándar

Entre las propiedades del coeficiente de correlación cabe destacar:

1. Siempre se encuentra entre los valores -1 y 1 (Fig. 11)

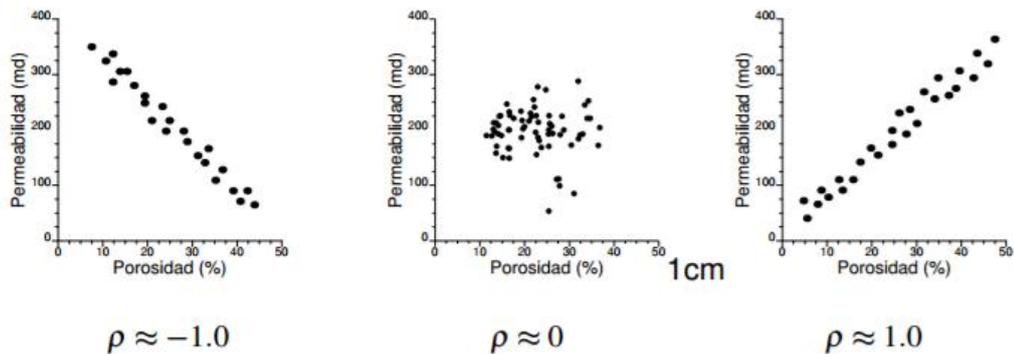


Fig.11. Posibles valores a tomar por el coeficiente de correlación. (Fuente: Cassiraga, 2015)

2. El coeficiente de correlación mide la dependencia lineal
3. Es sensible a valores fuera de la nube principal (Fig. 12)

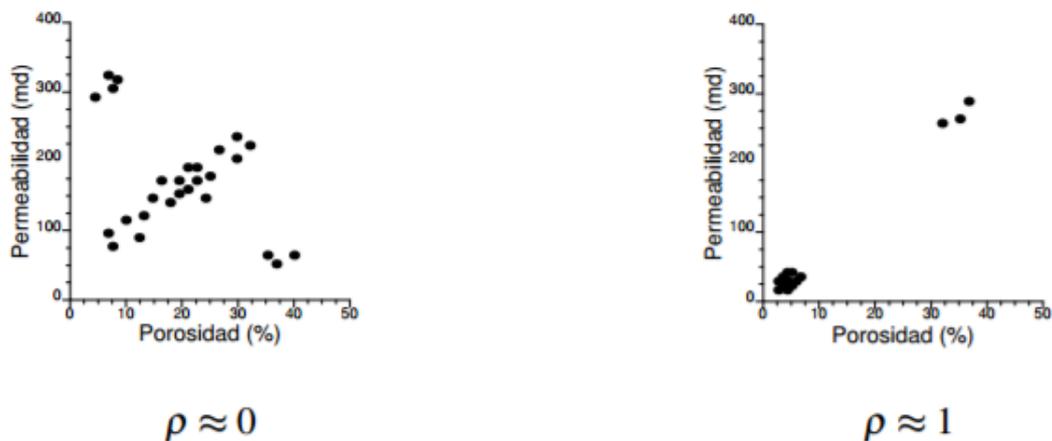


Fig. 12. Sensibilidad del coeficiente de correlación a valores extremos

2.3.3.3. El momento de inercia

El momento de inercia (Fig. 13) calcula la distancia al punto en el cual serían idénticos x e y , obteniendo una representación estadística de toda la nube de puntos en cuanto a su diferencia con respecto a la recta de igualdad $x=y$.

$$M^o = \frac{1}{N} \sum_{i=1}^n \left(\frac{x_i - y_i}{\sqrt{2}} \right)^2 = \frac{1}{2N} \sum_{i=1}^n (x_i - y_i)^2$$

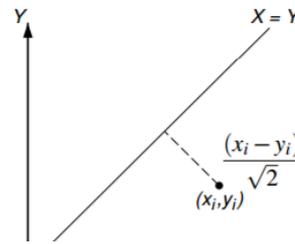


Fig. 13. Momento de inercia. (Fuente: Cassiraga, 2015)

Si el momento de inercia es cero, significa que hay igualdad total, correlación máxima entre las dos variables que se están analizando. Según va creciendo la dispersión alrededor de la nube el momento de inercia también crece.

2.4. Descripción espacial

2.4.1. Introducción

E. Cassiraga (2015) indica que si algo distingue a las bases de datos de las ciencias del terreno de otras es que, además de la medición, la información está asociada a una localización en el espacio. Hay características espaciales que pueden ser de gran interés para resolver los problemas a los que se hace frente día a día, como pueden ser la localización de los valores extremos, la tendencia general seguida por un parámetro, el grado de continuidad, etc. Todas estas características pueden servir de ayuda para una mejor comprensión del fenómeno modelado y para derivar información acerca de él.

El objetivo principal cuando se habla de analizar o describir espacialmente un parámetro es describir y cuantificar la relación entre las medidas que de él se ha tomado en distintas posiciones, o incluso se podrá extender el análisis a la comparativa de diferentes parámetros.

Todo este análisis se realiza estudiando los valores de estos atributos y comparándolos en dos posiciones simultáneamente, es decir, será un análisis bivariado, no siendo necesaria la aplicación de mayores órdenes. Por lo que se van a desarrollar una serie de técnicas capaces de obtener información acerca de la relación que existe entre las medidas de parámetros en dos posiciones en el espacio.

2.4.2. Visualización espacial de datos

La descripción espacial de la información comienza por una representación gráfica que tenga en cuenta la localización de los datos.

Hay varias formas de representar gráficamente la distribución espacial de un atributo:

- Mapas con localización de los datos y mapas de isolíneas.
- Mapas de símbolos
- Mapas de indicadores

2.4.2.1. Mapas con localización de los datos y mapas de isóneas

La manera más sencilla y completa de visualizar los datos es posicionar los parámetros de acuerdo a sus coordenadas sobre un mapa. Son los llamados mapas con localización de datos (Fig. 14a). A partir de él se pueden trazar isóneas, obteniendo un mapa de isóneas, que no es más que un mapa interpolado (Fig. 14b)

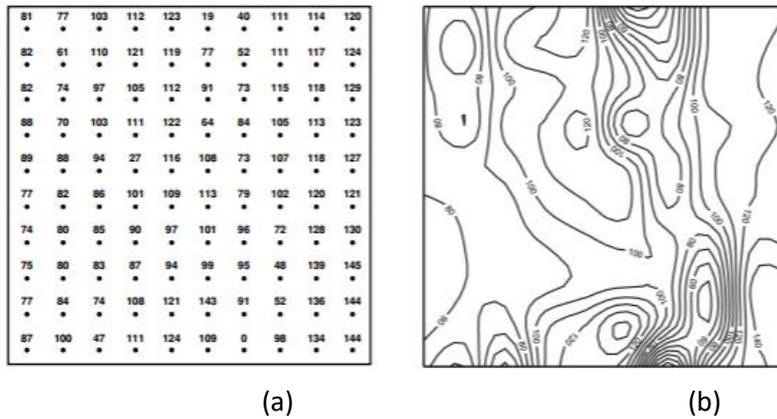


Fig. 14. Mapa de localización de los datos (a) y mapa de isóneas (b)
(Fuente: Isaaks.E, Srivastaba.R, 1989)

2.4.2.2. Mapas de símbolos

Otra alternativa son los mapas de símbolos (Fig. 15), que generalmente son los más usados en geoestadística. En estos mapas se asigna a cada píxel un color de acuerdo a una escala, la cual tiene que ver con el valor de la variable.

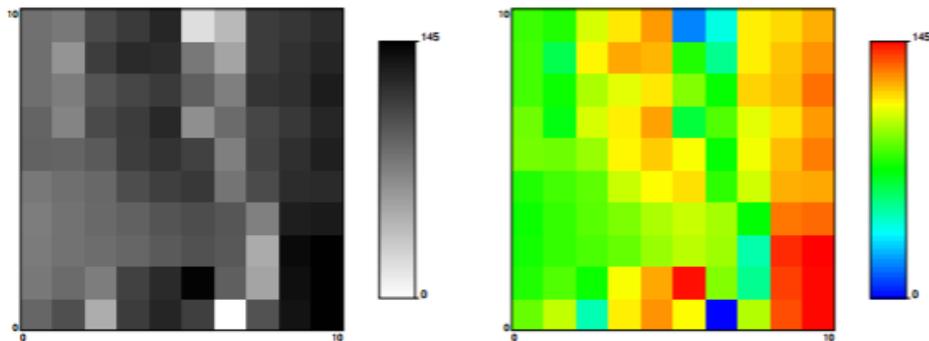


Fig. 15. Mapa de símbolos. (Fuente: Isaaks.E, Srivastaba.R, 1989)

2.4.2.3. Mapas indicadores

Por último, también se podría trabajar con mapas de indicadores (Fig. 16). Los indicadores son una variable binaria que surge de establecer una serie de umbrales y de dar a cada variable un valor si está por encima de dicho umbral u otro valor si está por debajo.

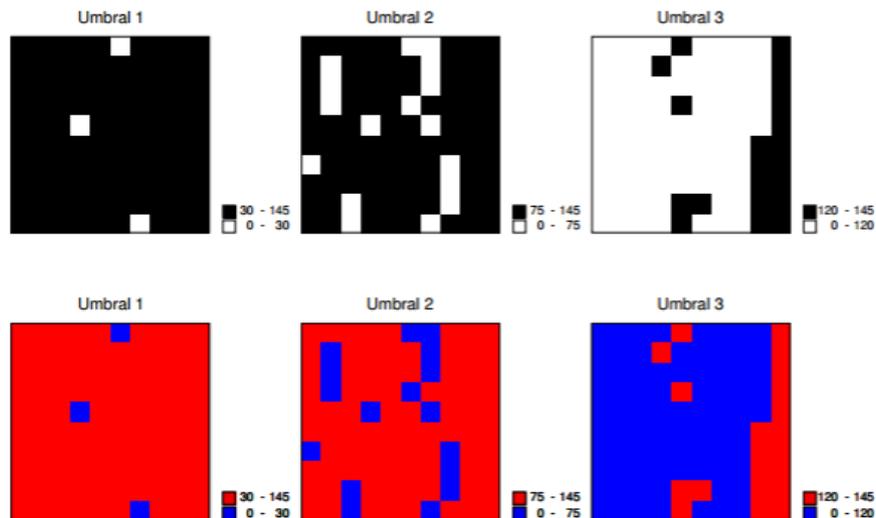


Fig. 16. Mapas indicadores. (Fuente: Isaaks.E, Srivastaba.R, 1989)

2.4.3. Continuidad espacial

Uno de los conceptos más importantes en la geoestadística es la continuidad espacial. En el libro *An Introductio to Applied Geostatistics* (1989), la continuidad espacial se describe a partir de la Figura 17, generada con un algoritmo matemático, la cual podría representar una variable en dos dimensiones, por ejemplo, la porosidad del terreno.

Al mirar esa imagen y los colores, llama la atención que los píxeles con colores más parecidos se encuentran asociados en sentido horizontal más que en vertical, es decir los valores cambian más rápidamente si se recorre en sentido vertical que en horizontal. Se dice por ello que ese parámetro es más continuo en sentido horizontal y por tanto, menor en sentido vertical.

Por continuidad espacial se debe pensar en la similitud de los valores de una variable cuando se observan en dos posiciones. Si estos valores se dispersan de forma aleatoria, como sucede en la Figura 17 b, se puede observar que ha desaparecido la estructura, ese ordenamiento en sentido horizontal ya no está, se ha perdido completamente la continuidad espacial, a esto se le llama ruido blanco.

Lo primero que se debe saber en geoestadística es distinguir entre ambos campos, ser capaces de caracterizarlo de alguna forma matemática, es decir, obtener una ecuación que permita decir si ese campo está ordenado de una forma determinada.

Para ello, lo primero que se puede hacer es realizar un análisis univariado de los datos, cuyos resultados se pueden observar en la tabla que se encuentra a la derecha de la Figura 17. En dichas tablas, se puede observar el número de datos, la media, la varianza, etc. Y ambas son iguales ya que estos estadísticos no tienen en cuenta la posición en el espacio. Es decir, la estadística univariada no me sirve para caracterizar la continuidad espacial.

Afortunadamente, las herramientas de la estadística bivariada, utilizadas para describir la relación entre dos variables, sí que permiten caracterizar la continuidad espacial ya que pueden utilizarse para describir la relación entre el valor de una variable y los valores de la misma variable en localizaciones cercanas, la relación en cuanto a su correlación.

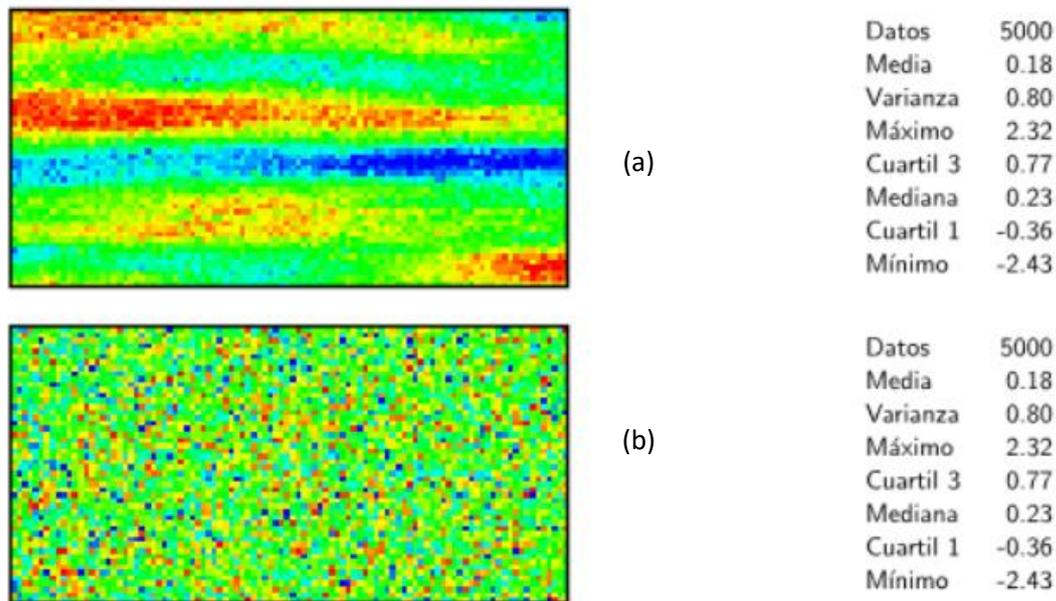


Fig.17. Ejemplo de continuidad espacial (a) y ruido blanco (b).
(Fuente: Cassiraga, 2015)

Para ello lo que se va a utilizar es lo que se llama el método gráfico de los diagramas de dispersión. Un diagrama de dispersión, como bien se dijo en el apartado 2.3.2.4, es un diagrama x-y donde se representan los valores de una variable frente los de otra.

Los estadísticos bivariados clásicos sirven para analizar la información contenida en un diagrama de dispersión.

2.4.3.1. Diagrama de dispersión tipo h

Al adaptar dicho método gráfico a la geoestadística se obtiene el diagrama de dispersión tipo h. Éste compara los valores de la misma variable en dos posiciones diferentes en el espacio separadas por un vector h, y todas las parejas representadas tendrán la particularidad de estar separadas ese mismo vector h, ya que lo que se busca es la relación entre las medidas del parámetro registrado en el espacio.

Para ello, se ordena de alguna forma los valores registrados, siguiendo la metodología de buscar parejas separadas por un mismo vector, es decir que siga la misma dirección y módulo. La nomenclatura utilizada es la que se presenta en la Figura 18:

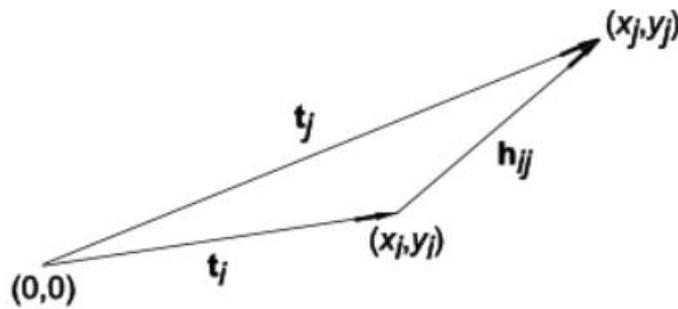


Fig.18. Nomenclatura vector. (Fuente: Isaaks.E, Srivastaba.R, 1989)

La localización del punto i de coordenadas (x_i, y_i) se especifica a través del vector t_i ; la localización del punto j de coordenadas (x_j, y_j) se especifica a través del vector t_j ; la separación entre el punto i y el punto j se denota como $t_j - t_i$ o como $(x_j - x_i, y_j - y_i)$; el vector que va de i a j como h_{ij} y el que va de j a i como h_{ji} (vectores separación)

Una vez clasificados los datos según la separación, se obtienen unos subgrupos de datos los cuales se estudiarán estadísticamente. Primero se dibujará para cada subconjunto un diagrama de dispersión, como se puede observar en la Figura 22.

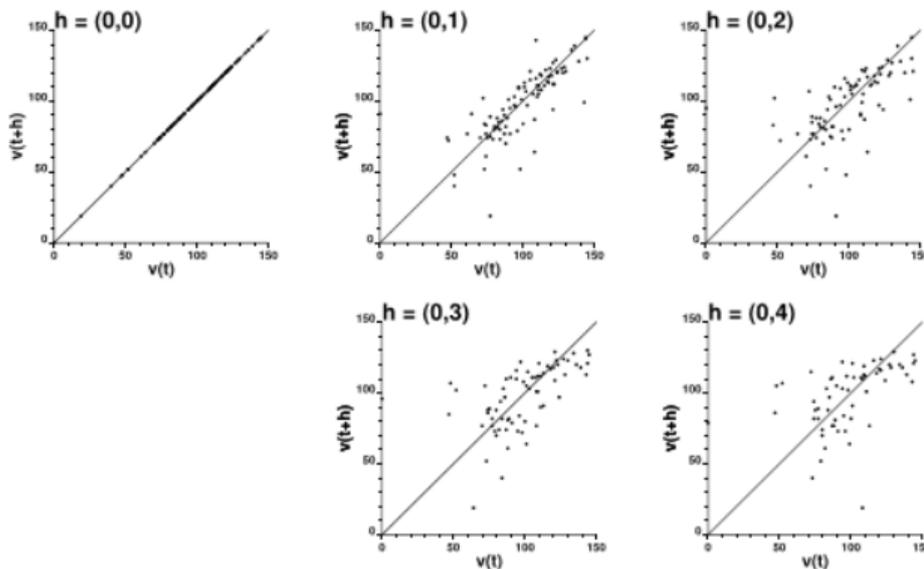


Fig.19. Diagrama de dispersión tipo h para distintos vectores de separación h .

(Fuente: Isaaks.E, Srivastaba.R, 1989)

En la Figura 19 se dibujan todos los valores asociados con la cola del vector y en el eje y se dibujan los valores que se encuentran en la cabeza del vector para un vector separación h igual a 0, 1, 2, 3, 4 en este caso, por eso se pone la etiqueta $V(t+h)$. Siendo V el parámetro y siendo los valores de dentro del paréntesis la coordenada. Además, t es una coordenada genérica y h es el valor fijado.

Como se habla de la misma variable, la recta en la que $x=y$, está a 45° ya que tienen la misma escala ambos ejes.

Observando los diagramas se puede comprobar que mientras en el primer diagrama todos los puntos caen en $X=Y$, ya que son parejas del mismo valor, en el resto no, lo que indica que cuanto más lejos están las estaciones asociadas, la dispersión de los puntos alrededor de la recta $x=y$ es mayor, la nube de puntos es más difusa. Luego a medida que te alejas de la recta $x=y$ los valores registrados son más diferentes en términos estadísticos. En términos de continuidad espacial esto indica que cuanto más se aleja, la continuidad será menor.

2.4.3.2. Medidas de continuidad espacial

Para estudiar la correlación entre dos muestras de una misma variable o dos muestras de distintas variables podemos calcular los siguientes estadísticos:

- La Covarianza (C)
- El Coeficiente de correlación (φ)
- Momento de inercia (Υ)

Se calcula para cada diagrama de dispersión la covarianza, el coeficiente de correlación y el momento de inercia, ya que se quiere conocer la relación entre ellos. Pero como se tiene un distinto grupo de datos para cada vector separación h , se aplicará el mismo algoritmo para cada uno de los conjuntos hallados.

Un conjunto se distingue del otro a partir del vector separación, transformándose estos vectores escalares en funciones de h :

- Función de Covarianza ($C(h)$)
- Función de correlación o correlograma ($\varphi(h)$)
- Variograma ($\Upsilon(h)$)

Por ejemplo, si se toman todas las parejas de puntos calculadas en la figura 19 y se halla el coeficiente de correlación, la covarianza y el momento de inercia, aplicando las fórmulas descritas en el análisis bivariado, da como resultado los datos de la Figura 20 y Tabla 3.

Distancia h	Coeficiente de correlación ρ	Covarianza C	Momento de inercia γ
(0,0)	1.000	686.4	0.000
(0,1)	0.742	448.8	312.8
(0,2)	0.590	341.0	479.2
(0,3)	0.560	323.8	521.4
(0,4)	0.478	291.5	652.9

Tabla 3. Coeficiente de correlación, covarianza y momento de inercia obtenido del ejemplo.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

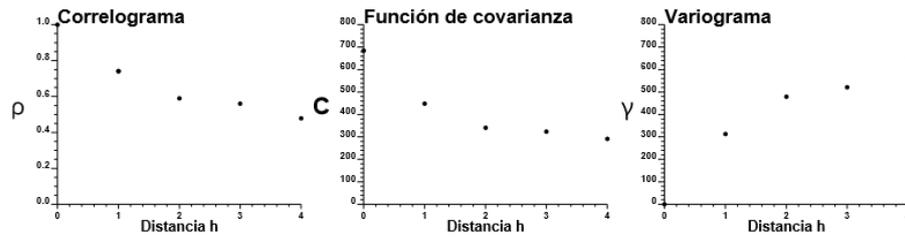


Fig.20. Correlograma. Función de covarianza y variograma del ejemplo.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

En la Tabla 3 se puede observar que en el análisis correspondiente al vector separación 0, el coeficiente de correlación da 1, es perfecta, lineal. El momento de inercia sin embargo es cero ya que todos los puntos están en la recta $x=y$. Además, La covarianza tiene un valor alto cuando la correlación es alta.

A medida que vamos aplicando lo mismo a los siguientes, se va obteniendo que el coeficiente de correlación y covarianza van disminuyendo según va disminuyendo la correlación entre los datos. En cambio, con el momento de inercia ocurre lo contrario, se parte de correlación perfecta, con valor cero, y según la nube se va haciendo más difusa las distancias se van haciendo más grandes a la recta $x=y$ y el momento de inercia va creciendo.

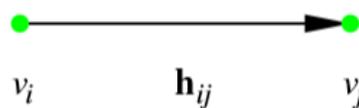
En resumen, a medida que va aumentando la distancia, el coeficiente de correlación y covarianza disminuyen, mientras que el momento de inercia o variograma aumenta.

Con ello se tienen tres estadísticos, funciones diferentes capaces de capturar la misma propiedad de un conjunto de datos, la continuidad espacial, con los cuales se puede medir como están relacionadas las medidas del parámetro en dos estaciones separadas una distancia determinada.

La expresión de la covarianza para el caso a estudiar va a variar, ya que no se está hablando de dos variables, sino de la misma variable en posiciones diferentes, es decir de la función de covarianza. Por lo tanto la fórmula a utilizar será:

$$C(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} v_i \cdot v_j - m_{-\mathbf{h}} \cdot m_{+\mathbf{h}}$$

$$m_{-\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{i|\mathbf{h}_{ij}=\mathbf{h}} v_i \quad m_{+\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{j|\mathbf{h}_{ij}=\mathbf{h}} v_j$$



Donde:

i, j son dos posiciones diferentes de la misma variable v

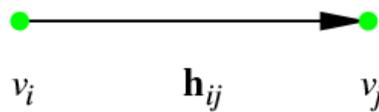
m_{-h} = media de todos los valores que se encuentran a la cola del vector

m_{+h} = media de todos los valores que se encuentran a la cabeza del vector

La función de correlación o correlograma será, por tanto:

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{\sigma_{-h} \cdot \sigma_{+h}}$$

$$\sigma_{-h}^2 = \frac{1}{N(\mathbf{h})} \sum_{i|\mathbf{h}_{ij}=\mathbf{h}} v_i^2 - m_{-h}^2 \quad \sigma_{+h}^2 = \frac{1}{N(\mathbf{h})} \sum_{j|\mathbf{h}_{ij}=\mathbf{h}} v_j^2 - m_{+h}^2$$



donde:

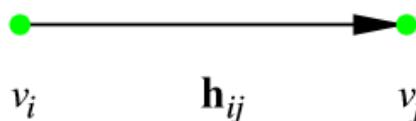
$C_{uv}(\mathbf{h})$: covarianza anteriormente calculada

σ_{u-h} desviación estándar de todos los valores de la cola del vector

σ_{v+h} desviación estándar de todos los valores de la cabeza del vector

Y por último el variograma:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} (v_i - v_j)^2$$



Estas tres fórmulas son las que contienen los programas de análisis de la continuidad espacial. Y se encargan de leer todas las coordenadas espaciales de los datos y sus correspondientes valores, dando como solución representaciones de lo ocurrido.

2.4.3.2.1. Variogramas cruzados de dispersión h

En este caso, se extiende el estudio a dos variables diferentes en distintas posiciones, de manera que se repite el mismo análisis pero en lugar de tener v_i y v_j , que serían la misma variable en las mismas posiciones, se coloca en la cabeza del vector la segunda variable. Ahora el eje de igualdad ya no estará a 45° porque cada eje tendrá su escala. Además, para el vector separación cero, al no estar comparando una variable consigo misma, sino dos variables distintas, los puntos ya no quedarán sobre la recta $x=y$. El resto no varía. Además, según va aumentando la separación también va aumentando la dispersión alrededor de la nube.

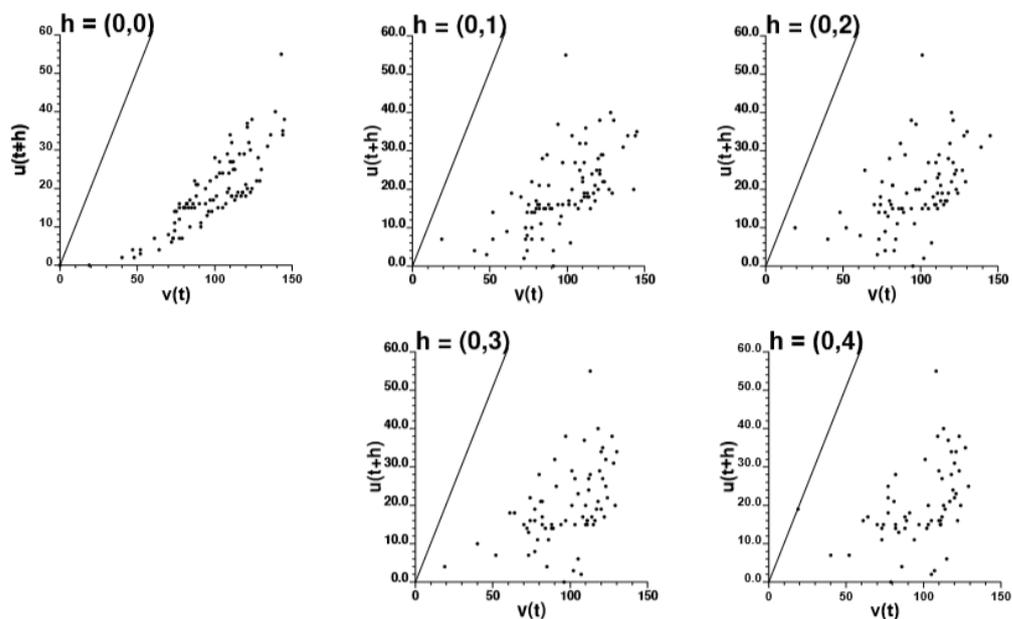


Fig. 21. Diagramas cruzados de dispersión tipo h para distintos vectores de separación h .
(Fuente: Isaaks.E, Srivastaba.R, 1989)

También se puede obtener la correlación calculando el coeficiente de correlación, la covarianza y el momento de inercia obteniendo funciones decrecientes en las dos primeras, y decreciente en la tercera (Tabla 4).

Distancia h	Coefficiente de correlación ρ	Covarianza C	Momento de inercia γ
(0,0)	0.84	218.3	38.7
(0,1)	0.60	144.0	54.2
(0,2)	0.45	94.2	80.7
(0,3)	0.36	73.1	89.5
(0,4)	0.28	60.1	111.0

Tabla 4. Coeficiente de correlación, covarianza y momento de inercia obtenido del ejemplo. (Fuente: Isaaks, E, Srivastaba, R, 1989)

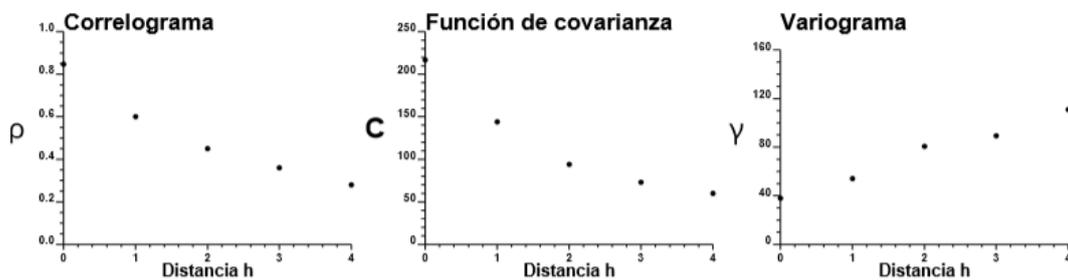


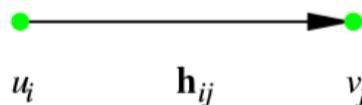
Fig.22. Correlograma. Función de covarianza y variograma del ejemplo. (Fuente: Isaaks, E, Srivastaba, R, 1989)

Este análisis que compara variables diferentes se le llama análisis cruzadas.

La función de covarianza cruzada será:

$$C_{uv}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} u_i \cdot v_j - m_{u-\mathbf{h}} \cdot m_{v+\mathbf{h}}$$

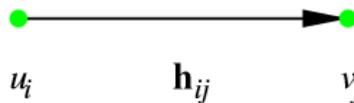
$$m_{u-\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{i|\mathbf{h}_{ij}=\mathbf{h}} u_i \quad m_{v+\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{j|\mathbf{h}_{ij}=\mathbf{h}} v_j$$



La función de correlación cruzada o correlograma cruzado:

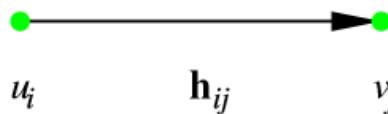
$$\rho_{uv}(\mathbf{h}) = \frac{C_{uv}(\mathbf{h})}{\sigma_{u-\mathbf{h}} \cdot \sigma_{v+\mathbf{h}}}$$

$$\sigma_{u-\mathbf{h}}^2 = \frac{1}{N(\mathbf{h})} \sum_{i|\mathbf{h}_{ij}=\mathbf{h}} u_i^2 - m_{u-\mathbf{h}}^2 \quad \sigma_{v+\mathbf{h}}^2 = \frac{1}{N(\mathbf{h})} \sum_{j|\mathbf{h}_{ij}=\mathbf{h}} v_j^2 - m_{v+\mathbf{h}}^2$$



Y por último el variograma cruzado será:

$$\gamma_{uv}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij}=\mathbf{h}} (u_i - u_j)(v_i - v_j)$$



Cuando se realiza un análisis cruzado lo que se hace es buscar parejas separadas un determinado vector, pero de dos variables diferentes.

2.4.3.2.2. Datos irregularmente espaciados

A continuación, se va a generalizar esta técnica al caso de datos irregularmente espaciados ya que normalmente no se realizan muestreos del espacio de forma regular, no siguen una configuración geométrica regular. Para encontrar estas parejas de un conjunto de datos irregularmente espaciados, se establecen unas tolerancias tanto angulares como lineales como se puede observar en la Figura 23.

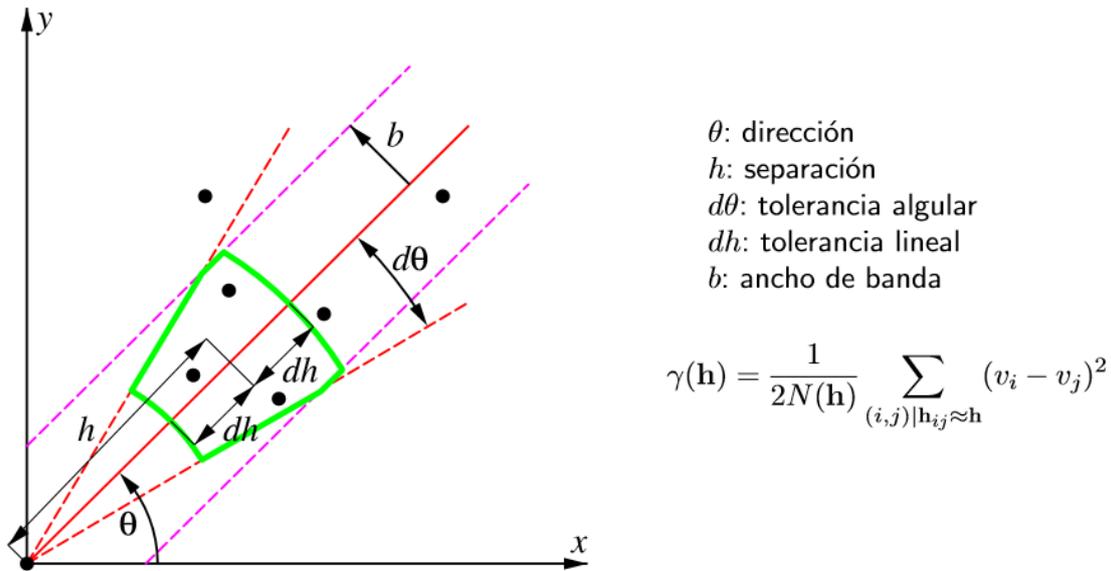


Fig. 23. Tolerancias en la búsqueda de datos irregularmente espaciados.
(Fuente: Cassiraga, 2015)

Uniendo la tolerancia lineal (dh) en el vector con la tolerancia en el ángulo ($d\theta$) se determina un área, con ello se puede ampliar el número de parejas encontradas con el que se mejorará el valor de los estadísticos a calcular. Para que el análisis direccional no pierda sensibilidad al tomar un área muy grande, se establece una tercera tolerancia que se llama ancho de banda que es una separación con respecto a la dirección de interés que establece unos límites, fuera de los cuales no se aceptarán parejas de valores.

Estas tres tolerancias son elegidas por el modelador y suele depender de cada problema y la calidad de las funciones que se obtienen. Es un problema de prueba y error.

Se debe tener en cuenta que los estadísticos a calcular, como por ejemplo el variograma, dependen del número de parejas que se tienen, cuantas más parejas, más representativo es la función.

2.4.4. Práctica de la descripción espacial.

La expresión del variograma permite cuantificar la continuidad espacial para una serie de distancias de separación y direcciones determinando así el variograma experimental.

Los métodos geoestadísticos requieren conocer el variograma para cualquier distancia de separación y cualquier dirección en la cual es considerada dicha separación, por lo que es necesario ajustar un modelo al variograma experimental (Fig. 24).

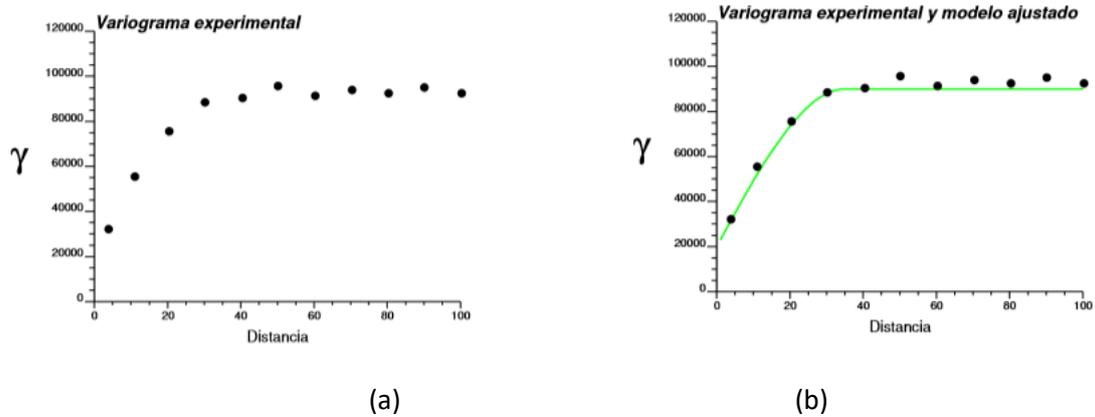


Fig. 24. Variograma experimental (a) y variograma experimental con modelo ajustado (b)
(Fuente: Cassiraga, 2015)

De manera que el objetivo final cuando estamos haciendo el análisis de la continuidad espacial es llegar a obtener unas funciones continuas en cuanto al espacio.

2.4.4.1. Terminología

Para comenzar, se va a aclarar la terminología que se va a utilizar (Fig. 25):

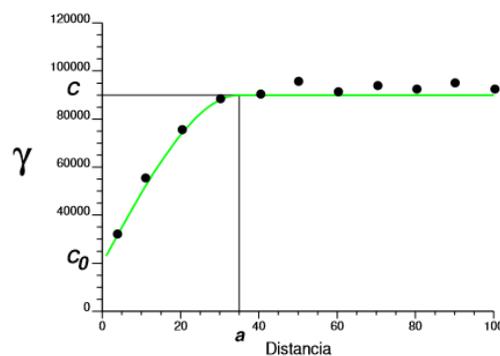


Fig.25. Terminología del variograma experimental con el modelo ajustado.
(Fuente: Cassiraga, 2015)

- A (alcance): distancia a la cual el variograma alcanza la meseta.
- C (meseta): valor del variograma cuando la distancia de separación se hace igual al alcance.
- C_0 (efecto pepita): valor del variograma para pequeñas escalas.

Un variograma tipo vale cero en el origen por definición (momento de inercia vale cero en el origen) pero para un vector separación pequeño en general pega un salto y empieza a crecer

hasta un valor estable llamado meseta. El salto en el origen es lo que se conoce como efecto pepita.

En el eje de las x se tiene el vector separación, representando para cada vector separación cual es la relación que hay entre las medidas separadas ese vector.

Por ejemplo, si se tiene un vector separación de 60 unidades de longitud, para el cual el variograma vale 1000, se le está dando una cuantificación de la relación que existe entre las estaciones separadas 60 unidades de longitud. Lo mismo para otros vectores h , incluso para vectores pequeños. Por ejemplo, si se toma h igual a 3 cm, se debería poder obtener un valor de la continuidad espacial para esos 3 cm, aunque nunca hay estaciones separadas 3cm en medidas de campo o hidrología. Pero como se está hablando de funciones continuas, se podrá extrapolar y conocer para vectores pequeños. Por ello en función de los puntos que se dispongan en la rama ascendente del variograma se puede intentar extender hacia los vectores pequeños el comportamiento de la función. El llamado efecto pepita serían los valores de la función que caracterizan la continuidad a pequeña escala.

El valor constante, llamado meseta, se alcanza para un vector separación llamado alcance, es decir, cuando llego a una determinada longitud, el variograma se estabiliza y permanece constante, esa longitud a la que se llega a un valor constante, se le denomina alcance y el valor constante meseta. A partir de dicho valor, se pierde completamente la correlación de los datos, no hay continuidad espacial.

Estos tres parámetros (meseta, alcance y efecto pepita) definirán la función que se quiere ajustar.

2.4.4.2. Metodología

Los pasos a seguir para el análisis de continuidad espacial son:

1. Estudio de los datos y cálculo del variograma omnidireccional.

El variograma omnidireccional es un variograma medio que no distingue direcciones. Éste, al ser un promedio, es único y es de utilidad al permitir ajustar inicialmente algunos de los parámetros que se usarán durante el análisis de la continuidad espacial, como por ejemplo el vector separación con el cual buscar parejas.

Por ejemplo, se dispone de un conjunto de datos, a los cuales se les calcula el variograma omnidireccional, utilizando como primer vector separación 5 unidades de longitud, y se van incrementando dicho valores 10, 15, 20, 25... y así hasta casi 100. Siendo las parejas que se encuentran y su correspondiente variograma omnidireccional los que se muestran en la Tabla 5 y Figura 26:

Número de parejas	Distancia de separación	$\gamma(h)$
22	2.1	11294.1
488	5.4	42671.4
1720	10.4	51932.4
1856	14.8	71141.8
3040	20.3	70736.9
2412	24.9	86745.2
3550	30.1	84077.8
2816	34.8	99986.6
4092	40.3	89954.4
3758	44.9	86155.0
4248	50.1	98319.3
3920	55.0	94415.1
5324	60.2	88848.9
4442	64.8	96309.2
5478	70.2	96397.3
4696	74.8	90704.6
5762	80.2	92560.6
5084	84.9	88104.0
5666	90.1	95530.9
4458	94.8	101174.8
2890	98.8	94052.1

Tabla 5. Núm. parejas, distancia separación y variograma.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

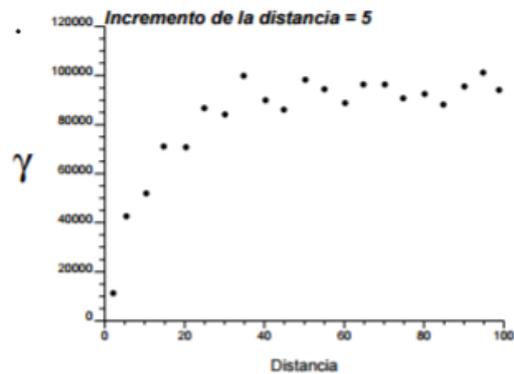


Fig.26. Variograma omnidireccional (h=5)
(Fuente: Isaaks.E,Srivastaba.R, 1989)

Se hace lo mismo pero para un vector separación 10 unidades de longitud, y se van incrementando dicho valores h=10, 20, 30.....hasta 100, obteniendo:

Número de parejas	Distancia de separación	$\gamma(h)$
178	3.6	32544.3
3044	11.0	55299.8
5140	20.4	75224.6
6238	30.2	88418.6
7388	40.5	90544.1
7954	50.1	95689.1
9782	60.3	91285.2
10060	70.3	93809.2
10628	80.3	92357.8
10454	90.1	95010.5
4856	97.8	97349.3

Tabla 5. Núm. parejas, distancia separación y variograma.
(Fuente: Isaaks.E, Srivastaba.R, 1989)

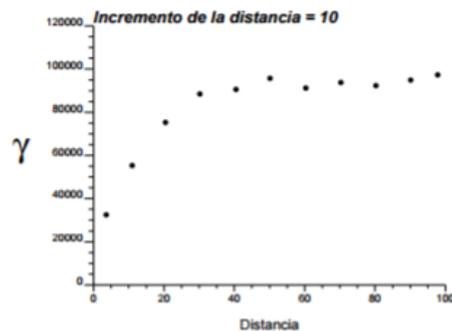


Fig.27. Variograma omnidireccional (h=10).
(Fuente: Isaaks.E, Srivastaba.R, 1989)

Si se tuviese que elegir entre ambos variogramas, se elegiría el segundo de ellos ya que las oscilaciones son menores, la nube de puntos es más estable.

Por todo ello el variograma tiene este único fin, el de detectar el vector separación inicial que produce la nube de puntos más estable posible, es decir, aquel que a simple vista cumple las características de un variograma tipo, es decir, un salto en el origen, una rama ascendente la cual se estabiliza para un cierto valor.

En realidad, este cálculo se habrá hecho con una determinada tolerancia en el vector, de forma que se empieza a buscar parejas con esa distancia (en nuestro ejemplo h=5) \pm la tolerancia indicada, siendo cada distancia entre las parejas diferentes, pero cumpliendo ese criterio. Con ello se hace la media aritmética de dichas distancias obteniendo el valor deseado. Tras esto se busca con el siguiente valor (h=10) \pm la misma tolerancia y así con todos.

2. Construcción de los variogramas direccionales.

Cuando se obtiene un variograma, las funciones obtenidas corresponden a una determinada dirección, si se busca el análisis en otra dirección, se debe repetir todo el proceso anterior en otra dirección, obteniendo así otras funciones que corresponderán a la otra dirección. Si estas funciones obtenidas son muy parecidas entre sí, el fenómeno estudiado es isótropo, es decir, se mire en la dirección que se mire, el fenómeno se comportará siempre igual.

Para ello se calculan los variogramas direccionales, para poder observar si en el fenómeno a estudiar existe o no anisotropía. Para ello, se planteará la obtención de los variogramas en varias direcciones típicas y se compararán unos con otros.

Un variograma tiene una rama ascendente que corresponde a los vectores dentro de los cuales existe relación entre las estaciones de medida, pero a partir de la distancia "a" (alcanse), ya no hay correlación. Por lo que para saber si el fenómeno de estudio es isótropo o anisótropo, se encontrará esa distancia en cada dirección analizada a partir del cual la correlación es nula. Si este valor se encuentra a la misma distancia en cualquier dirección, es decir si su unión forma un círculo (al representar su valor con dicha dirección), se estará hablando de un fenómeno isótropo. Pero si la forma es elíptica, es decir, cambia de valor según la dirección, el fenómeno será anisótropo. Por lo tanto, el análisis consistirá en realizar 7-9 variogramas experimentales en las 7-9 direcciones (Fig. 28).

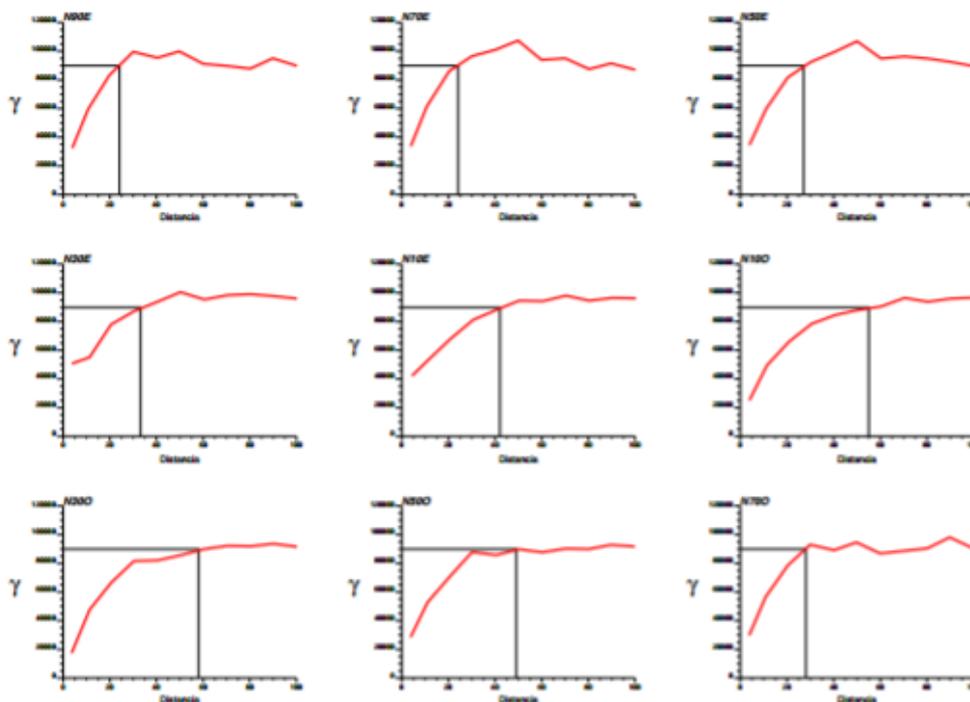


Fig.28. Variogramas direccionales y sus correspondientes alcances
(Fuente: Cassiraga, 2015)

El semieje mayor de la elipse, que no tiene porqué coincidir con ninguna de las direcciones omnidireccionales, indicará la dirección de máxima continuidad. Indicando por tanto el semieje menor, la dirección de mínima continuidad.

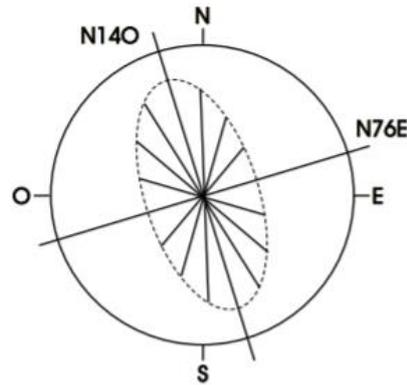


Fig.29. Elipse de anisotropía. (Fuente: Cassiraga, 2015)

Si el fenómeno es isótropo servirá un único variograma para todas las direcciones, pero si es anisótropo, se tendrá que encontrar cuales son las direcciones extremas, que estarán alineadas con el semieje mayor y menor de la elipse. Esto se llama elipse de anisotropía (Fig.29).

Una vez que se tiene identificadas ambas direcciones, se les tiene que ajustar una función, un modelo que permita su utilización en el esquema de las técnicas geoestadísticas de estimación y/o simulación.

Un modelo de variograma es una función matemática expresada en función de la separación, la dirección y unos pocos parámetros a determinar durante el proceso de la modelización. Las funciones aptas para ser modelo deben cumplir con ciertas reglas que garantizan la existencia de una solución única y estable de los sistemas de ecuaciones lineales subyacentes a las técnicas de estimación y/o simulación.

Los modelos básicos que se suelen usar son:

- Modelo efecto pepita $\rightarrow \gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ 1 & \text{para los otros casos} \end{cases}$
- Modelo esférico $\rightarrow \gamma(h) = \begin{cases} 1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a}\right)^3 & \text{si } h \leq a \\ 1 & \text{para los otros casos} \end{cases}$
- Modelo exponencial $\rightarrow \gamma(h) = 1 - \exp\left(\frac{-3h}{a}\right)$
- Modelo gaussiano $\rightarrow \gamma(h) = 1 - \exp\left(\frac{-3h^2}{a^2}\right)$

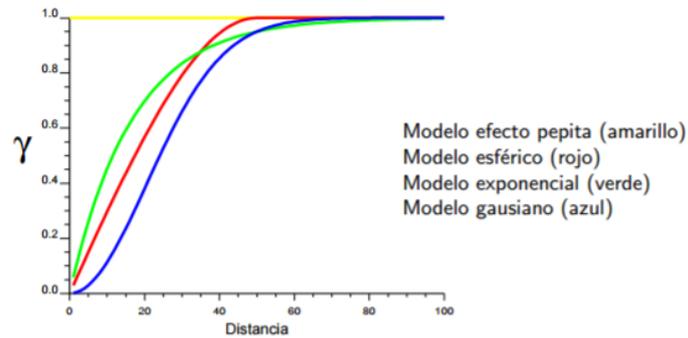


Fig.30. Modelos posibles a ajustar en el variograma.
(Fuente: Cassiraga, 2015)

Los modelos de la Figura 30 están representados para cuando la meseta es 1, pero se puede extender a cualquier valor de la meseta.

Una vez ajustadas las dos funciones para las dos direcciones de máxima y mínima continuidad, se acaba el análisis de la continuidad espacial por lo que se tiene ya caracterizada la continuidad espacial del fenómeno de estudio. A partir de aquí, ya se puede pasar a las aplicaciones de estimación y simulación.

2.4.4.3. Estructuras imbricadas

Cuando la nube de puntos a la que se tiene que realizar el ajuste es complicada, se puede combinar linealmente los modelos anteriormente expuestos. Eso es lo que se denomina estructura imbricada, es decir, el variograma final puede ser una combinación lineal de n variogramas parciales. Esto es en teoría, ya que en la práctica habrá que aplicar el principio de parsimonia, es decir, cuantos menos variogramas se necesite combinar para conseguir el ajuste mejor, como mucho un efecto pepita y una estructura, máximo dos, aunque en la teoría no haya límite en el número a combinar.

Un ejemplo de estructura imbricada sería el que se muestra en la Figura 31:

$$\gamma(h) = 22000 + 40000 \text{ Esf}_{30}(h) + 45000 \text{ Esf}_{150}(h)$$

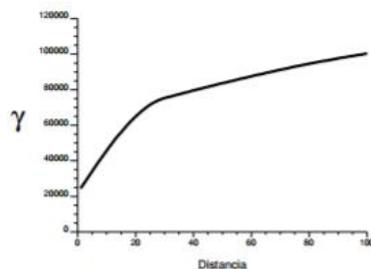


Fig.31. Estructura imbricada.
(Fuente: Cassiraga, 2015)

En este ejemplo se observan dos ramas ascendentes, las cuales se consiguen combinando linealmente dos funciones, dos estructuras esféricas. En este caso se tiene como salto en el origen 22000, a continuación, una primera rama ascendente, la cual es una función esférica con alcance 30 y un peso de 40000. Y finalmente, se añade otra igual de 45000. La suma total de la meseta da 107000 para un alcance de 150.

2.4.3.3. Modelos de anisotropía

Los fenómenos naturales son generalmente anisótropos, pero hay diferentes tipos de anisotropía.

Cuando se tiene un comportamiento anisótropo, llegando a la misma meseta por las dos direcciones, la de mínima y la de máxima continuidad, se dice que la anisotropía es geométrica (Fig. 32).

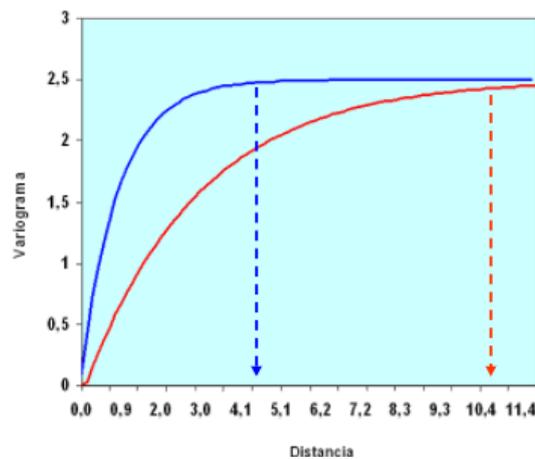


Fig.32. Modelo de anisotropía geométrica. (Fuente: Slideshare, 2015)

La función que tarda más en llegar a la meseta es la dirección de máxima continuidad, ya que la memoria se extiende a más distancia. Los diferentes alcances serán el semieje mayor y semieje menor de la elipse de anisotropía.

Puede ocurrir que se tenga el mismo alcance en todas las direcciones, pero distinta meseta, a esto se le denomina Anisotropía zonal (Fig. 33) En la naturaleza es muy raro que se de este tipo de comportamientos.

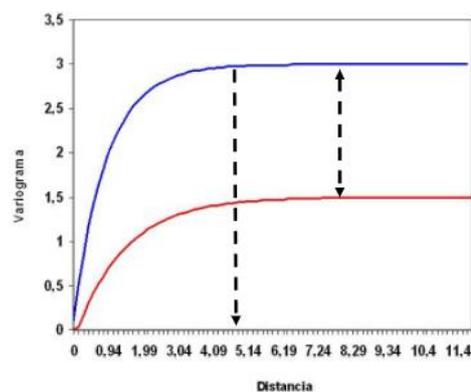


Fig.33. Modelo de anisotropía zonal. (Fuente: Slideshare, 2015)

Lo que si se encuentra fácilmente en la naturaleza, es la combinación de la anisotropía geométrica más la zonal. Se produce un desplazamiento en los alcances, pero sin llegar a la misma meseta (Fig. 34):

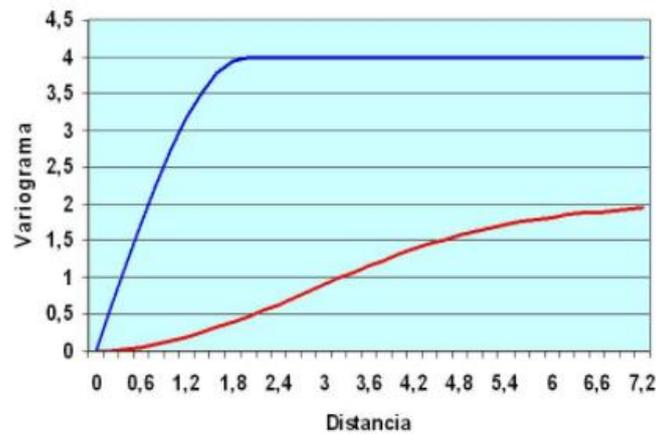


Fig.34. Modelo de combinación de anisotropía geométrica y zonal.
(Fuente: Slideshare, 2015)

Con esto se termina la primera parte de un estudio geoestadístico, que es el análisis de la información. Los primeros pasos recurren a la estadística univariada y bivariada, donde únicamente se obtiene información del valor de la variable, pero no de su localización. Con ello y el análisis de la continuidad espacial se ha completado el análisis, produciendo como resultado una ecuación, que es el variograma donde está contenida toda la información acerca de la conectividad que hay entre las estaciones de medida en las distintas direcciones en el espacio.

2.5. El modelo de función aleatoria

2.5.1. Introducción

Se va a estudiar las bases del modelo matemático que subyace a todos los algoritmos de geoestadística, este es un modelo probabilístico estocástico que se llama modelo de función aleatoria.

Hasta ahora lo que se ha conseguido es describir la información, pero raramente es el objetivo de un estudio geoestadístico. El principal interés suele estar en las inferencias acerca de las características de la realidad a partir de la información registrada, es decir, conseguir una representación de la realidad.

El análisis estadístico descriptivo es un primer paso que permite obtener una serie de estadísticos a partir de los cuales podemos construir un modelo (una representación) acerca de la realidad desconocida que se quiere investigar.

Suponiendo que se tiene el siguiente conjunto de mediciones de la figura 35, los cuales son las estaciones de medida. El sitio en el que se ubican es la localización y el color de los puntos es el valor de la variable según la escala.

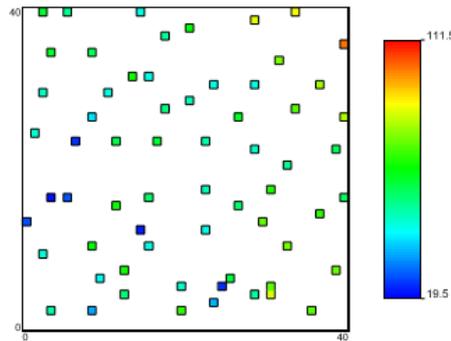


Fig.35. Ejemplo datos modelo función aleatoria
(Fuente: Elaboración propia)

¿Qué sucede con el parámetro o atributo que ha sido registrado en las localizaciones donde no ha sido medido? Esta es la razón por la que es necesario la realización de modelos, el insuficiente conocimiento de un determinado parámetro en el cual la información disponible no es exhaustiva. Si no se propone un modelo que explique cómo varía entre dos estaciones el parámetro en el que se está interesado, es imposible obtener de él una representación.

Un modelo es una representación de la realidad que desconocemos. La realidad es única, no hay ambigüedad, en cambio, nuestras representaciones pueden ser infinitas. La incertidumbre es un elemento que debe ser introducido en el ejercicio de modelación, y uno de los objetivos serían buscar maneras de cuantificar esa incertidumbre. La geoestadística ofrece una serie de herramientas para cuantificarla.

2.5.2. Tipos de modelos

Edward H. Isaaks (1989) y R. Mohan Srivastava (1989) citan que existen modelos determinísticos y probabilísticos.

Los modelos determinísticos se basan en la física del problema, como puede ser la mecánica newtoniana o las leyes del movimiento. Asocian a cada localización no muestreada un sólo valor estimado. Se dice que refleja un conocimiento “perfecto” del fenómeno estudiado.

Sin embargo, los modelos probabilísticos se basan en los datos medidos, construyendo todo a partir de dichas medidas registradas en la realidad. A diferencia de los modelos determinísticos que devuelven un solo valor en cada estimación, los probabilísticos asocian a cada localización no muestreada un conjunto de posibles valores con sus correspondientes probabilidades de ocurrencia. Por ello al final lo que se obtiene es una función de distribución de probabilidades. Es por ello que un modelo probabilístico refleja un conocimiento imperfecto del fenómeno estudiado.

Por ejemplo, se dispone de un conjunto de datos que podría ser, por ejemplo, la altura en función de la distancia de un objeto, ya sea una pelota que se deja caer y va botado en el suelo (Figura 36). El movimiento se va amortiguando por el rozamiento hasta que la pelota se para. Solamente se conocen 7 medidas, y se necesita ajustar un modelo que explique el comportamiento continuo de ese fenómeno. Con las leyes de Newton y el modelo determinístico, podemos obtener el valor perfecto de la realidad.

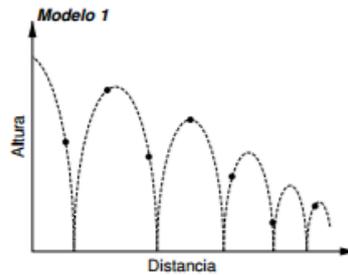


Fig.36. Ejemplo modelo determinístico I.
(Fuente: Isaaks.E,Srivastaba.R, 1989)

Sin embargo, esos 7 datos podrían ser por ejemplo la tasa de interés que un banco toma todos los jueves de cada semana (Figura 37). El banco ajusta su tasa una vez por semana cada viernes. Como se mantiene constante cada semana forman escalones, y si se toma la medida cada semana, este escalón variará.

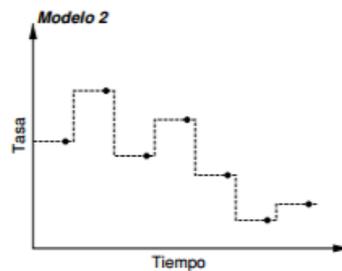


Fig.37. Ejemplo modelo determinístico II.
(Fuente: Isaaks.E,Srivastaba.R, 1989)

En cambio, si los 7 puntos corresponden a la concentración de un material pesado como el zinc, y se quiere conocer a partir de esas 7 medidas la concentración a x cm de profundidad donde no se ha medido, se necesitaría construir un modelo probabilístico para construir una función de distribución de probabilidades de los valores de la concentración del zinc en ese punto (Figura 38). No es posible obtener, como en el caso anterior, un único valor para cada posición en el espacio, tengo muchos valores con su probabilidad de ocurrencia.

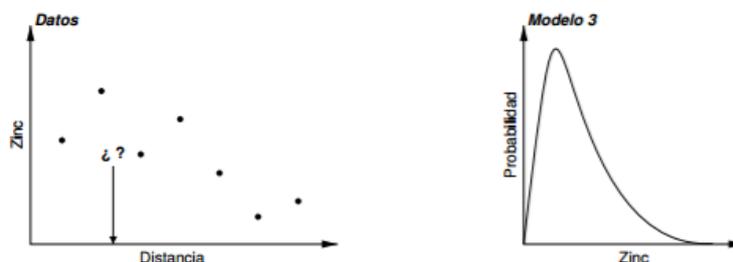


Fig.38. Ejemplo modelo probabilístico
(Fuente: Goodvaerts, 1997)

2.5.3. El modelo de función aleatoria

2.5.3.1. Variable aleatoria

Para poder modelar lo anterior, se tendrá que hacer uso del concepto de variable aleatoria. Una variable aleatoria es una variable cuyos valores son generados aleatoriamente de acuerdo a algún mecanismo probabilístico.

Cuando se trata de un parámetro que varía de forma continua, es decir, que sus valores pueden ser cualesquiera dentro de un rango, el comportamiento de ese atributo si lo modelamos como una variable aleatoria, viene dado por una función de distribución de probabilidades que cuando la escribimos de manera acumulada queda:

$$F(u; z) = Prob\{Z(u) \leq z\} \forall z$$

la cual da la probabilidad de que la variable aleatoria Z en la localización u sea menor o igual que un determinado valor umbral z .

Una función aleatoria es un conjunto de variables aleatorias $Z(u)$, generalmente correlacionadas entre sí, definidas para cada localización u sobre el dominio de estudio A tal que $\{Z(u), \forall u \in A\}$.

En geoestadística se aproxima a la realidad suponiendo que los valores registrados por las estaciones de medida son el resultado de muestrear en cada posición en el espacio una variable aleatoria. Como se disponen de muchas estaciones de medida, se muestrean muchas variables aleatorias que constituyen lo que se llama una función aleatoria.

Se supone que se dispone un dominio de trabajo (Figura 39) que podría ser un acuífero, una cuenca hidrográfica, etc.



Fig.39. Dominio de trabajo.
(Fuente: Cassiraga, 2015)

Siempre se harán unos modelos, para ello se discretiza el dominio de trabajo en x celdas (figura 40) y en el centro de cada celda se dispone de una posible estación de medida o una posible necesidad de informar al modelo del valor de una variable.

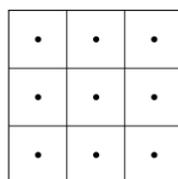


Fig.40. Dominio de trabajo discretizado.
(Fuente: Cassiraga, 2015)

Según Cassiraga (2015), en geoestadística se supone que en cada pixel el comportamiento de la variable a registrar es descrito por una función de distribución de probabilidad. Si se quiere obtener valores de las variables aleatorias para todos los pixeles, tendré que muestrear cada una de las funciones y obtener dichos valores.

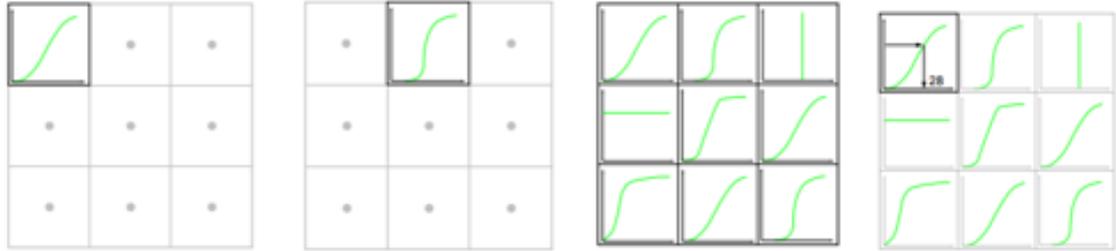


Fig.41. Obtención de variables aleatorias a través de las funciones de distribución de probabilidad. (Fuente: Cassiraga, 2015)

Una variable regionalizada se interpreta como una función aleatoria $Z(u)$. Cada muestra u observación de la función aleatoria $Z(u)$ se llama realización de $Z(u)$.

28	38	32	43	23	59
71	66	45	89	31	36
16	24	19	68	26	57

Fig.42. Diferentes realizaciones de la misma función aleatoria. (Fuente: Cassiraga, 2015)

Esta aproximación permite hablar de incertidumbre dado que se trabaja con una aproximación estocástica probabilística. En cada una de las celdas, se ha asumido que se cuenta con una distribución de probabilidades, es decir, se dispone de todos los valores posibles que puede tomar la variable con sus probabilidades asociadas. Luego se tiene un modelo de la incertidumbre del atributo para cada localización.

En resumen, si se mide en una misma localización en el espacio muchas veces el atributo que se quiere conocer, se obtendrían muchas medidas. A partir de estas, se podría construir una función de distribución de probabilidades, que sería un modelo de incertidumbre local, ya que está asociada a esa localización.

Por otro lado, como se puede tener varias imágenes de mi atributo, ese conjunto de medidas es lo que me permite valorar la incertidumbre espacial del atributo, es decir, una imagen permite visualizar el tipo de relación que hay en el dominio de trabajo entre los valores que toma la variable.

2.5.3.1.1. Ley espacial de la función aleatoria

En resumen, el modelo de función aleatoria que es el utilizado en geoestadística, consiste en abordar la realidad pensando que los valores que toman los parámetros son el resultado de muestrear variables aleatorias, una en cada posición en la que se está interesado es decir, el conjunto de todas esas variables aleatorias, conforman lo que se llama una función aleatoria.

Si se tienen un conjunto de N localizaciones u_k , $k = 1, \dots, N$ corresponde un vector de N variables aleatorias $\{Z(u_1), \dots, Z(u_N)\}$, que esta caracterizado por la función de distribución acumulada N -variada siguiente:

$$F(u_1, \dots, u_N; z_1, \dots, z_N) = \text{Prob}\{Z(u_1) \leq z_1, \dots, Z(u_N) \leq z_N\}$$

La ecuación anterior caracteriza la incertidumbre conjunta acerca de los N valores $z(u_1), \dots, z(u_N)$

El conjunto de todas las N -variadas funciones de distribución acumuladas para algún entero positivo N y para alguna elección de las localizaciones u_k , constituye la ley espacial de la función aleatoria $Z(u)$.

Con ello se podría indicar como están relacionados los valores del atributo (conocimiento exhaustivo de la medida de ese parámetro). En la práctica esto es imposible, el análisis se limita a calcular las funciones de distribución acumulada y los momentos para no más que dos localizaciones conjuntamente.

A continuación se definen las funciones y estadísticos que se irán utilizando en el posterior análisis:

- Función de distribución acumulada univariada:

$$F(u, z) = \text{Prob}\{Z(u) \leq z\}$$

- Función de distribución acumulada bivariada:

$$F(u_i, u_j; z_i, z_j) = \text{Prob}\{Z(u_i) \leq z_i, Z(u_j) \leq z_j\}$$

- Esperanza matemática o valor esperado también llamado deriva de $Z(u)$:

$$E\{Z(u)\} = m(u)$$

- La varianza es una medida de la dispersión de las realizaciones de $Z(u)$ en u .

$$\text{Var}\{Z(u)\} = \sigma^2(u) = E\{[Z(u) - m(u)]^2\} = E\{[Z(u)]^2\} - [m(u)]^2$$

- La covarianza centrada es una medida de la correlación entre las variables aleatorias $Z(u_i)$ y $Z(u_j)$.

$$C(u_i, u_j) = E\{[Z(u_i) - m(u_i)][Z(u_j) - m(u_j)]\} = E\{Z(u_i)Z(u_j)\} - m(u_i)m(u_j)$$

En el caso particular en que $u_i = u_j = u$

$$C(u_i, u_j) = \sigma^2(u)$$

- El correlograma

$$\rho(u_i, u_j) = \frac{C(u_i, u_j)}{\sqrt{C(u_i, u_i) \cdot C(u_j, u_j)}}$$

- El variograma es una medida de la disimilitud entre $Z(u_i)$ y $Z(u_j)$.

$$\gamma(u_i, u_j) = \frac{1}{2} E \{ [Z(u_i) - Z(u_j)]^2 \}$$

Todos estos estadísticos dependen de las posiciones.

2.5.3.2. La decisión de estacionariedad

Las funciones de distribución acumulada univariadas y bivariadas y sus correspondientes momentos son dependientes de la localización, por lo que su inferencia requeriría muchas realizaciones para cada localización u .

Por ejemplo (Figura 43), la inferencia de la $C(u_7, u_{13})$ entre las dos variables aleatorias $Z(u_7)$ y $Z(u_{13})$ separadas por un vector $h = (u_{13} - u_7)$, exigiría un conjunto de medidas repetidas en cada localización u_7 y u_{13} de las que no se dispone.

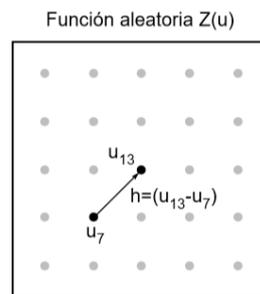


Fig.43. Datos disponibles cálculo de la covarianza.
(Fuente: Cassiraga, 2015)

Por lo que, si se está trabajando sobre un problema real y se quisiese calcular este estadístico, por ejemplo en el contexto hidrológico, se tiene una cuenca con pluviógrafos, y para un determinado momento solo se dispone de un conjunto de medidas. Para calcular la covarianza entre dos estaciones, si solo se tiene una medida, no se puede calcular utilizando las fórmulas para las cuales necesito n datos. Ese problema se solventa recurriendo a la decisión de estacionariedad. Se va a suponer que el comportamiento entre dos estaciones es el mismo que el que se da para todas las estaciones separadas ese mismo vector. La idea es utilizar todas las medidas separadas un vector $h = (u_{13} - u_7)$ dentro del dominio de estudio A. Esto exige que el fenómeno estudiado sea "homogéneo" dentro del dominio A o, en términos probabilísticos, el modelo de función aleatoria $Z(u)$ debe ser estacionario dentro de A.

En su lugar, la idea es utilizar todas las medidas separadas un vector $h = (u_{13} - u_7)$ dentro del dominio de estudio A.

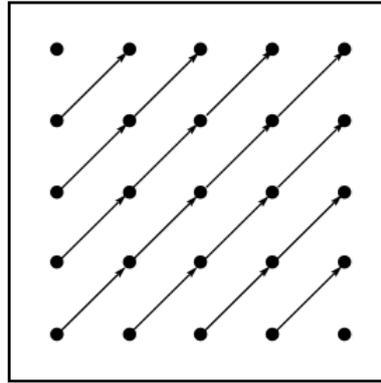


Fig.44. Datos disponibles tras decisión estacionariedad para cálculo de covarianza.
(Fuente: Cassiraga, 2015)

Por lo tanto, Una función aleatoria $Z(u)$ se dice ser estacionaria dentro de un dominio A si su función de distribución acumulada es invariante bajo traslación, esto es, cualquier par de vectores de variables aleatorias

$$\{Z(u_1), \dots, Z(u_N)\} \text{ y } \{Z(u_1 + h), \dots, Z(u_N + h)\}$$

tienen la misma función de distribución acumulada cualquiera sea el vector de traslación h , lo que se escribe:

$$F(u_1, \dots, u_N; z_1, \dots, z_N) = F(u_1 + h, \dots, u_N + h; z_1, \dots, z_N) \forall u_1, \dots, u_N \text{ y } h$$

La decisión de estacionariedad es limitada a las funciones de distribución acumulada univariada y bivariada y a sus primeros dos momentos, los cuales se pueden expresar ahora dependiendo sólo del vector separación h

- La función de distribución acumulada univariada

$$F(u; z) = \text{Prob}\{Z(u) \leq z\} \Rightarrow F(z) = \text{Prob}\{Z(u) \leq z\}$$

- La función de distribución acumulada bivariada

$$F(u_i, u_j; z_i, z_j) = \text{Prob}\{Z(u_i) \leq z_i, Z(u_j) \leq z_j\} \Rightarrow F(h; z_i, z_j) = \text{Prob}\{Z(u) \leq z_i, Z(u+h) \leq z_j\}$$

- El valor esperado

$$E\{Z(u)\} = m(u) \Rightarrow E\{Z(u)\} = m$$

- La varianza

$$\text{Var}\{Z(u)\} = \sigma^2(u) \Rightarrow \text{Var}\{Z(u)\} = \sigma^2$$

- La covarianza

$$C(h) = \text{Cov}\{Z(u)Z(u+h)\} = E\{Z(u)Z(u+h)\} - E\{Z(u)\}E\{Z(u+h)\} \\ = E\{Z(u)Z(u+h)\} - E\{Z(u)\}^2$$

$$C(0) = \text{Cov}\{Z(u)Z(u)\} = \text{Var}\{Z(u)\}$$

- El correlograma

$$\frac{\text{Cov}\{Z(u)Z(u+h)\}}{\sqrt{\text{Var}\{Z(u)\} \cdot \text{Var}\{Z(u+h)\}}} = \frac{C(h)}{C(0)}$$

- El variograma

$$\gamma(h) = \frac{1}{2} E\{[Z(u) - Z(u+h)]^2\}$$

Además, se dice que un modelo de función aleatoria es estacionario de segundo orden cuando:

1. El valor esperado $E\{Z(u)\}$ existe y es invariante dentro del área A
2. La función de covarianza $C(h)$ existe y sólo depende del vector separación h

Igualmente, la función de covarianza, el correlograma y el variograma de una función aleatoria estacionaria están relacionados de la siguiente manera:

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

$$\rho(h) = 1 - \frac{\gamma(h)}{C(0)}$$

$$C(h) \rightarrow 0 \text{ para } |h| \rightarrow \infty$$

$$\gamma(h) \rightarrow C(0) \text{ para } |h| \rightarrow \infty$$

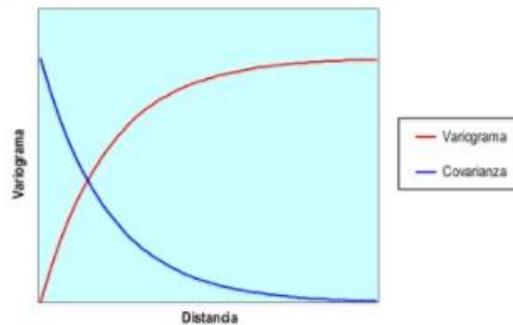


Fig.45. Relación covarianza y variograma.
(Fuente: Slideshare, 2015)

Como se puede observar, la covarianza y el variograma son complementarios, los dos describen lo mismo, la conectividad entre estaciones de medida. Se puede usar una u otra. En la práctica, se introducen los variogramas en el programa informático, pero internamente el programa transforma el variograma en una covarianza.

2.6. Técnicas de estimación

2.6.1. Introducción

Como bien se ha comentado en apartados anteriores, el objetivo de un estudio geoestadístico según Edward H. Isaaks en el libro *“An Introduction to Applied Geostatistics”* (1989), no es solamente describir la información y construir un modelo de continuidad espacial.

La información acerca de los datos y el modelo de continuidad espacial nos permiten realizar la estimación de los valores de una o varias variables donde ellas no fueron muestreadas.

2.6.1.1. Combinación lineal ponderada

Todos los métodos de estimación que serán descritos construyen cada estimador como una combinación lineal ponderada de la información disponible, esto es:

$$\text{Estimador} = z^* = \sum_{i=1}^n \lambda_i \cdot z_i$$

donde el asterisco indica que el valor obtenido es una estimación, z_1, \dots, z_n son los n datos disponibles y λ_i es el peso asignado al valor z_i .

2.6.1.2. Tipos de estimaciones

Existen diferentes aproximaciones para dar valor a los pesos, las cuales dan origen a distintas técnicas de estimación:

2.6.1.2.1. Estimación Global y local

Se habla de estimación global cuando se quiere estimar el parámetro en conjunto de todo el área de estudio. Por ejemplo, se quiere saber el total del metal de una mina (Para ver si es rentable o no su explotación) y lo obtengo como una combinación lineal de las cantidades observadas en los puntos de muestreo.

Una estimación local, sin embargo, es lo que se necesita después, una vez que se ha decidido que la mina es viable (a través de la estimación global), se necesitará una buena estimación de cómo se distribuye espacialmente el mineral (Ya que eso modificará el plan de extracción). En este caso, sí se necesita hacer una estimación de los valores que se van a encontrar en cada punto. Por ello, se habla de una estimación local, se quiere saber el valor de la variable en un punto o una zona determinada.

2.6.1.2.2. Estimación de una media y una distribución completa

La media es el estadístico más común para analizar una distribución de valores.

Presenta problemas para distribuciones muy sesgadas y para aplicaciones donde la simple media aritmética carece de significado como es el caso del cálculo de permeabilidades equivalentes.

Hay muchas aplicaciones que exigen una estimación de la variabilidad de un dado atributo, por ejemplo:

- a. La variabilidad espacial de la conductividad hidráulica es vital a la hora de una correcta descripción del flujo de agua subterránea en un acuífero o de petróleo en un yacimiento.
- b. Un estudio de contaminación de suelos requiere el conocimiento de la concentración de algunos metales pesados por encima o por debajo de determinados umbrales.

Cuando una determinada aplicación requiere la estimación de estadísticos distintos de la media la solución pasa por estimar la distribución completa de valores.

2.6.1.2.3. Estimación puntual y de bloque

También se habla de estimación puntual y de bloque, que están muy relacionadas con la estimación local y global, pero en realidad en estas, lo que hay que tener en cuenta es que los

datos son representativos de un volumen; no es lo mismo la cantidad de metal que hay en un testigo, que la transmisividad que se obtiene después de un ensayo de bombeo de un acuífero, el cual es un volumen relativamente grande.

Este volumen se le llama soporte, por lo tanto, se pueden tener datos de diferentes soportes y con ello hacer estimaciones de nuestra variable en diferentes soportes.

Si el volumen es la totalidad, sería la estimación global; pero también se puede hacer una estimación en una zona específica, este sería un soporte diferente al soporte de los datos. En este caso se hablaría de estimaciones puntuales, que es en lo que se va a centrar este estudio.

2.6.2. Técnicas de estimación de una sola variable

2.6.2.1. El Krigeado ordinario

Cassiraga (2015) dijo que la idea básica del krigeado ordinario es estimar el valor desconocido de un atributo z en el punto de coordenadas u , a partir de n valores conocidos de z , cuyas coordenadas son u_α , con $\alpha = 1, \dots, n$.

El estimador por krigeado ordinario tiene la forma siguiente:

$$z^*(u) = \sum_{\alpha=1}^n \lambda_\alpha z(u_\alpha)$$

donde el $*$ denota que se trata de un estimador y $\lambda_\alpha (u)$ son coeficientes de ponderación desconocidos a priori.

2.6.2.1.1. Objetivos

Los objetivos del Krigeado son obtener los valores de los coeficientes λ_α de manera que:

- El estimador sea insesgado, es decir, que la media del error m_R cometido sobre un conjunto de estimaciones sea igual a cero.
- Que la varianza σ_R^2 de esos errores sea mínima.

El error cometido en la estimación para una localización u se puede definir como:

$$r(u) = z^*(u) - z(u)$$

donde $z^*(u)$ es el valor estimado y $z(u)$ es el valor verdadero.

Por ello se puede definir un error en cada punto de mi área de estudio como la diferencia entre el valor estimado y el valor real, el cual se desconoce. Y se hará que la media del error sea cero y que la varianza de este error sea nula.

El problema suele ser que, en campo, se tienen los valores observados, pero al ir a calcular el error, el valor de z solo se conoce en pocos puntos. En estos, sí que se puede realizar dicha combinación lineal y obtener el error, pero en el resto no se sabe ya que el valor real es desconocido. Por ello, de primeras no se puede hacer nada con ello, a no ser que se construya un modelo, y a partir de las propiedades de ese modelo trabajar con esas premisas.

$$Z^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha}) \quad (1)$$

Por lo que, para poder avanzar con ese problema, es necesario trabajar con un modelo, dicho modelo a utilizar es el modelo de función aleatoria.

2.6.2.1.2. Modelo de función aleatoria

Como bien se ha dicho en el punto anterior, la solución pasa por construir un modelo de nuestros datos y trabajar con el error medio y la varianza para nuestro modelo. Y dicho modelo a construir es un modelo de función aleatoria, de forma que la expresión del estimador por Krigeado ordinario se modifica tal que tiene la siguiente forma:

Por ejemplo, se tiene una realidad que desconozco en la cual he observado tres puntos y entre medias no se sabe que hay. Lo que se quiere es hacer una estimación que tengan un error medio nulo y que esté lo más próximo a la realidad, es decir que el error cuadrático medio sea mínimo. El modelo de función aleatoria, en el cual solo tengo tres puntos de la realidad, se reemplaza por un conjunto de realidades. No se sabe cuál es la realidad, pero se sabe que pertenece a una de ellas. Además, dicho conjunto de realidades se puede construir con una serie de condiciones, por ejemplo, podría ser que la realidad son valores enteros entre 1-10 y que su media es 5. Esa es la condición que me define una función aleatoria, que serían todas las distribuciones que se puedan montar con números naturales entre 1 y 10 que tengan de media 5. Todas esas combinaciones corresponderían a mi función aleatoria.

Entonces, las condiciones que se introducen en la función aleatoria para modelar nuestra realidad son las condiciones de estacionariedad. Entonces lo que se dice es que todas estas realidades virtuales alternativas cumplen dos condiciones:

1. Son estacionarias de primer orden en el sentido que, si en un punto cualquiera del espacio analizo todos los posibles valores de esas realidades alternativas, el valor medio no depende de la posición, el valor medio es contante. Ese valor medio calculado en un punto a través de todas las alternativas es lo que se llama valor esperado de la variable aleatoria en ese punto y, es un valor m que es constante (no depende de u).
2. Condición de estacionariedad de segundo orden que establece que, si se coge dos puntos que están separados una distancia h , la covarianza entre esos dos puntos u y $u+h$ es una función que solo depende de h . Y esa función covarianza de h es una función que se puede definir ya que la media es un valor ahora conocido. Esto, lo que quiere decir, es que la covarianza se puede calcular para dos puntos separados una distancia h , y es un valor que se calcula a través de las realizaciones. Si además la función aleatoria es ergódica, entonces el valor medio calculado en cualquiera de las aplicaciones coincide con ese valor medio que es constante en cada punto, por lo que el valor medio en cualquier realización coincide con el valor obtenido a través de las realizaciones y la covarianza coincide con esa covarianza que es constante a través de las realizaciones.

Una vez establecidas estas condiciones, hay herramientas que me permiten generar estas realidades, es decir, una vez que se dice que se tiene una función aleatoria estacionaria con media m y con covarianza $c(h)$, hay herramientas que son capaces de construir esas realidades alternativas a esa realidad desconocida.

Cuando se usa un modelo de función aleatoria, se pasa de estimar el valor de la variable en el punto objetivo a partir de los datos, a estimarlo en todas esas alternativas. La colección de todos los valores que tenemos en un punto dado sería una variable aleatoria Z , los cuales tendrán una cierta función de distribución, y que se estimará como combinación de las funciones aleatorias en los puntos que corresponderían a las mismas ubicaciones donde se disponen los datos.

2.6.2.1.3. El error en la estimación

Defino una variable error, que ya no será una variable error de un único valor, sino que será una variable aleatoria, porque van a ser los errores cometidos en cada una de esas realidades alternativas. Con ello, ya se puede calcular su error medio y su varianza, haciendo que el error medio sea nulo y la varianza mínima, definiéndose como:

$$R(u) = Z^*(u) - Z(u) \quad (2)$$

donde $Z^*(u)$ y $Z(u)$ son las variables aleatorias que representan el valor estimado y el valor verdadero respectivamente. El error $R(u)$ es también una variable aleatoria.

El error medio m_R cometido en un conjunto de estimaciones:

$$M_R = E\{R(u)\} = E\{Z^*(u) - Z(u)\}$$

y la varianza del error $\sigma_R^2(u)$ de un conjunto de estimaciones:

$$\sigma_R^2 = Var\{R(u)\} = Var\{Z^*(u) - Z(u)\} \quad (3)$$

Como bien se ha indicado antes y de acuerdo a una aproximación probabilística, los objetivos del krigado ordinario, que son obtener los valores de los coeficientes λ_α de manera que la varianza del error cometido en un conjunto de estimaciones sea mínima y el estimador hallado sea insesgado, pueden escribirse matemáticamente como:

$$\text{Min } \sigma_R^2 = \text{min } Var\{R(u)\} = \text{min } Var\{Z^*(u) - Z(u)\} \quad (4)$$

bajo la restricción

$$m_R = E\{R(u)\} = E\{Z^*(u) - Z(u)\} = 0 \quad (5)$$

2.6.2.1.4. La condición de estimador insesgado

La condición de estimador insesgado (ec. 5) se expresa:

$$M_R = E\{R(u)\} = E\{Z^*(u) - Z(u)\} = 0$$

Reemplazando la expresión del estimador (ec. 1) en la del error (ec. 2):

$$E\{R(u)\} = E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha}) - Z(u)\right\} = \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(u_{\alpha})\} - E\{Z(u)\}$$

Asumiendo que la función aleatoria es estacionaria de primer orden, lo que implica que el valor de z_{α} es el mismo independientemente de donde estemos analizándolo y coincidirá con el valor medio del punto donde se está estimando, esto es que

$$E\{Z(u)\} = m(u) = \text{cte } \forall u$$

Y el valor esperado del error queda expresado como

$$E\{R(u)\} = \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z\} - E\{Z\}$$

de donde se deduce que para que

$$E\{R(u)\} = 0$$

se debe verificar que

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1 \quad (6)$$

2.6.2.1.5. La condición de varianza mínima

De forma general, la varianza de una combinación lineal de variables aleatorias se expresa:

$$\text{Var} \left\{ \sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha}) \right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C \{Z(u_{\alpha})Z(u_{\beta})\} \quad (7)$$

donde $C \{Z(u_{\alpha})Z(u_{\beta})\}$ es la covarianza entre las variables aleatorias para las localizaciones u_{α} y u_{β} .

La varianza del error cometido en un conjunto de estimaciones (ec. 3) es una combinación lineal de dos variables aleatorias, por lo que

$$\begin{aligned} \text{Var}\{R(u)\} &= C\{Z^*(u)Z^*(u)\} - C\{Z^*(u)Z(u)\} - C\{Z(u)Z^*(u)\} + C\{Z(u)Z(u)\} = \\ &C\{Z^*(u)Z^*(u)\} - 2C\{Z^*(u)Z(u)\} + C\{Z(u)Z(u)\} \end{aligned} \quad (1)$$

Ahora se derivará cada uno de estos términos:

- Primer término $\rightarrow C\{Z^*(u)Z^*(u)\}$

Este término corresponde a la autocovarianza de $Z^*(u)$, es decir, $\text{Var}\{Z^*(u)\}$.

Utilizando la definición del estimador (ec. 1) tenemos que de acuerdo a la expresión para una combinación lineal de variables aleatorias (ec. 7)

$$\begin{aligned} \text{Var}\{Z^*(u)\} &= \text{Var}\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})\right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C\{Z(u_{\alpha})Z(u_{\beta})\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C\{u_{\alpha} - u_{\beta}\} \end{aligned} \quad (2)$$

- Segundo término $\rightarrow C\{Z^*(u)Z(u)\}$

y se puede escribir como:

$$\begin{aligned} C\{Z^*(u)Z^*(u)\} &= C\left\{\left(\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})\right)Z(u)\right\} \\ &= E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})Z(u)\right\} - E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})\right\} E\{Z(u)\} \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(u_{\alpha})Z(u)\} - \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(u_{\alpha})\}E\{Z(u)\} \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} C\{Z(u_{\alpha})Z(u)\} = \sum_{\alpha=1}^n \lambda_{\alpha} C(u_{\alpha} - u) \end{aligned} \quad (10)$$

- Tercer término $\rightarrow C\{Z(u)Z(u)\}$

Este último término corresponde a la autocovarianza de la variable aleatoria $Z(u)$, y es igual a la varianza de $Z(u)$, que puede expresarse como:

$$C\{Z(u)Z(u)\} = \sigma^2 \quad (11)$$

Reemplazando las ecuaciones (9), (10) y (11) y reordenando en la expresión de la varianza del error (ec. 8) tenemos:

$$\sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C\{u_{\alpha} - u_{\beta}\} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C\{u_{\alpha} - u\} \quad (12)$$

que debe ser minimizada, para lo cual aplicamos el criterio de la primera derivada.

2.6.2.1.6. Minimización de la varianza del error

La expresión a minimizar es la siguiente:

$$\sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C \{u_{\alpha} - u_{\beta}\} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C \{u_{\alpha} - u\} + 2\mu \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right)$$

donde μ es el parámetro de Lagrange.

El último término de esta ecuación no altera la expresión para la varianza de los errores ya que es igual a 0. A partir de esta expresión se deben hallar las $n + 1$ primeras derivadas (n con respecto a los coeficientes de ponderación y una con respecto al parámetro de Lagrange).

Derivando con respecto a μ e igualando el resultado a cero se obtiene la ecuación que corresponde a la condición de estimador insesgado.

Al derivar con respecto a λ_1 como el primer término no depende de λ_1 su derivada es cero.

El segundo término puede expandirse y si despreciamos todos los términos que no incluyen a λ_1 tenemos:

$$\begin{aligned} \frac{\partial \left(\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(u_{\alpha} - u_{\beta}) \right)}{\partial \lambda_1} &= \frac{\partial \left(\lambda_1^2 C(u_1 - u_1) + 2\lambda_1 \sum_{\beta=2}^n \lambda_{\beta} C(u_1 - u_{\beta}) \right)}{\partial \lambda_1} \\ &= 2\lambda_1 C(u_1 - u_1) + 2 \sum_{\beta=2}^n \lambda_{\beta} C(u_1 - u_{\beta}) \\ &= 2 \sum_{\beta=1}^n \lambda_{\beta} C(u_1 - u_{\beta}) \end{aligned} \quad (13)$$

El tercer término contiene un término que es función de λ_1 , luego:

$$\frac{\partial (\sum_{\alpha=1}^n \lambda_{\alpha} C(u_{\alpha} - u))}{\partial \lambda_1} = \frac{\partial (\lambda_1 C(u_1 - u))}{\partial \lambda_1} = C(u_1 - u) \quad (14)$$

Con el último término sucede igual y su derivada se escribe:

$$\frac{\partial (\mu (\sum_{\alpha=1}^n \lambda_{\alpha} - 1))}{\partial \lambda_1} = \frac{\partial (\mu \lambda_1)}{\partial \lambda_1} = \mu \quad (15)$$

De acuerdo a las ecuaciones (13), (14) y (15), la derivada de σ_R^2 queda:

$$\frac{\partial \sigma_R^2}{\partial \lambda_1} = 2 \sum_{\beta=1}^n \lambda_{\beta} C(u_1 - u_{\beta}) - 2C(u_1 - u) + 2\mu$$

que igualando a cero y ordenando se transforma en:

$$\sum_{\beta=1}^n \lambda_{\beta} C(u_1 - u_{\beta}) + \mu = C(u_1 - u)$$

De igual forma se pueden obtener las derivadas con respecto a los demás coeficientes de ponderación λ tal que el resultado es:

$$\begin{aligned} \frac{\partial \sigma_R^2}{\partial \lambda_1} = 0 &\implies \sum_{\beta=1}^n \lambda_\beta C(\mathbf{u}_1 - \mathbf{u}_\beta) + \mu = C(\mathbf{u}_1 - \mathbf{u}) \\ &\vdots \\ \frac{\partial \sigma_R^2}{\partial \lambda_\alpha} = 0 &\implies \sum_{\beta=1}^n \lambda_\beta C(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu = C(\mathbf{u}_\alpha - \mathbf{u}) \\ &\vdots \\ \frac{\partial \sigma_R^2}{\partial \lambda_n} = 0 &\implies \sum_{\beta=1}^n \lambda_\beta C(\mathbf{u}_n - \mathbf{u}_\beta) + \mu = C(\mathbf{u}_n - \mathbf{u}) \end{aligned}$$

2.6.2.1.7. Sistema de Krigeado ordinario

Las n ecuaciones anteriores sumadas a la obtenida al derivar con respecto al multiplicador de Lagrange, que es la condición necesaria para que el estimador sea insesgado, constituyen un sistema de $n+1$ ecuaciones lineales denominado sistema de krigeado ordinario, y se escribe:

$$\left\{ \begin{aligned} \sigma_R^2 &= \sum_{\beta=1}^n \lambda_\beta C\{u_\alpha - u_\beta\} + \mu = C(u_\alpha - u) & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_\beta &= 1 \end{aligned} \right. \quad (16)$$

donde $C(h)$ es la función de covarianza de la función aleatoria $Z(u)$ que se obtiene de la información disponible.

En forma matricial, la ecuación del krigeado queda:

$$\underbrace{\begin{pmatrix} C(\mathbf{u}_1 - \mathbf{u}_1) & \cdots & C(\mathbf{u}_1 - \mathbf{u}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(\mathbf{u}_n - \mathbf{u}_1) & \cdots & C(\mathbf{u}_n - \mathbf{u}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}}_C \cdot \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix}}_\lambda = \underbrace{\begin{pmatrix} C(\mathbf{u}_1 - \mathbf{u}) \\ \vdots \\ C(\mathbf{u}_n - \mathbf{u}) \\ 1 \end{pmatrix}}_D$$

de donde se obtiene que la solución del sistema es:

$$\lambda = C^{-1} \cdot D$$

La matriz C caracteriza la correlación entre los datos que intervienen en la estimación y

el vector D explica la continuidad entre la localización de cada dato con la localización a estimar.

A la vista queda que los pesos de ponderación son proporcionales a D, que es la covarianza entre los datos y el punto a estimar, que de alguna manera mide la proximidad estructural entre los datos y mi punto a estimar. Estos pesos son proporcionales a esta proximidad estructural y lo que hará es favorecer, dar más peso a aquellos puntos que estructuralmente estén más cerca (Se dice "estructuralmente" porque la covarianza no nos mide la distancia euclídea, sino que es función de la correlación espacial)

La matriz C además, es una matriz de redundancia, indica que si se dispone de datos que están muy próximos estructuralmente, quizás tenga menos información que dos datos que no están tan próximos, indican una redundancia en los datos. En este caso, se reducirán los valores asignados a dichos pesos.

2.6.2.1.8. La expresión para la varianza del error

Para conocer el valor mínimo que toma la varianza de los errores cometidos multiplicamos cada una de las n primeras ecuaciones del sistema de ecuaciones del kriging ordinario (ec. 16) por λ_α , obteniendo:

$$\lambda_\alpha \sum_{\beta=1}^n \lambda_\beta C \{u_\alpha - u_\beta\} + \mu = \lambda_\alpha C \{u_\alpha - u\}, \quad \alpha = 1, \dots, n$$

sumando estas n ecuaciones se llega a:

$$\sum_{\alpha=1}^n \lambda_\alpha \sum_{\beta=1}^n \lambda_\beta C \{u_\alpha - u_\beta\} + \sum_{\alpha=1}^n \lambda_\alpha \mu = \sum_{\alpha=1}^n \lambda_\alpha C \{u_\alpha - u\}$$

$$\sum_{\alpha=1}^n \lambda_\alpha \sum_{\beta=1}^n \lambda_\beta C \{(u_\alpha - u_\beta)\} = \sum_{\alpha=1}^n \lambda_\alpha C \{u_\alpha - u\} - \sum_{\alpha=1}^n \lambda_\alpha \mu$$

Dado que la suma de los coeficientes λ_α es igual a 1, el último término es igual a μ quedando:

$$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C \{u_\alpha - u_\beta\} = \sum_{\alpha=1}^n \lambda_\alpha C \{u_\alpha - u\} - \mu$$

que se puede sustituir en la ecuación de la varianza del error (12) obteniendo la siguiente expresión para la varianza de la estimación por kriging ordinario:

$$\sigma_{KO}^2(u) = \sigma^2 - \sum_{\alpha=1}^n \lambda_\alpha C \{u_\alpha - u\} - \mu \quad (17)$$

Se obtiene con ello que la varianza del error es la varianza de la función aleatoria, la cual sería la varianza a priori, pero reducida. Ésta se reduce como una combinación lineal de las covarianzas entre los datos y el punto a estimar. Si hay mucha proximidad entre mis datos y dicho punto, se reducirá mucho, y si hay poca proximidad, poco.

2.6.2.2. Kriging ordinario y efecto de parámetros del modelo de continuidad espacial

Se estudiará a continuación el efecto de los diferentes parámetros del modelo de continuidad espacial (variograma):

1. Sobre la estimación de un solo punto
2. Sobre la estimación de un campo completo

2.6.2.2.1. Efecto de los diferentes parámetros sobre la estimación de un punto

En este ejemplo, se tiene el valor estimado (40,07) y la varianza de la estimación (0,162) del punto rojo, obtenidos a partir de los datos de partida, que serán los valores de los puntos verdes y entre paréntesis el peso que resulta de la estimación por kriging ordinario (Fig. 46).

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{1500}(h)$$

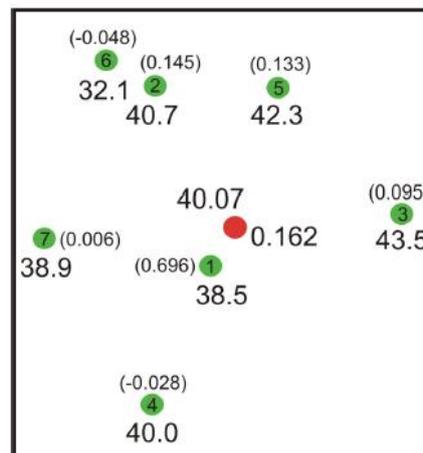


Fig.46. Caso inicial del ejemplo de estimación de un punto.

(Fuente: Cassiraga, 2015)

Inicialmente se dispone de un variograma que es isotrópico, con alcance de 1500 m (alcance mucho mayor que la zona de estudio), por lo que se podría decir que todos los datos están ya correlacionados entre sí, más cuanto más próximos estén. Al ser la correlación isotrópica, el punto más próximo estructuralmente será el punto 1 que será el que recibe el mayor peso. Luego los puntos 7, 5, 3, 4, los cuales están aproximadamente igual. Cabe destacar que los puntos 2 y 6, que están aproximadamente igual de lejos que el 7, tienen una gran diferencia en cuanto al valor de los pesos, y esto se debe a que tenemos una redundancia entre los datos, la suma de los pesos es el equivalente al que se obtendría en uno solo. El efecto de dar más peso al de delante con respecto al de atrás se llama efecto pantalla. El 4 en cambio, tiene un peso negativo. Se ha impuesto que el sumatorio de los pesos sea 1 pero no que los pesos sean

únicamente positivos. Por ello, la posibilidad de que los pesos sean negativos nos permite extrapolar fuera de los datos que tenemos.

EFECTO DE LA MESETA

$$\gamma(h) = 0.0 + 10.0 \cdot \text{Esf}_{1500}(h)$$

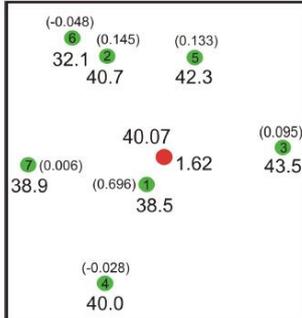


Fig.47. Efecto de aumentar la meseta.
(Fuente: Cassiraga, 2015)

Si se multiplica el variograma por 10, se obtiene el mismo variograma que antes, pero con todos los valores multiplicados por 10. Los pesos son exactamente los mismos, por lo tanto, el estimador es el mismo. Simplemente se escala el variograma por un factor, los pesos quedarán igual ya lo que se multiplica en un caso, se divide en la matriz de redundancia. Lo que si cambia, es la varianza de la estimación que queda aumentada por 10 (Fig. 47).

EFECTO DEL EFECTO PEPITA

El efecto pepita es una pérdida de continuidad a una distancia minúscula. Si se tiene efecto pepita es porque los datos tienen una cierta variabilidad intrínseca para distancias muy cortas, es decir, no se puede esperar mucha continuidad a distancias pequeñas.

$$\gamma(h) = 0.5 + 0.5 \cdot \text{Esf}_{1500}(h)$$

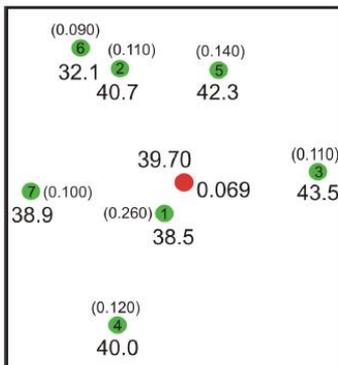
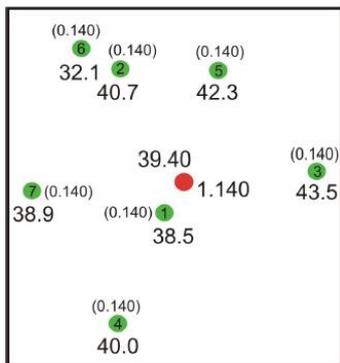


Fig.48. Efecto del aumento del efecto pepita.
(Fuente: Cassiraga, 2015)

En este caso, lo que sucede es, que la suma total de uno y el otro se han separado. El 50% de la continuidad se pierde en los primeros cm y el otro 50% se pierde en los 1500 m (Fig. 48).

$$\gamma(h) = 1.0 + 0.0 \cdot \text{Esf}_{1500}(h)$$

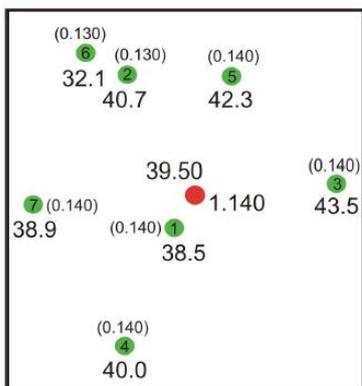


Si el efecto pepita sube a uno, tenemos que toda la continuidad se pierde en los primeros cm. Todos tienen la misma información. Estructuralmente todos estarán igual de lejos. No hay ningún dato que debería recibir más peso que otro (Fig. 49)

Fig.49. Efecto del efecto pepita puro.
(Fuente: Cassiraga, 2015)

EFEECTO DEL ALCANCE

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{100}(h)$$



Si el alcance es 1500 todos los datos están muy correlacionados, en cambio, si el alcance es solo 100, solamente hay dos datos correlacionado con el punto a estimar, todos los demás están a más de 100 m (Fig. 50)

Los dos datos muy cercanos no reciben sus valores originales ya que hay una pequeña redundancia.

El resto recibirán el mismo peso ya que están a más de 100 metros y no hay correlación.

Fig.50. Efecto de disminuir el alcance.
(Fuente: Cassiraga, 2015)

EFECTO DE LA FORMA

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Gau}_{1500}(h)$$

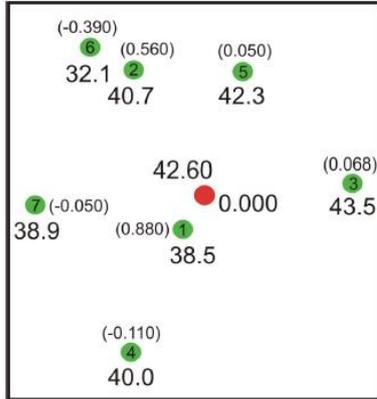


Fig. 51. Efecto de la forma.
(Fuente: Cassiraga, 2015)

El variograma gaussiano, es un variograma que tiene una derivada segunda continua en el origen, el variograma de una distancia muy pequeña es igual a cero. Con este, la pérdida de continuidad es inexistente, es muy continuo. Esta tipología lo que obliga es que se ajuste como un plano entre los datos, dando mucho peso a los datos que están próximos. Y si hay datos que están apantallados, los pesos negativos también serán muy altos, ya que intenta como ajustar algo que pase por todos los datos bajo cualquier situación (Fig. 51)

EFECTO DE LA ANISOTROPÍA

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{1500,150,120}(h)$$

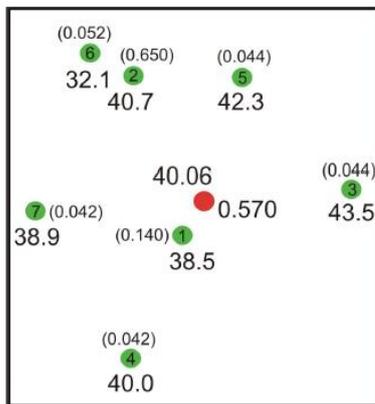


Fig. 52. Efecto de la anisotropía.
(Fuente: Cassiraga, 2015)

En los casos anteriores, la variable era isotrópica, la continuidad era la misma en todas las direcciones. En este caso se hace un estudio en el que se tiene un modelo esférico que tiene un alcance de 1500 en una dirección de 120°. Es decir, en este caso se está diciendo que hay más continuidad a 120° con respecto al eje de las x, es 10 veces más continua que en la dirección perpendicular. Por ello, el punto que está próximo euclídeamente, está distante estructuralmente. Por ello el 2 recibe un peso mucho mayor que el 1 puesto que el variograma es más continuo en una dirección de 120°.

2.6.2.2.2. Efecto de los diferentes parámetros sobre la estimación de un campo completo

En este caso se tendrá la estimación de todo un campo, se está estimando, por ejemplo, el espesor de un yacimiento petrolífero, en el que se tienen medidas del espesor productivo, aquel que contiene petróleo.

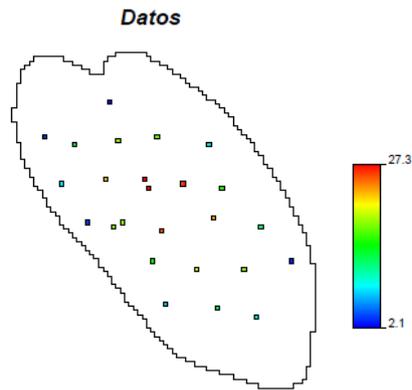


Fig. 53. Datos iniciales del ejemplo de la estimación de un campo completo. (Fuente: Cassiraga, 2015)

CASO INICIAL

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{1500}(h)$$

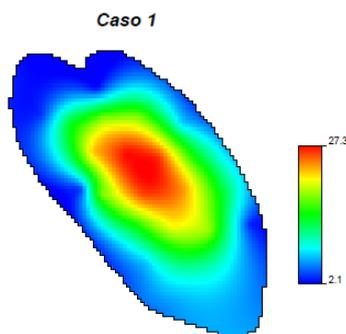


Fig. 54. Caso inicial del ejemplo. (Fuente: Cassiraga, 2015)

Si se hace una estimación en cada punto de una malla superpuesta a este mapa, con un variograma esférico, un alcance de 1500m y sin preferencias en la dirección, se puede observar que los valores estimados a partir de los datos obtiene los valores más altos en el centro y a medida que nos alejamos, el espesor de la capa productiva disminuye.

EFFECTO PEPITA PURO

$$\gamma(h) = 1.0 + 0.0 \cdot \text{Esf}_{1500}(h)$$

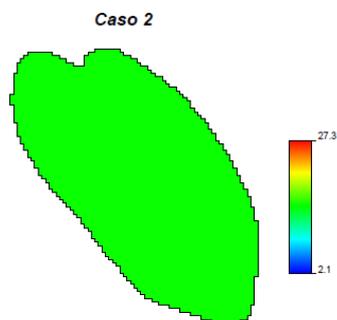
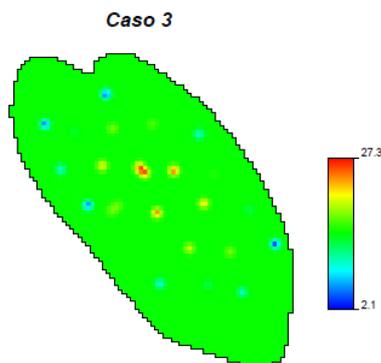


Fig. 55. Efecto del efecto pepita puro. (Fuente: Cassiraga, 2015)

Si hay efecto pepita puro, es decir no hay continuidad, los datos me dan información del punto donde estoy, pero en cuanto se desplazan algo no hay continuidad. No hay información adicional (Fig. 55)

EFFECTO DEL ALCANCE

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{100}(h)$$

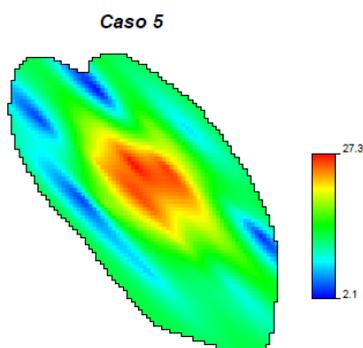


Si hay una pequeña correlación de 100 m en vez de 1500, cada dato tiene solo una pequeña influencia en la estimación de los datos entre los puntos que están próximos. Influye solamente en los puntos que están más próximos, los que están a más de 100 m proporcionan la misma información y el valor estimado sería el valor medio. (Fig. 56).

Fig. 56. Efecto del alcance.
(Fuente: Cassiraga, 2015)

EFFECTO DE LA ANISOTROPÍA

$$\gamma(h) = 0.0 + 1.0 \cdot \text{Esf}_{1500,300,135}(h)$$



Si es anisotrópico a 135°, los datos propagan su información de acuerdo a esta anisotropía (Fig. 57).

Fig. 57. Efecto de la anisotropía.
(Fuente: Cassiraga, 2015)

2.6.2.3. Ejercicio ejemplo de estimación

Calcular los pesos que permitan hacer una estimación.

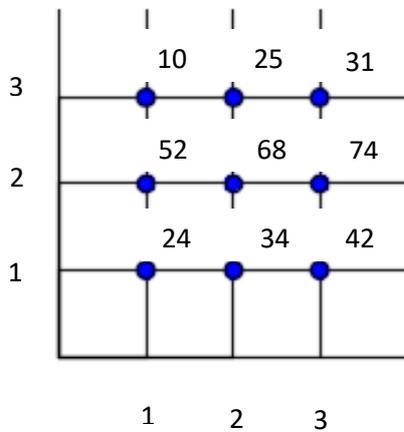


Fig. 58. Ejercicio ejemplo de estimación. Conjunto de datos sobre malla regular. (Fuente: Elaboración propia)

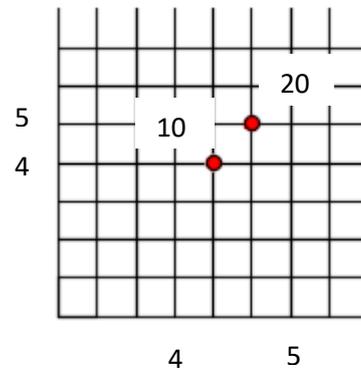


Fig. 59. Ejercicio ejemplo de estimación. Datos disponibles de porosidad (%). (Fuente: Elaboración propia)

De partida se parte de un conjunto de 9 datos en una malla regular. Siendo más continuos aquellos que se encuentran en la dirección horizontal que vertical.

Datos de la porosidad(%):

$U_1 = (4,4) \rightarrow$ porosidad del 10%

$U_2 = (5,5) \rightarrow$ porosidad del 20%

$U_0 = (5,4) \rightarrow$ porosidad????

RESOLUCIÓN:

Para hacer el cálculo usando variogramas:

$$Z^*(5,4) = \lambda_1 \cdot z(4,4) + \lambda_2 \cdot z(5,5)$$

$$\begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \mu \end{Bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & 1 \\ \gamma_{21} & \gamma_{22} & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{Bmatrix} \gamma_{10} \\ \gamma_{20} \\ 1 \end{Bmatrix}$$

Necesitamos conocer esos datos del variograma, $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$

$$\gamma_{11} = \gamma(0,0) = 0 \text{ (vector separación es nulo)}$$

$$\gamma_{22} = \gamma(0,0) = 0$$

En este caso, para calcular el variograma de (1,2) y (2,1) el vector que separa el dato 1 y el 2 es el vector (1,1), se busca todas las parejas en los datos que están separadas el vector 1,1. Obteniendo:

$$\begin{aligned}\gamma_{12} = \gamma(1,1) &= \frac{1}{2N(u=4)} (24 - 68)^2 + (34 - 74)^2 + (52 - 25)^2 + (68 - 31)^2 \\ &= 704\end{aligned}$$

Obteniendo los mismos valores para el caso de γ_{21}

$$\begin{aligned}\gamma_{21} = \gamma(-1, -1) &= \frac{1}{2N(u=4)} (24 - 68)^2 + (34 - 74)^2 + (52 - 25)^2 + (68 - 31)^2 \\ &= 704\end{aligned}$$

Con todo ello, ya hemos obtenido la matriz de redundancia de los variogramas de los datos. Ahora falta calcular la matriz del variograma entre cada uno de los datos y el punto a estimar:

$$\begin{aligned}\gamma_{10} = \gamma(1,0) &= \frac{1}{2N(=6)} (24 - 34)^2 + (34 - 42)^2 + (52 - 68)^2 + (68 - 74)^2 \\ &\quad + (10 - 25)^2 + (25 - 31)^2 = 60\end{aligned}$$

$$\begin{aligned}\gamma_{20} = \gamma(0, -1) &= \frac{1}{2N(=6)} (10 - 52)^2 + (52 - 24)^2 + (25 - 68)^2 + (68 - 34)^2 \\ &\quad + (31 - 74)^2 + (74 - 42)^2 = 680\end{aligned}$$

Según el enunciado se tenía más continuidad horizontal que vertical, luego las magnitudes son correctas.

Con ello ya tenemos todos los elementos que nos permiten calcular todos los valores de los λ , para ello planteamos el siguiente sistema de ecuaciones:

$$(0 \cdot \lambda_1) + (700 \cdot \lambda_1) + \mu = 60$$

$$(700 \cdot \lambda_2) + (0 \cdot \lambda_2) + \mu = 680$$

Resolviendo el sistema de ecuaciones se obtiene:

$$\mu = 20$$

$$\lambda_1 = 0.94$$

$$\lambda_2 = 0.096$$

Esa diferencia de pesos viene inducida a que hay mucha más continuidad horizontal que vertical

Por lo tanto el valor estimado sería:

$$Z^* = (0.96 \cdot 10) + (0.04 \cdot 20) = 104$$

Y la varianza de la estimación:

$$\sigma_{ko}^* = (0.96 \cdot 60) + (0.04 \cdot 680) = 20$$

2.6.3. Técnicas de estimación de una variable incorporando información secundaria

Se va a estimar una variable en un punto o una zona de estudio a partir de los datos que obtenemos acerca de la misma variable. Pero en este caso se va a incorporar una información secundaria, de la cual se dispone multitud de datos, estando correlacionados con la variable principal, la de interés.

Para ello se aprovechará y se hará uso de la correlación cruzada entre las distintas variables con la variable de interés para mejorar el conocimiento acerca de esta última.

En este caso los conceptos de covarianza, variograma... se extiende ya no a comparar parejas de valores de la misma variable, sino a comparar parejas de valores de distintas variables.

El concepto de correlación espacial se mantiene, pero ahora es una correlación cruzada, me está midiendo que información tiene la variable secundaria para mejorar la estimación de la variable primaria.

2.6.3.1. Tipos de datos

Se habla de forma genérica de dos tipos de datos: datos duros y datos blandos.

2.6.3.1.1. Información directa, principal o dura

Cuando se habla de datos duros, "Hard data", se hablan de datos que miden directamente la variable de interés, que son los datos de la variable primaria. Estos datos en general tienen incertidumbre pequeña o despreciable. Por ejemplo, variables primarias típicas suelen ser la transmisividad, la conductividad hidráulica, la concentración de un contaminante, ...

2.6.3.1.2. Información indirecta, secundaria o blanda

Los datos secundarios, indirectos o soft son aquellos que se refieren a las variables secundarias. Son datos que miden indirectamente el parámetro de estudio, pero pueden servir para mejorar esa estimación. Como variable secundaria, suelen utilizarse la porosidad cuando se está estudiando la conductividad hidráulica, información geofísica,.....

2.6.3.2. El Cokrigado

El paradigma del cokrigado es el mismo que el del krigado. Se busca realizar una estimación lineal de nuestra variable a partir de los datos de un estimador insesgado y que tienen una varianza mínima.

Además, este permite incorporar en el proceso de estimación una variable secundaria que no es exhaustivamente conocida.

Desde el punto de vista de la correlación espacial, es la técnica de estimación más potente ya que en la construcción del estimador intervienen los valores de la variable principal, la o las variables secundarias y todos los patrones de continuidad espacial entre variables.

Solamente se tiene en cuenta la correlación dos a dos, es decir, por parejas de variables. No se tiene en cuenta la correlación global con todas las variables. Para ello hay otras técnicas, como por ejemplo la geoestadística multipunto, donde se estudian dos puntos o varias variables

consideradas simultáneamente, pero que en este caso no es el método que se está llevando a cabo.

2.6.3.2.1. El estimador general por cokrigado

En este caso no se va a realizar la derivación de las ecuaciones, pero el proceso a seguir sería básicamente el mismo que se hizo para el krigado ordinario, haciendo que el estimador sea insesgado (valor esperado del error sea nulo) y la varianza de la estimación sea mínima.

A continuación, se muestra la fórmula del cokrigado simple:

$$Z_{CK}^{(1)*}(u) = \sum_{\alpha=1}^{n_1} \lambda_{\alpha_1} [Z_1(u_{\alpha_1}) - m_1(u_{\alpha_1})] + \sum_{i=2}^{N_v} \sum_{\alpha_i=1}^{n_i} \lambda_{\alpha_i} [Z_i(u_{\alpha_i}) - m_i(u_{\alpha_i})]$$

Donde $Z_{CK}^{(1)*}(u)$ es el estimador de la variable principal por cokrigado en la localización (u) , $m_1(u)$ y $m_i(u_{\alpha_i})$ son los valores esperados de las variables aleatorias $Z_1(u)$ y $Z_i(u_{\alpha_i})$ respectivamente, λ_{α_1} y λ_{α_i} son los pesos asignados a los datos primarios y secundarios respectivamente, y n_1 , N_v y n_i son el número de datos de la variable principal, el número de variables secundarias y el número de datos de la variable secundaria que intervienen en la estimación de la localización (u) respectivamente.

2.6.3.2.2. El estimador y la varianza por cokrigado ordinario

El estimador por cokrigado ordinario considerando solo una variable secundaria es:

$$Z_{CKO}^{(1)*}(u) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} Z_1(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} Z_2(u_{\alpha_2})$$

y la varianza de la estimación es:

$$Z_{CKO}^2(u) = \sigma_1^2 - \mu_1 \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} C_1(u_{\alpha_1} - u) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} C_{12}(u_{\alpha_2} - u)$$

2.6.3.2.3. El sistema de cokrigado ordinario

Los coeficientes de ponderación λ_{α_1} y λ_{α_2} se obtienen resolviendo el siguiente sistema de ecuaciones lineales:

$$\left\{ \begin{array}{l} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) + \mu_1 = \\ \quad C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \quad \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_2(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_2}) + \mu_2 = \\ \quad C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}), \quad \alpha_2 = 1, \dots, n_2 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} = 1 \\ \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} = 0 \end{array} \right.$$

2.6.3.2.3.1. EL ESTIMADOR Y LA VARIANZA POR COKRIGEADO ORDINARIO ESTANDARIZADO

El hecho de que se imponga que la suma de los pesos de la variable secundaria sea igual a 0, es como quitarle peso a ésta, lo que hace que no juegue un gran papel en la estimación. Por ello, lo que se propuso fue la opción de estandarizar la media del cokrigeado ordinario.

Lo que se propone es transformar la variable secundaria, conociendo las medias (constante en el espacio, m_1 para la variable primaria y m_2 para la variable secundaria) por una nueva variable (la variable original (Z_2) menos su media (m_2) más la media de la primaria (m_1) que tiene exactamente la misma media que la primaria. Con ello se consigue que el valor esperado de dicha variable sea igual a m_1 , por lo que la nueva variable tiene la misma media que la variable original.

Al tener la misma media las dos variables, la condición de no sesgo que se puede imponer es que la suma de todos los pesos sea igual a 1, no se separan los pesos de la variable primaria y la secundaria.

Aplicando lo anteriormente expuesto, el estimador por cokrigeado ordinario estandarizado considerando sólo una variable secundaria es:

$$Z_{CKO}^{(1)*}(u) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} Z_1(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} [Z_2(u_{\alpha_2}) - m_2 + m_1]$$

y la varianza de la estimación es:

$$Z_{CKO}^2(u) = \sigma_1^2 - \mu_1 \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} C_1(u_{\alpha_1} - u) - \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} C_{12}(u_{\alpha_2} - u)$$

2.6.3.2.3.2. EL SISTEMA DE COKRIGEADO ORDINARIO ESTANDARIZADO

Los coeficientes de ponderación λ_{α_1} y λ_{α_2} se obtienen resolviendo el siguiente sistema de ecuaciones lineales:

$$\left\{ \begin{array}{l} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) + \mu = \\ \quad C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \quad \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_2(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_2}) + \mu = \\ \quad C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}), \quad \alpha_2 = 1, \dots, n_2 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} = 1 \end{array} \right.$$

2.6.3.3. Un ejemplo de cokrigado ordinario

Se quiere estimar la variable primaria u, teniendo valores de la primaria en 1 y 2. Y valores de la secundaria (v) en 1,2 y 3.

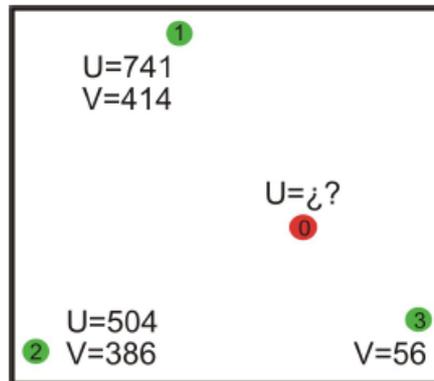


Fig. 60. Datos ejemplo cokrigado ordinario.
(Fuente: Isaaks.E,Srivastaba.R, 1989)

Hay 5 datos, por lo que la matriz de covarianzas será de 5x5 mas las condiciones de no sesgo.

1. Se calculan los variogramas entre u y v, que no es la mismo que entre v y u, por lo que hay que hacer las diferentes combinaciones. El modelo de continuidad espacial que se obtiene tras una inferencia de dichos variogramas es:

$$\begin{aligned}\gamma_u(h) &= 440000 + 70000 \text{Esf}(h) + 95000 \text{Esf}(h) \\ \gamma_{uv}(h) &= 47000 + 50000 \text{Esf}(h) + 40000 \text{Esf}(h) \\ \gamma_v(h) &= 22000 + 40000 \text{Esf}(h) + 45000 \text{Esf}(h)\end{aligned}$$

2. Construcción del sistema de cokrigado

$$\begin{bmatrix} C_{11}^{UU} & C_{12}^{UU} & C_{13}^{UV} & C_{14}^{UV} & C_{15}^{UV} & 1 & 0 \\ C_{21}^{UU} & C_{22}^{UU} & C_{23}^{UV} & C_{24}^{UV} & C_{25}^{UV} & 1 & 0 \\ C_{31}^{UV} & C_{32}^{UV} & C_{33}^{VV} & C_{34}^{VV} & C_{35}^{VV} & 0 & 1 \\ C_{41}^{UV} & C_{42}^{UV} & C_{43}^{VV} & C_{44}^{VV} & C_{45}^{VV} & 0 & 1 \\ C_{51}^{UV} & C_{52}^{UV} & C_{53}^{VV} & C_{54}^{VV} & C_{55}^{VV} & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \eta_1 \\ \eta_2 \\ \eta_3 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} C_{10}^{UU} \\ C_{20}^{UU} \\ C_{30}^{UV} \\ C_{40}^{UV} \\ C_{50}^{UV} \\ 1 \\ 0 \end{bmatrix}$$

$$\lambda = \begin{bmatrix} 605000 & 99155 & 137000 & 49715 & 57615 & 1 & 0 \\ 99155 & 605000 & 49715 & 137000 & 45554 & 1 & 0 \\ 137000 & 49715 & 107000 & 49623 & 57158 & 0 & 1 \\ 49715 & 137000 & 49623 & 107000 & 45164 & 0 & 1 \\ 57615 & 45554 & 57158 & 45164 & 107000 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 134229 \\ 102334 \\ 70210 \\ 52697 \\ 75887 \\ 1 \\ 0 \end{bmatrix}$$

Este es el sistema que resultaría del kriging ordinario sin estandarizar, en el que simplemente se ha puesto que la suma de los pesos de la variable primaria sea 1 y los de la variable secundaria sean 0.

Se sustituyen los valores de cada una de las covarianzas que se calcularían a partir de la expresión del modelo ajustado anteriormente, y finalmente, se invierte, obteniendo así los pesos con los valores de los multiplicadores de Lagrange.

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \mathbf{C}^{-1} \cdot \mathbf{D} = \begin{bmatrix} 0.512 \\ 0.488 \\ -0.216 \\ -0.397 \\ 0.666 \\ 205963 \\ 13823 \end{bmatrix}$$

$$\begin{cases} u_0^* = \sum_{i=1}^{n_1} \lambda_i u_i + \sum_{i=1}^{n_2} \eta_i v_i = 398 \\ \sigma_{CKO}^2 = 681549 \end{cases}$$

El estimador es igual a 398 y la varianza del error 681549.

2.6.3.4. Cokriging colocalizado

Cuando la información secundaria está muy densamente muestreada el algoritmo del cokriging tiene los siguientes inconvenientes:

1. El contraste entre la densidad de muestreo de la variable principal y la secundaria hace que el sistema de ecuaciones de CK tienda a ser muy inestable.
2. Los datos secundarios que están muy cerca de la localización de la variable principal a ser estimada, "apantallan" el efecto de los demás datos blandos.

Una solución a estos problemas es construir un estimador de la variable principal en una dada localización reteniendo sólo el dato "colocalizado" de la variable secundaria, aproximación conocida como cokriging colocalizado.

Este es aplicable solo cuando la variable secundaria es conocida en todos los puntos de la zona de estudio. Por ejemplo, cuando se ésta estudiando el fenómeno de la precipitación y se aprovecha la correlación precipitación-elevación, usando un modelo digital del terreno. Es decir, se tiene información exhaustiva de la variable secundaria, y para estimar la primaria, se va a usar solamente el dato secundario colocalizado, es decir el dato en la misma posición donde se está hallando la estimación.

2.6.3.4.1. Estimador general por cokrigado colocalizado

El estimador general por cokrigado colocalizado tiene la forma siguiente:

$$Z_{CKC}^{(1)*} - m_1(u) = \sum_{\alpha=1}^{n_1} \lambda_{\alpha_1} [Z_1(u_{\alpha_1}) - m_1(u_{\alpha_1})] + \sum_{i=1}^{N_v} \lambda_i [Z_i(u) - m_i(u)]$$

Donde $Z_{CKC}^{(1)*}$ es el estimador de la variable principal por cokrigado colocalizado en la localización (u), $m_1(u)$ y $m_i(u)$ son los valores esperados de las variables aleatorias $Z_1(u)$ y $Z_i(u)$ respectivamente, λ_{α_1} y λ_{α_i} son los pesos asignados a los datos primarios y secundarios respectivamente, y n_1 y N_v son el número de datos de la variable principal y el número de variables secundarias que intervienen en la estimación de la localización (u) respectivamente.

2.6.3.4.2. Cokrigado colocalizado ordinario estandarizado

Se puede hacer la misma estandarización que se ha visto anteriormente en el punto 2.6.3.2.2 para obtener como condición que la suma de los pesos sumen 1, operando entonces de forma similar se obtiene que el estimador por cokrigado colocalizado ordinario estandarizado considerando sólo una variable secundaria es:

$$Z_{CKCO}^{(1)*}(u) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} Z_1(u_{\alpha_1}) + \lambda_2 [Z_2(u) - m_2 + m_1]$$

y el sistema de ecuaciones cuya solución nos da los coeficientes de ponderación es:

$$\begin{cases} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \lambda_2 C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_0) \\ \mu = C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \quad \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u} - \mathbf{u}_{\beta_1}) + \lambda_2 C_2(0) + \mu = C_{12}(0) \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} + \lambda_2 = 1 \end{cases}$$

2.6.3.4.3. Cokrigado colocalizado bajo un modelo de Markov

A través de una aproximación de tipo markoviana la función de covarianza cruzada entre dos variables se puede expresar:

$$C_{12}(h) \cong \frac{C_{12}(0)}{C_{11}(0)} C_{11}(h) = \rho_{12}(0) C_{11}(h)$$

Este nos indica que la covarianza que se tiene para una cierta distancia entre la variable 1 y 2 es proporcional a la covarianza que tiene la variable primaria multiplicada por la correlación que existe entre las dos variables.

Si la correlación es muy buena, la covarianza será muy parecida a la variable 1.

Si la correlación es muy mala, la covarianza será cero, al no haber correlación espacial.

En términos de correlograma

$$\rho_{12}(h) \cong \rho_{12}(0) \rho_{11}(h)$$

o en términos de variograma

$$\gamma_{12}(h) \cong \rho_{12}(0) \gamma_{11}(h)$$

Donde $\rho_{12}(0)$ es el coeficiente de correlación lineal entre las variables.

2.6.3.4.3.1. EL ESTIMADOR Y EL SISTEMA DE COKRIGEADO COLOCALIZADO BAJO UN MODELO DE MARKOV

El estimador y el sistema de ecuaciones del cokrigado colocalizado bajo un modelo de Markov en su forma estandarizada son:

$$\frac{Z_{CKCSM}^{(1)*}(u) - m_1}{\sigma_1} = \sum_{\alpha=1}^{n_1} \lambda_{\alpha_1} \frac{[Z_1(u_{\alpha_1}) - m_1]}{\sigma_1} + \lambda_2 \frac{[Z_1(u) - m_2]}{\sigma_2}$$

$$\left\{ \begin{array}{l} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \lambda_2 \rho_{12}(0) \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_0) = \\ \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \quad \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} \rho_{12}(0) \rho_1(\mathbf{u} - \mathbf{u}_{\beta_1}) + \lambda_2 = \rho_{12}(0) \end{array} \right.$$

Donde $\rho_{12}(0)$ es el coeficiente de correlación lineal entre las variables.

2.6.3.4.3.2. CASOS PARTICULARES

Si $\rho_{12}(0) = 0$ entonces $\lambda_2 = 0$ y el dato secundario colocalizado es ignorado por lo que queda que la estimación del cokrigado coincide con la estimación del krigado ordinario. La variable secundaria no afecta para nada y esto es la estimación del krigado simple, quedando:

$$\frac{Z_{CKCSM}^{(1)*}(u) - m_1}{\sigma_1} = \sum_{\alpha=1}^{n_1} \lambda_{\alpha_1} \frac{[Z_1(u_{\alpha_1}) - m_1]}{\sigma_1}$$

Si la correlación es perfecta, $\rho_{12}(0) = 1$ entonces $\lambda_2 = 1$ y $\lambda_{\alpha_1} = 0 \quad \forall \alpha_1 = 1, \dots, n_1$, por lo que el estimador corresponde al dato secundario colocalizado estandarizado, esto es:

$$\frac{Z_{CKCSM}^{(1)*}(u) - m_1}{\sigma_1} = \frac{Z_2(u) - m_2}{\sigma_2}$$

2.6.4. Evaluación de la incertidumbre local

Los métodos de estimación presentados hasta ahora son aptos para la inferencia de un valor medio de los atributos modelizados en las localizaciones donde no los conocemos.

En este caso, en lugar de estimar el valor esperado, se estimarán todos los posibles valores que serían coherentes con mis datos en el punto donde se hace la estimación y ver qué posibilidad tienen de que esos valores ocurran. Es decir, la estimación de la distribución completa de valores de una variable de interés permite disponer de un modelo de incertidumbre acerca de esa variable.

A partir de un modelo de incertidumbre se pueden derivar diferentes estimadores óptimos según los criterios de optimalidad elegidos y/o realizar una evaluación del riesgo involucrado en un proceso de toma de decisión relativo al medio que está siendo analizado.

Por todo ello, el objetivo de la evaluación de la incertidumbre es la modelización de funciones de distribución de probabilidad.

2.6.4.1. Valor estimado VS función de distribución de probabilidad.

Anteriormente, en el kriging, cuando en un punto no se disponía de información, se buscaba el valor medio, el valor esperado en ese punto de todos los posibles valores. Ahora, de dicha realidad, de la que solo se conocen unos pocos puntos, se extraen unos parámetros estadísticos (media y variograma) y con ellos se construyen todas las posibles distribuciones de valores y, a partir de ellas, se podrá ir analizando en cada punto cual es la distribución de probabilidad de los valores que son coherentes con mis datos y con el modelo de función aleatoria (Fig. 62).

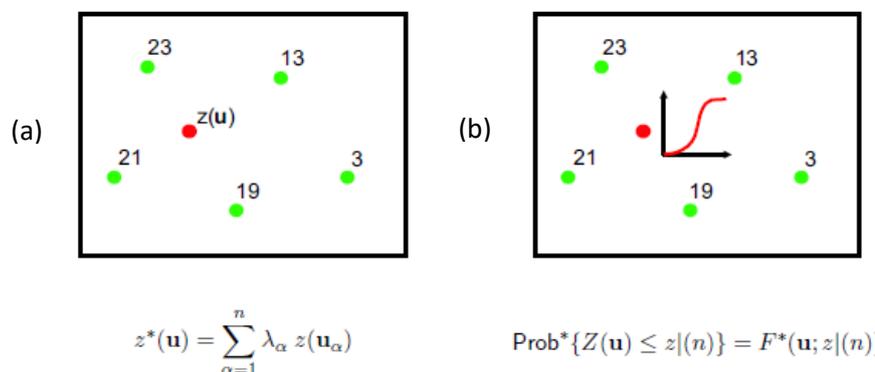


Fig.62. Valor estimado (a) y función de distribución de probabilidad(b). (Fuente: Cassiraga, 2015)

En el kriging se obtenían un valor estimado y una incertidumbre debido a ese error de estimación. Cuando se tiene la función de distribución completa, también se va a obtener la varianza de esa distribución. Por tanto, el valor medio correspondería con el valor del kriging y la varianza de la función de distribución también coincidiría. Hay ocasiones en que la varianza no es lo que interesa obtener, sino obtener un rango intercuartil, saber cuál sería la distribución del 10% de los valores más bajos y el 10 % de los valores más altos, por ejemplo. Esta sería otra forma de estimar la incertidumbre que se tiene sobre la variable, esos intervalos se pueden calcular, cosa que con el kriging no.

Lo que se va a buscar es evaluar dicha incertidumbre, modelizar la posible distribución de probabilidades del atributo a estudiar en el punto a estimar.

2.6.4.2. Utilización de funciones de distribución condicional

El conocer esta función de distribución, la cual variará en el espacio para cada valor de z y condicionada a los datos disponibles, nos dará más versatilidad a la hora de elegir un valor estimado.

Se podrá retener el valor esperado, que será el mismo que el valor del krigeado y se podrá retener la mediana, que estará minimizando el error absoluto medio. También se podrá obtener cuantiles para diferentes valores de probabilidad y las varianzas condicionales.

Además, también se podrá realizar estimación de probabilidades, tanto de excedencia ($Prob\{Z_u > Z_k\}$), como intervalos de probabilidad ($Prob\{Z(u) \in [Z_{k-1}, Z_k]\}$) o valores medios por encima o por debajo de un determinado umbral.

2.6.4.3. Estimación de funciones de distribución

Hay dos formas de realizar las estimaciones: Una es postular a priori cual es la forma de la distribución y después calcular sus parámetros. Método paramétrico. Y la otra opción es calcular el valor de la probabilidad de excedencia para unos cuantos umbrales, y montar una distribución que será por tramos, por puntos.

El paramétrico está basada en un modelo multigaussiano. En la aproximación multigaussina, como bien se ha dicho anteriormente, se asume un modelo a-priori acerca de la forma de las funciones de distribución buscadas. Después, se estiman los parámetros de la distribución de acuerdo al modelo asumido. Dichos parámetros junto con el modelo definido a-priori se utilizan a continuación para derivar probabilidades y/o valores de la variable. Los resultados dependerán fuertemente del modelo asumido.

En cambio, en el método de estimación por aproximación indicadora (método no paramétrico), no se asume un modelo a-priori acerca de la forma de las funciones de distribución buscadas. En esta, se estiman algunos puntos de la distribución los cuales son interpolados y extrapolados convenientemente para obtener la distribución completa. Los resultados en este caso dependerán fuertemente de la extrapolación elegida para conocer la distribución más allá del primer y último punto estimado.

2.6.4.3.1. Aproximación multigaussiana

El modo más fácil de construir las funciones de distribución acumuladas condicionadas es asumir un modelo para la ley espacial (distribución multivariada) de la función aleatoria $Z(u)$.

Esta ley debe ser tal que todas las funciones locales de distribución $F(u; z|(n))$ tengan la misma expresión analítica y estén completamente definidas por unos pocos parámetros.

La solución al problema de estimar las funciones buscadas pasa por determinar los valores de sus parámetros.

El modelo multigaussiano consiste en asumir que la función aleatoria $Z(u)$ que describe la variable de estudio sigue una distribución multinormal, es decir tiene unas características asimilables a una distribución multinormal, esto es:

- Cada distribución univariada sea normal
- Las distribuciones de orden superior corresponden a una distribución multinormal. Esto supone que se deberían comprobar la normalidad bivariada, trivariada, N-Variada de los datos.

2.6.4.3.1.1. COMPROBANDO LA NORMALIDAD

Para comprobar la normalidad simplemente se necesita construir el histograma de los datos y representar sus frecuencias acumuladas sobre un diagrama de probabilidad normal. Si la distribución es normal, las frecuencias acumuladas quedarán representadas gráficamente sobre una línea recta.

2.6.4.3.1.2. TRANSFORMADA NORMAL

Cuando los datos no sigan una distribución normal es necesario realizar una transformación a los mismos que se denomina transformada normal. Tras esto ya se podrá calcular la media y la varianza ya que ya se dispondrá de un modelo de función aleatorio que tenga la media, la varianza y los variogramas de los datos transformados.

Lo que se hace es transformar cada dato individualmente en un nuevo dato, de forma que estos si sigan una distribución gaussiana.

Los n datos originales de la variable z son transformados en los n valores de una variable Y tal que estos tengan un histograma normal estandarizado, es decir que:

$$Y(u_\alpha) = \phi(z(u_\alpha))$$

Donde $\phi(\cdot)$ es la función transformada normal.

En la práctica la transformada normal supone los tres pasos siguientes:

1. Los n datos originales $z(u_\alpha)$ son ordenados de menor a mayor

$$[z(u_\alpha)]^{(1)} \leq \dots \leq [z(u_\alpha)]^{(K)} \leq \dots \leq [z(u_\alpha)]^{(n)}$$

2. Se calcula la frecuencia acumulada experimental de cada dato $z(u_\alpha)$ haciendo:

$$p_k^* = \frac{k - 0.5}{n}$$

Donde k es el número de orden, 0.5 es un valor de centrado y N es el número total de datos

3. Se obtiene la transformada normal del dato $z(u_\alpha)$ según:

$$y(u_\alpha) = G^{-1} [F^*(z(u_\alpha))] = G^{-1} (p_k^*)$$

donde G es la función de distribución normal estándar

Como ejemplo de obtención de la transformada normal, se dispone de los siguientes datos y su interpretación física (Tabla 6 y Fig. 63):

Datos originales (Paso 1)	Frecuencia acumulada (Paso 2)	Transformada normal (Paso 3)
0.135	0.05	-1.6449
0.275	0.15	-1.0364
0.570	0.25	-0.6745
0.910	0.35	-0.3853
1.310	0.45	-0.1257
1.325	0.55	0.1257
1.485	0.65	0.3853
2.120	0.75	0.6745
2.805	0.85	1.0364
3.925	0.95	1.6449

Tabla 6. Datos del ejemplo para la obtención de la transformada normal

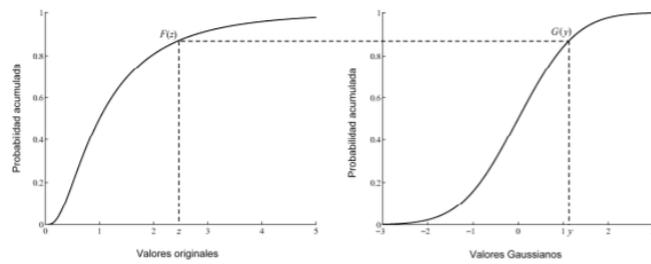


Fig. 63. Transformada de la función de distribución acumulada. (Fuente: Emery, 2013)

Se dispone de la función de distribución acumulada de nuestros datos sin ninguna transformación (gráfica de la izquierda) y en la gráfica situada a su derecha se tiene la función acumulada de distribución de una variable genérica de media 1 y varianza 0 gaussiana. Lo que se hace es para cada dato ver la probabilidad experimental que está asociada y asociarle el valor de la gaussiana que tiene la misma probabilidad.

Esto también se puede hacer al revés ya que una vez hecho esto, se hacen los cálculos en la transformada gaussiana (se calcula la media (que será cero) y el variograma), y una vez acabados estos, se transforma nuevamente a los “originales”.

Una de las ventajas que muestra esta transformada es que resta importancia a sus valores extremos, les quita peso.

El problema se puede dar cuando los datos son iguales, si se tienen datos repetidos. Cuando eso ocurre, sobre todo en minería, se le añade una pequeña perturbación, un pequeño error, permitiendo así diferenciarlos.

2.6.4.3.1.3. COMPROBANDO LA BINORMALIDAD

La transformada normal de la variable Z da como resultado una nueva variable Y que por construcción sigue una distribución (univariada) normal, esto es:

$$Prob \{Y(u) \leq y_p\} = G(u; y_p); \quad \forall y_p$$

donde G(·) es la función de distribución acumulada normal estándar, Y(u) se asume que está estandarizada, esto es, con media cero y varianza unidad.

Esta es una condición necesaria pero no suficiente para que la distribución multivariada de Y sea multinormal.

La siguiente condición necesaria es la que supone que la función de distribución acumulada bivariada de cada par de valores Y(u), Y(u+h), $\forall u, \forall h$, es normal, esto es:

$$Prob \{Y(u) \leq y_p, Y(u+h) \leq y_p'\} = G(h; y_p, y_p'); \quad \forall y_p, y_p'$$

La forma más común de comprobar la bigausianidad es verificando que la función de distribución acumulada bivariada experimental para cada conjunto de pares de datos separados por el mismo vector h , $\{y(u_\alpha), y(u_\alpha + h), \alpha = 1, \dots, N(h)\}$, se ajusta al modelo teórico de una distribución normal bivariada.

2.6.4.3.1.4. LA MULTINORMALIDAD

La normalidad univariada y bivariada son condiciones necesarias pero no suficientes para asegurar la normalidad multivariada, para lo cual deberíamos comprobar la normalidad trivariada, ..., N-variada de la transformada normal de los datos.

El procedimiento a seguir es el mismo que para el caso bivariado, esto es, se procede a comparar frecuencias multivariadas experimentales con las correspondientes frecuencias teóricas derivadas analíticamente.

A pesar de que se han establecido expresiones teóricas para las funciones de distribución acumulada N-variada, es muy difícil realizar una inferencia de las frecuencias multivariadas experimentales.

En la práctica, si se han verificado la normalidad univariada y bivariada, se asume que el formalismo multigaussiano puede ser utilizado.

2.6.4.3.1.5. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO MULTINORMAL

Si se dispone de un modelo multinormal, se deben estimar los parámetros de cada una de las funciones de distribución acumuladas en cada localización u en donde no conocemos el atributo, condicionadas a la información cercana, esto es:

$$G(u; y)|(n) = G\left(\frac{y - E\{Y(u)|(n)\}}{\sqrt{\text{Var}\{Y(u)|(n)\}}}\right)$$

donde $G(\cdot)$ es la función de distribución acumulada normal estándar.

Tal estimación se reduce a inferir los valores de los dos parámetros de una distribución normal (la media y la varianza), los cuales son asimilados al estimador $z_{KS}^*(u)$ y a la varianza $\sigma_{KS}^*(u)$ por krigeado simple de los n datos de la variable en estudio respectivamente, tal que:

$$[G(u; y)|(n)]^* = G\left(\frac{y - y_{KS}^*(u)}{\sigma_{KS}^*(u)}\right)$$

2.6.4.3.1.6. TRANSFORMADA INVERSA

Una vez que los modelos locales de incertidumbre han sido calculados en el espacio de la variable transformada normal Y , es necesario realizar la transformación inversa a la normal para expresarlos en el espacio correspondiente a la variable original Z

Recordemos que un valor z' es transformado en un valor y' correspondiente a su transformada normal, es decir que $y' = \phi(z') = G^{-1}[F(z')]$, donde $G^{-1}(\cdot)$ es la inversa de la función de distribución acumulada normal estándar y $F(z')$ es el valor original de la función de distribución acumulada.

La transformada inversa consiste en que el valor de la función de distribución acumulada condicionada en una localización u se lee a partir de la función de distribución acumulada condicionada normal utilizando el umbral y' sabiendo que el valor de Z debe ser z' y que

$$F(u; z'|n) = G(u; y'|n)$$

2.6.4.3.1.7. UN EJEMPLO DE KRIGEADO MULTIGAUSSIANO

Utilizando la aproximación multigaussiana, inferir las funciones locales de distribución de la profundidad del techo del reservorio.

Datos

Asumir que la función de distribución a-priori de la profundidad es normal con los siguientes parámetros:

Min	Media	Des. est.	Max
2388	2400	7.7	2413

Tabla 7. Datos del ejemplo krigeado multigaussiano (Mínimo, media, desviación estándar y máximo)

El variograma asumido para la profundidad es:

$$\gamma(h) = 60.0 + 1800 \text{Gau}_{a_{max}=2100, a_{min}=900, \alpha=135}(h)$$

El coeficiente de correlación lineal entre la profundidad y la amplitud es 0.7

Primero se considera la información de tres pozos y después se incorporará la información secundaria (amplitud sísmica)

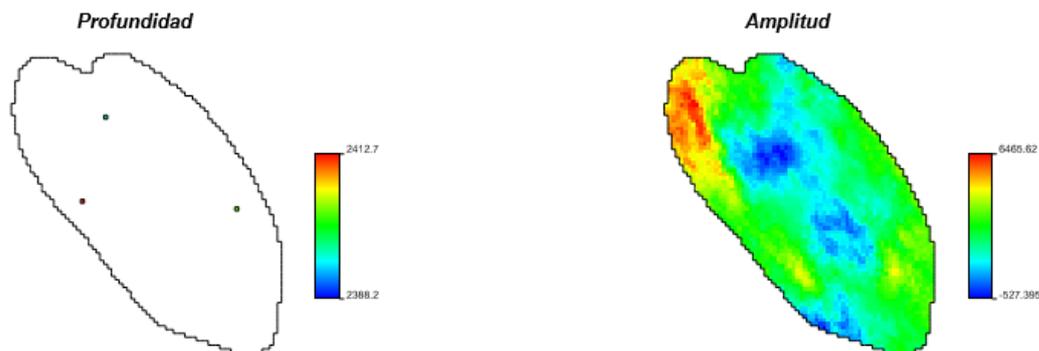


Fig.64. Datos del ejemplo del Krigeado Multigaussiano. Información primaria (a) e información secundaria (b).
(Fuente: Cassiraga, 2015)

Resolución

El variograma es gaussiano, con un alcance de 2100m en la dirección de máxima continuidad, orientada 135° con respecto al eje horizontal en dirección antihoraria, y con un alcance de 900 en la dirección de mínima continuidad. Además el variograma tiene una contribución de 1800m a la variabilidad total del fenómeno y con un efecto pepita de 60.

Para su resolución, se emplea el método del cokrigado colocalizado, donde se considera el punto dado en la estimación y se aprovecha el coeficiente de correlación de 0.7 para poder con el modelo de Markov deducir directamente cuales son las covarianzas cruzadas y las covarianzas de las variables secundarias.

En la primera parte no se utiliza la amplitud, por lo que en cada punto se hace una estimación por krigado simple de la profundidad del techo del reservorio (figura 63). También se puede observar la desviación estándar de la estimación (Figura 64). Si suponemos que hacemos uso de ese modelo multigaussiano, ya tenemos en cada uno de los puntos de estimación la distribución local, que es una distribución gaussiana que tiene esta media y esta varianza.

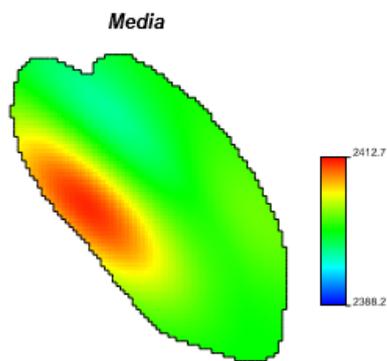


Fig.65. Estimaciones por Krigado simple. Media.
(Fuente: Cassiraga, 2015)

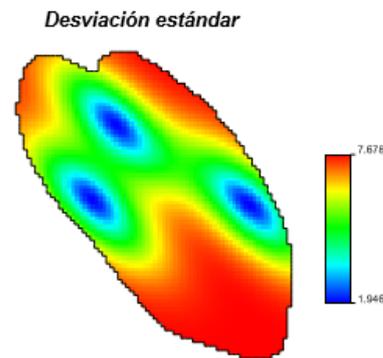


Fig.66. Estimaciones por Krigado simple. Desviación estándar
(Fuente: Cassiraga, 2015)

De estas distribuciones gaussianas se puede rescatar por ejemplo los cuantiles, y ver cuál será el valor mínimo absoluto que puedo observar en cada punto.

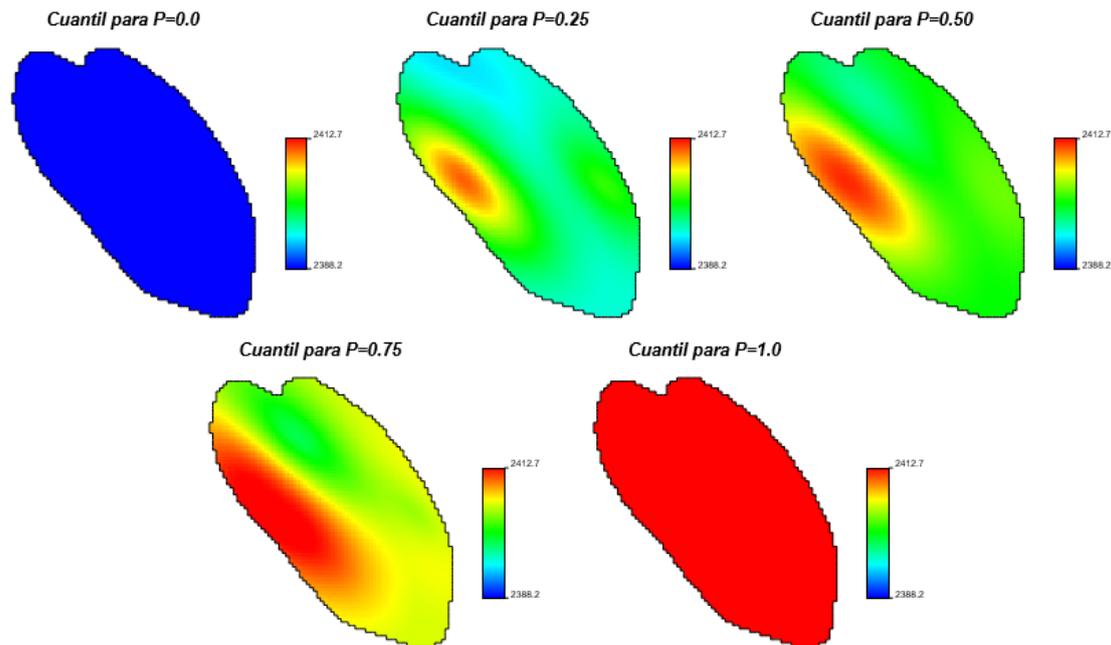


Fig.67. Cuantiles (0, 0.25, 0.5, 0.75, 1) obtenidos mediante KS.
(Fuente: Cassiraga, 2015)

En la figura 67 se puede observar que para un valor de 0,25, valor que está sesgado hacia los valores más pequeños, daría un valor conservador de lo que se está midiendo. Pero si se escoge el 75%, que sería coger de la distribución de gauss los valores que quedan a la derecha del 75%, se ve que tres cuartos son valores por debajo, es decir, tiene un 75 % de probabilidad de que la cota real esté por debajo de ese valor, por lo que éste sería, una estimación optimista de mis reservas, ya que tendría un mayor volumen de ellas.

Para una distribución del 50%, al ser la distribución gaussiana, el cuantil de 0,5 coincide con la media, por lo que ese campo tiene que ser exactamente igual al campo de la media.

También se podría establecer un umbral para ver la probabilidad de que el fenómeno de estudio esté por debajo de un determinado valor. Por ejemplo, en este caso, la probabilidad de que esté por debajo de 2405 es muy alta.

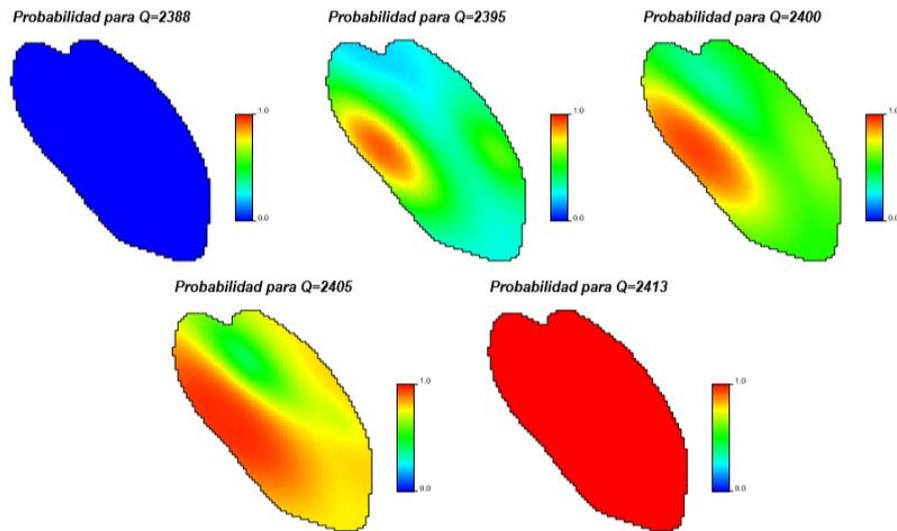


Fig.68. Probabilidades de que el fenómeno se encuentre por debajo de un valor determinado obtenidos mediante KS. (Fuente: Cassiraga, 2015)

Si se hace con cokrigado colocalizado (CKC) usando la amplitud, los valores del valor esperado y la varianza cambian ya que serían de la distribución gaussiana. En este caso, se podrían obtener mapas de estimaciones que serían conservadores (0.25), medios (0.5) y altos (optimista) (0.75).

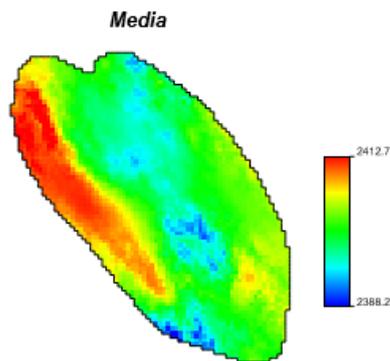


Fig.69. Estimaciones por coKrigado colocalizado. Media. (Fuente: Cassiraga, 2015)

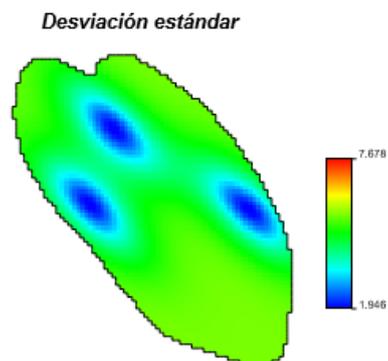


Fig.70. Estimaciones por coKrigado colocalizado. Desviación estándar. (Fuente: Cassiraga, 2015)

Dependiendo del uso que se le vaya a dar se utilizará o bien cuantiles o bien probabilidades. Si tengo un umbral que es mi referencia y necesito saber si se está por encima o por debajo de ese umbral, entonces esas probabilidades me sirven para marcar un área de análisis de donde se tiene que tomar más información.

Siendo los cuantiles calculados mediante CKC:

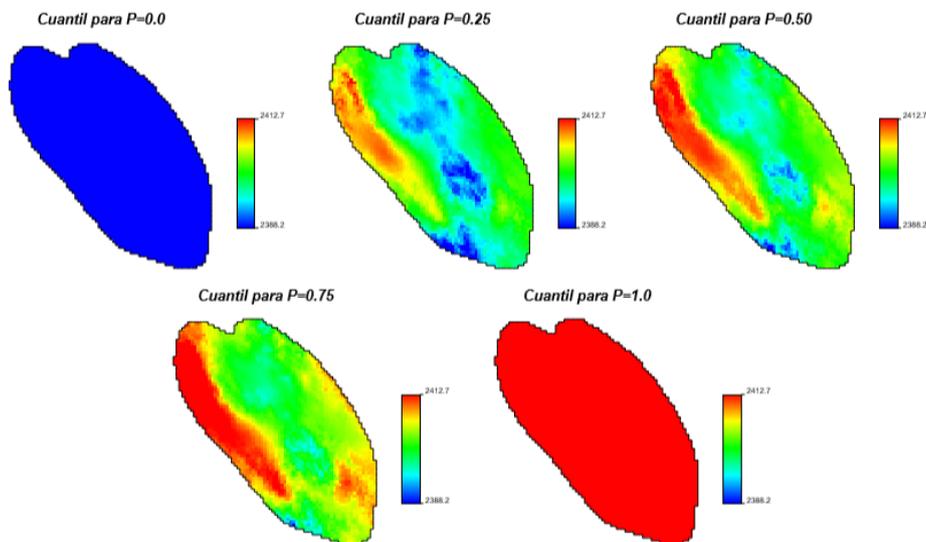


Fig. 71. Cuantiles obtenidos mediante cokrigado colocalizado. (Fuente: Cassiraga, 2015)

Y estableciendo los mismos umbrales anteriores, se observa en la figura 72 las probabilidades de no superación de dicho umbral:

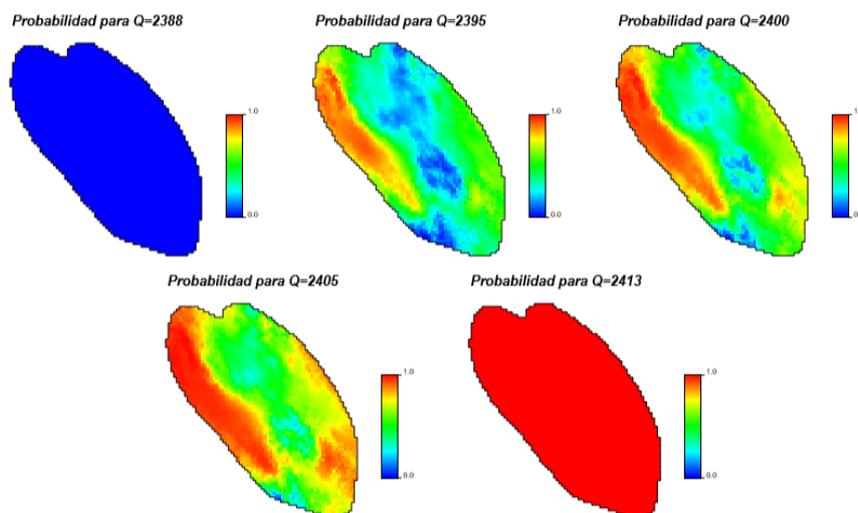


Fig. 72. Probabilidades de que el fenómeno se encuentre por debajo de un valor determinado obtenidos mediante cokrigado colocalizado. (Fuente: Cassiraga, 2015)

A la vista de los resultados anteriores, se observa que si las estimaciones se realizan mediante cokrigado colocalizado, los resultados obtenidos son más precisos que aquellos obtenidos mediante Krigado simple. Esta conclusión tiene coherencia ya que ha sido añadida una segunda variable en el cálculo de la estimación que ayuda a establecer una relación más clara entre los datos.

2.6.4.2.2. Aproximación indicadora

La función indicadora se basa en el conocimiento del valor mínimo y máximo de los datos. En estos se definirán unos umbrales, desde z_1 hasta z_n , y se estimará la probabilidad de que las variables estén por debajo de cada uno de esos valores. A partir de ahí se construirá una función (que no tiene que ser gaussiana) de interpolación entre los datos, que habrá que extrapolar hasta el mínimo y el máximo.

Para ello se utilizará tantas variables auxiliares como umbrales quiero calcular para esa función de probabilidad. Y esas auxiliares son las variables indicadoras.

Las variables indicadoras simplemente consisten en transformar mi variable original en variables indicadoras para tantos umbrales como valores quiero calcular de esta variable de distribución.

¿Qué es lo que se hace cuando se calcula la proporción real de valores por debajo de un umbral a partir de un conjunto de datos exhaustivo?

Para su explicación se considera el siguiente ejemplo (figura 73), en el cual se está interesado en realizar una estimación de la proporción de valores por debajo del umbral -3.5.

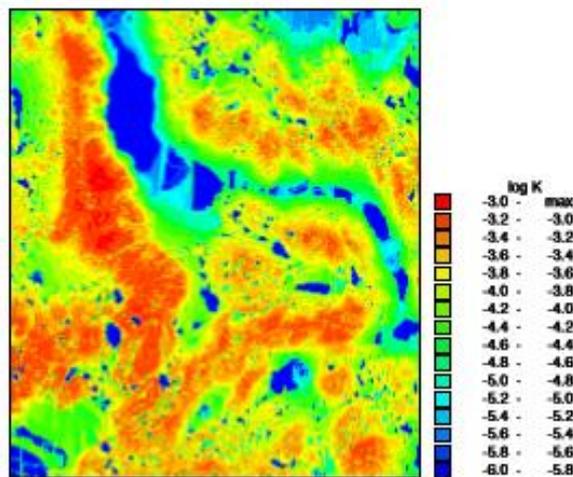


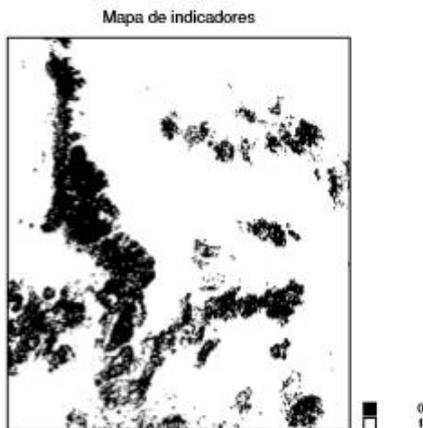
Fig.73. Datos iniciales ejemplo explicativo de estimación de proporción de valores por debajo en un umbral.
(Fuente: Isaaks,E,Srivastaba.R, 1989)

Una sencilla aproximación es contar cuantos datos son menores o iguales que -3.5 y dividir este número por el número total de datos,

$$F(-3.5) = \frac{\text{Número de valores por debajo de } -3.5}{\text{Número total de valores}} = \frac{63335}{78000} = 0.81$$

La idea de conteo puede ser descrita matemáticamente a través de la variable indicadora, esto es:

$$I(\mathbf{u}; z_k) = \begin{cases} 1 & \text{si } Z(\mathbf{u}) \leq z_k \\ 0 & \text{si } Z(\mathbf{u}) > z_k \end{cases}$$



Es decir, se transformará ese mapa de valores continuos, en valores binarios, que será uno si se está por debajo de dicho umbral y cero si se está por encima (el 1 es blanco y el 0 negro). (Figura 74).

La probabilidad de estar por debajo del 3.5 es del 81%. Esto se ha hecho contando, pero se puede hacer directamente si se trabaja en la función indicadora calculando el promedio de esta función:

Fig.74. Mapa de indicadores del ejemplo.
(Fuente: Isaaks.E,Srivastaba.R, 1989)

$$F(-3.5) = \frac{\sum_{j=1}^n i_j}{n} = \frac{63335}{78000} = 0.81$$

De la noción de conteo se ha pasado al concepto de indicador y de la noción de proporción por debajo de un umbral al concepto de indicador medio.

2.6.4.2.2.1. ESTIMANDO UNA DISTRIBUCIÓN ACUMULADA GLOBAL

Este procedimiento puede ser repetido para cualquier umbral z_k de manera que la proporción de valores deseada puede expresarse de una forma general como sigue:

$$F(z_k) = \frac{1}{n} \sum_{j=1}^n i_j(z_k)$$

ϕ	i(5)	i(10)	i(15)	i(20)	i(25)
2.1	1	1	1	1	1
3.7	1	1	1	1	1
8.3	0	1	1	1	1
12.5	0	0	1	1	1
15.0	0	0	1	1	1
18.1	0	0	0	1	1
18.3	0	0	0	1	1
23.1	0	0	0	0	1
27.8	0	0	0	0	0
Media	0.22	0.33	0.55	0.77	0.88

Para el caso de $\overline{i(15)} = 0.55 =$ Valor medio indicadores

= Proporción de datos donde la $\phi \leq 15$

= $Prob\{\phi \leq 15\}$

Fig.75. Valores medios de Indicadores para diferentes umbrales. (Fuente: Isaaks.E,Srivastaba.R, 1989)

La distribución obtenida es una distribución global de los valores de la variable estudiada.

Cuando los datos disponibles no son exhaustivos la fórmula para calcular la proporción de valores por debajo de un dado umbral se escribe:

$$F(z_k) = \sum_{j=1}^n \lambda_j i_j(z_k)$$

donde λ_j son pesos que ponderan cada dato indicador.

El caso de un conjunto de datos exhaustivo corresponde al caso en el cual los pesos asignados a cada dato son todos iguales a $\frac{1}{n}$

2.6.4.2.2. ESTIMANDO DISTRIBUCIONES LOCALES

En muchas aplicaciones prácticas la estimación de una distribución global es insuficiente, por lo que es necesario la inferencia de la distribución de valores sobre áreas pequeñas, esto es, distribuciones locales.

El procedimiento para estimar distribuciones locales se basa en el mismo concepto que el utilizado para inferir distribuciones globales: obtener valores medios de indicadores. Solo que estos valores serán ahora valores medios locales, los cuales se calculan a partir de la información cercana a la localización de la distribución buscada.

Para determinar los valores medios indicadores se pueden utilizar métodos clásicos de estimación local como son el método de los polígonos, de la inversa al cuadrado o el Krigado. Cuando el Krigado es aplicado a la variable indicadora, hablamos de Krigado indicador.

2.6.4.2.3. KRIGADO INDICADOR A TRAVÉS DE UN EJEMPLO

Se va a realizar en la figura 76 una estimación de la probabilidad del valor de la variable en el punto rojo, la cual vale entre 0 y 30.

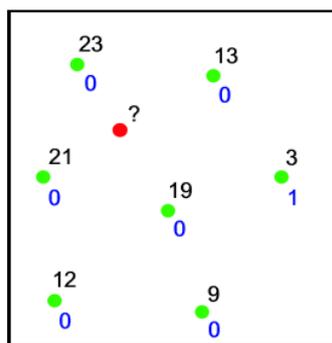


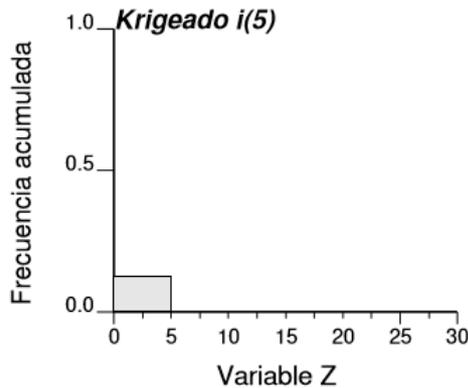
Fig. 76. Datos del ejemplo de krigado indicador.
(Fuente: Cassiraga, 2015)

Para ello se va a estimar la probabilidad de que esté por debajo de 5, 10, 15, 20, 25, 30.

Lo primero será que los datos originales que están en negro los transformamos en indicadores, en este caso, si se comienza con la probabilidad de que esté por debajo de 5, solo hay un dato y todos los demás son ceros ya que son valores mayores

Con los datos transformados, se calcula el variograma. Y con este variograma se hace una estimación de la variable indicadora mediante Krigado simple, de la cual se obtiene su valor esperado, que coincide con la probabilidad de no ser superado.

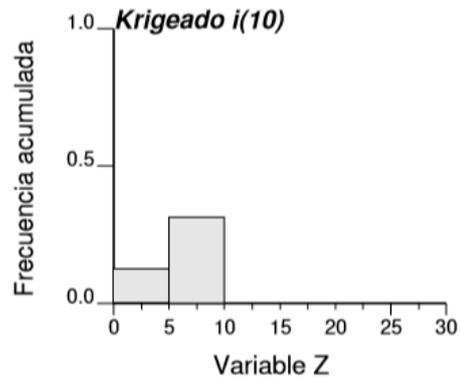
Como resultado se obtiene en la figura 77 que la posibilidad de que la porosidad sea menor que 5% es del 7%, valor bajo esperado ya que básicamente todos son ceros.



$$i^*(5) = \text{Prob}^*\{Z(u_0) \leq 5\} = 0.07$$

O

Fig.77. Probabilidad de que la porosidad sea menor del 5%. (Fuente: Cassiraga, 2015)



$$i^*(10) = \text{Prob}^*\{Z(u_0) \leq 10\} = 0.27$$

Fig.78. Probabilidad de que la porosidad sea menor del 10%. (Fuente: Cassiraga, 2015)

En la figura 78 se toma el siguiente umbral 10. Datos por debajo de 10 solo dos. Por ello se transforman los datos en indicadores, se calcula el variograma de la nueva variable indicadora y se hace una estimación por krigado en ese punto usando los unos y los ceros. Y esa estimación nos dará la probabilidad de que se encuentre por debajo de 10, que en este caso es del 27% (Figura 72)

Se repite lo mismo para el umbral 15,20,25 dando como resultado 50% ,71% y 100% respectivamente.

Con ello se obtiene una distribución de probabilidad genérica, ni gaussiana, ni rectangular ni nada por el estilo. Con ello ya se podrá calcular su valor esperado, su varianza....

2.6.4.2.2.4. LA APROXIMACIÓN INDICADORA PASO A PASO

1. Decidir los umbrales y los valores
2. Transformar los valores originales en tantas variables indicadoras como en umbrales han sido elegidos
3. Calcular los variogramas indicadores para cada umbral
4. Hacer el Krigado indicador para cada umbral
5. Corrección de los problemas de relaciones de orden.

1. SELECCIÓN DEL NÚMERO DE UMBRALES Y SUS VALORES

El número de umbrales seleccionados influye directamente sobre el aspecto final de la distribución inferida.

Cuanto más umbrales tomemos, más puntos de la distribución serán estimados. Pero esto está limitado en la práctica por el número de datos disponibles y por el tiempo necesario para realizar las interpolaciones correspondientes.

El valor de los umbrales puede ser fijado según su significancia física

Cuando no existe algún umbral que tenga un interés práctico particular, un criterio general muy utilizado es tomar como umbrales los nueve valores correspondientes a los deciles de la distribución experimental de la información disponible.

Lo que se busca es tener los umbrales próximos en la zona más probable para intentar resolver esa parte con más precisión. Y un poco más alejados en la zona con menos precisión. A no ser que se esté interesado en los extremos

2.CODIFICACIÓN INDICADORA DE LA INFORMACIÓN

Una vez seleccionados los K umbrales, la información disponible debe ser codificada en vectores de K probabilidades acumuladas:

$$Prob\{Z(u) \leq z_k | \text{información específica local en } u\} \quad k = 1, \dots, K$$

La función de distribución discreta anterior representa la información local acerca de z en u antes de realizar cualquier corrección o actualización considerando la información cercana y se la denomina función local previa.

Dependiendo de la naturaleza de la información local hay diferentes tipos de funciones locales previas, entre ellas las correspondientes a dato duro o dato de intervalo

- Dato duro

Un dato duro $z(u_\alpha)$ es una medida precisa del atributo modelizado, esto es, la incertidumbre acerca de su valor se considera nula o despreciable.

La función de distribución previa local viene dada según la definición de función indicadora como:

$$i(\mathbf{u}_\alpha; z_k) = \begin{cases} 1 & \text{si } z(\mathbf{u}_\alpha) \leq z_k \\ 0 & \text{si no} \end{cases} \quad k = 1, \dots, K$$

de forma que un dato duro está representado por un vector completo de indicadores duros ya que conocemos el valor exacto del dato y los indicadores solo toman los valores 0 o 1.

- Dato de intervalo

Un dato de intervalo $z(u_\alpha)$ es una medida imprecisa del atributo modelizado, esto es, la incertidumbre acerca de su valor no puede ser considerada nula, por ejemplo, debido a errores en la medición del dato.

La información en la localización u_α es modelizada como un intervalo de posibles valores para z tal que $z(u_\alpha) \in (a_\alpha, b_\alpha]$ que equivale a decir que $a_\alpha < z(u_\alpha) \leq b_\alpha$

La función de distribución previa local viene dada según la definición de función indicadora como:

$$i(u_\alpha; z_k) = \begin{cases} 1 & \text{si } b_\alpha \leq z_k \\ \text{no definida} & \text{si } z_k \in (a_\alpha, b_\alpha] \\ 0 & \text{si } a_\alpha > z_k \end{cases} \quad k = 1, \dots, K$$

de forma que un dato de intervalo está representado por un vector incompleto de indicadores duros ya que no conocemos el valor exacto del dato pero los indicadores sólo toman los valores 0 o 1.

3. CÁLCULO DE LOS VARIOGRAMAS INDICADORES

El krigado indicador requiere caracterizar la continuidad espacial de la variable indicadora.

Para cada uno de los umbrales seleccionados es necesario obtener un variograma lo cual hace que el krigado ordinario conlleve una tarea de modelización más ardua que los demás algoritmos de estimación

En contrapartida es una técnica muy potente ya que permite modelizar de forma diferente los patrones de continuidad espacial dentro de cada uno de los rangos de valores de la variable establecidos por los umbrales seleccionados.

Los variogramas indicadores experimentales no sufren tantos problemas como los no indicadores a causa de los valores erráticos, ya que la variable indicadora sólo puede valer 0 o 1.

Los modelos válidos para ajustar a los variogramas indicadores experimentales son los mismos que para los variogramas correspondientes a variables no indicadoras

4. KRIGEADO INDICADOR

Una vez que todas las distribuciones previas locales han sido construidas, el objetivo es inferir el conjunto de funciones de distribución acumuladas condicionadas locales posteriores en cada localización u , esto es:

$$F(u; z_k | (n)) = Prob\{Z(u) \leq z_k | (n)\} \quad k = 1, \dots, K$$

Estas funciones están condicionadas a la información $|(n)$ dentro de un vecindario $W(u)$ centrado en la localización u que está siendo estimada, y está integrada por indicadores $i(u_\alpha; z_k)$.

Estimar las probabilidades posteriores en una localización u supone realizar el krigado de los indicadores disponibles.

Al igual que en el caso de las técnicas de estimación no indicadoras existen varios tipos de krigado según la forma de la media indicadora y según el tipo de información que interviene en el proceso de estimación.

El estimador por krigado indicador

El estimador general por krigado tiene la forma siguiente:

$$Z^*(u) - m(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) [Z(u_\alpha) - m(u_\alpha)]$$

El estimador anterior expresado en términos de la variable indicadora es:

$$[I(u; z_K)]^* - E\{I(u; z_K)\} = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u; z_K)[I(u; z_K) - E\{I(u; z_K)\}]$$

donde $\lambda_{\alpha}(u; z_K)$ son los pesos asignados a cada dato indicador $i(u; z_K)$ interpretado como el resultado de una variable aleatoria $I(u; z_K)$.

Krigeado indicador ordinario

El krigeado indicador ordinario asume que la media indicadora es desconocida y puede fluctuar dentro de un entorno $W(u)$ alrededor de la localización estimada, esto es:

$E\{I(u'; z_k)\} = \text{constante pero desconocida } \forall u' \in W(u)$

El estimador por krigeado indicador ordinario se escribe:

$$[I(u; z_K)]_{KIO}^* = \sum_{\alpha=1}^n \lambda_{\alpha}(u; z_K) I(u; z_K)$$

El sistema de ecuaciones del krigeado indicador ordinario es:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C_I(u_{\alpha} - u_{\beta}; z_K) + \mu(u; z_K) = C_I(u_{\alpha} - u; z_K), & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}(u; z_K) = 1 \end{cases}$$

5. CORRECCIÓN DE LOS PROBLEMAS DE RELACIONES DE ORDEN

Las probabilidades posteriores $[F(u; z_k | (n))]^*$ estimadas en cualquier localización u deben estar entre 0 y 1 y ser una serie de valores no decrecientes, esto es:

$$[F(u; z_k | (n))]^* \in [0, 1]$$

$$[F(u; z_k | (n))]^* \leq [F(u; z_k' | (n))]^* \quad \forall z_k' > z_k$$

Dado que las probabilidades anteriores son calculadas una independiente de la otra muchas veces las relaciones de orden no se cumplen por lo que deben ser corregidas antes de utilizar los modelos de incertidumbre construidos.

Las dos razones principales que pueden ser causa de este tipo de violaciones son la ocurrencia de pesos negativos del krigeado y la falta de datos en alguno de los umbrales seleccionados

2.7. Las técnicas de simulación

A través de las técnicas de krigeado multigaussiano y krigeado indicador se pueden construir modelos de incertidumbre local que son funciones de distribución acumuladas condicionadas específicas para cada localización.

En muchas aplicaciones de las ciencias del terreno es necesario una evaluación de la incertidumbre acerca de los valores de un atributo en muchas localizaciones tomadas conjuntamente. A este tipo de incertidumbre se la denomina incertidumbre espacial.

La evaluación de la incertidumbre espacial se lleva a cabo a través de la generación de múltiples realizaciones de la distribución espacial del atributo de interés, proceso que se denomina representación estocástica o simplemente simulación.

Las representaciones generadas se utilizan luego como entradas a funciones de transferencia y estudiando la distribución de los valores de respuesta.

2.7.1. Estimación versus simulación

Los algoritmos de estimación dan como resultado una única representación del atributo interpolado, esto es, la distribución espacial desconocida de un parámetro z es representada por un único grupo de valores estimados $\{z^*(u), u \in D\}$, siendo D el dominio del problema. Cada valor estimado $z^*(u)$ considerado independientemente de sus valores vecinos es el “mejor” estimador posible de acuerdo a un determinado criterio de optimalidad.

En cambio, se llama simulación estocástica al proceso de generar múltiples campos de una variable z modelizada como una función aleatoria $Z(u)$, esto es $\{z^l(u), u \in D\}$, siendo D el dominio del problema y con $l = 1, \dots, L$ denotando la l -ésima realización de los valores de z .

Considerada como un todo, una imagen estimada hace honor a los datos experimentales, pero no al histograma experimental o a las características experimentales de continuidad espacial, mientras que las imágenes generadas en la simulación sí que lo hacen a ambas tres (Figura 79).

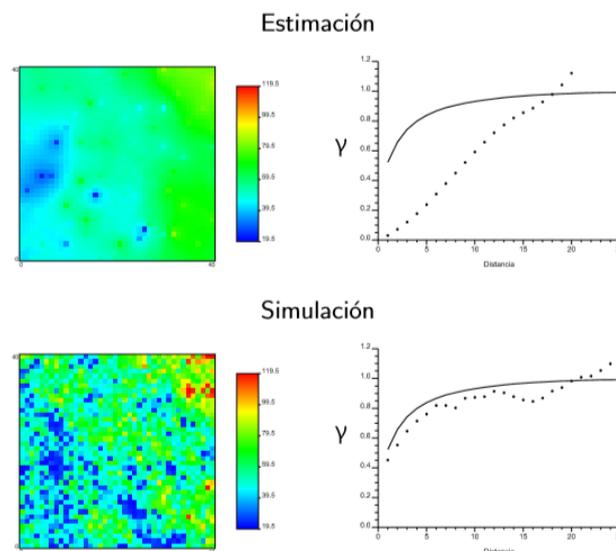


Fig.79. Estimación versus simulación. (Fuente: Cassiraga, 2015)

Los algoritmos de estimación, además, tienden a suavizar la variación espacial del atributo interpolado, es decir, los valores pequeños son sobreestimados y los altos subestimados. Además, este suavizado no es uniforme, es mínimo cerca de los datos y se acentúa cuando nos alejamos de ellos, en consecuencia, un mapa interpolado presenta una mayor variabilidad en las áreas más densamente muestreadas, por ello no son óptimos para aplicaciones sensibles a la presencia de valores extremos y a sus patrones de continuidad espacial.

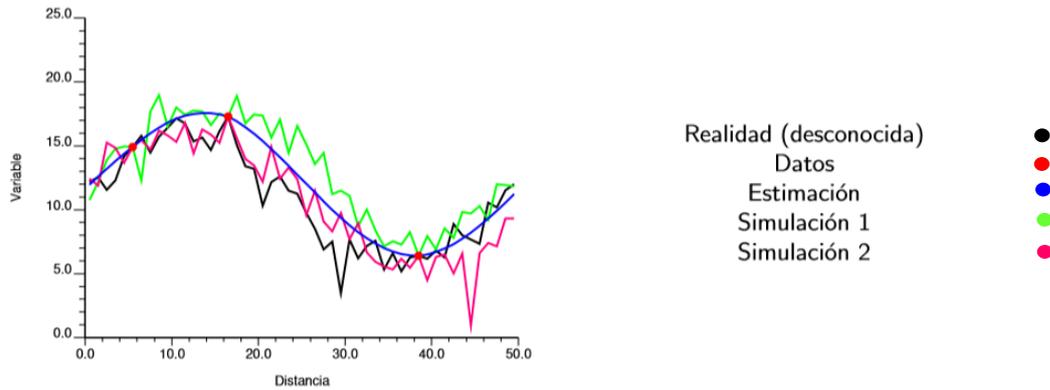


Fig. 80. Estimación versus simulación. Suavizado de la estimación. (Fuente: Cassiraga, 2015)

En cuanto a la evaluación de la incertidumbre se observa en la figura 80 los algoritmos de estimación permiten construir modelos locales de incertidumbre, en cambio, el conjunto de múltiples imágenes de la realidad desconocida generadas por simulación estocástica provee una medida visual y cuantitativa de la incertidumbre espacial del atributo modelizado.

2.7.2. Simulación secuencial

En el libro *An Introduction to applied Geostatistics* (1989), describe que la simulación secuencial es una técnica de simulación muy poderosa, además, sus bases teóricas son fáciles de comprender y pueden ser adaptadas para resolver diferentes problemas de simulación.

2.7.2.1. Teoría

Se considera que $Z(u_j), j = 1, \dots, N$ es un conjunto de variables aleatorias definidas en N localizaciones u_j dentro del área de estudio A . Entonces el objetivo es generar realizaciones conjuntas de estas N variables aleatorias condicionadas al conjunto de datos $z(u_\alpha), \alpha = 1, \dots, n$, es decir:

$$z^{(l)}(u_j), j = 1, \dots, N \quad l = 1, \dots, L$$

Tal generación equivale a muestrear la función de distribución acumulada condicional N -variada siguiente:

$$F(u_1, \dots, u_N; z_1, \dots, z_N | (n))$$

La simulación conjunta de los valores de Z en dos localizaciones u_1 y u_2 , es decir:

$$z^{(l)}(u_1), z^{(l)}(u_2), \quad l = 1, \dots, L$$

puede llevarse a cabo muestreando la función de distribución acumulada condicional bivariada

$$F(u_1, u_2; z_1, z_2 | (n) = \text{Prob}\{Z(u_1) \leq z_1, Z(u_2) \leq z_2 | (n)\}$$

Aplicando el teorema de Bayes la función de distribución acumulada condicional bivariada queda:

$$F(u_1, u_2; z_1, z_2 | (n) = F(u_2; z_2 | (n + 1))F(u_1; z_1 | (n))$$

donde $|(n + 1)$ denota el condicionamiento a los n datos $z(u_\alpha)$ más el valor generado de $Z(u_1)$, es decir, $z^{(l)}(u_1)$.

La aplicación recursiva del teorema de Bayes permite expresar la función de distribución acumulada condicional N-variada como un producto de N funciones de distribución acumuladas condicionales univariadas, es decir:

$$F(u_1, \dots, u_N; z_1, \dots, z_N | (n)) = F(u_N; z_N | (n + N - 1)) \cdot F(u_{N-1}; z_{N-1} | (n + N - 2)) \dots \dots$$

$$\dots F(u_2; z_2 | (n + 1)) \cdot F(u_1; z_1 | (n))$$

donde, por ejemplo, $F(u_N; z_N | (n + N - 1))$ es la función de distribución acumulada condicional dados los n datos y las N - 1 realizaciones de Z, es decir, $z^{(l)}(u_j), j = 1, \dots, N - 1$

2.7.2.2. Algoritmo de simulación secuencial

Para generar una realización del vector aleatorio $Z(u_j), j = 1, \dots, N$ vía simulación secuencial seguiremos los siguientes N pasos:

1. Generación de un valor $z(u_1)$ a partir de la función de distribución acumulada condicional de $Z(u_1)$ dados los valores de los datos (n)
2. Generación de un valor $z(u_2)$ a partir de la distribución condicional de $Z(u_2)$ dado que $Z(u_1) = z(u_1)$ y dados los valores de los datos (n)
- ⋮
- N. Generación de un valor $z(u_N)$ a partir de la distribución condicional de $Z(u_N)$ dado que $\{ Z(u_1) = z(u_1), Z(u_2) = z(u_2), \dots, Z(u_{N-1}) = z(u_{N-1}) \}$ y dados los valores de los datos (n).

Todo ello se obtendrá de la siguiente forma:

A. Seleccionar aleatoriamente una localización donde la variable aún no ha sido simulada.

B. Construir la función local de distribución de probabilidades condicionada a los datos disponibles y a los nodos previamente simulados.

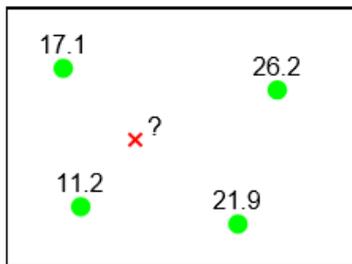


Fig.81. Paso A. Selección localización.
(Fuente: Cassiraga, 2015)

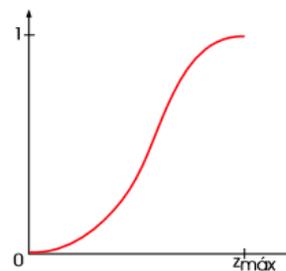


Fig.82. Paso B. Construcción función local de distribución de probabilidades condicionada.
(Fuente: Cassiraga, 2015)

C. Muestrear la función local de distribución de probabilidades condicionada para obtener un valor de la variable

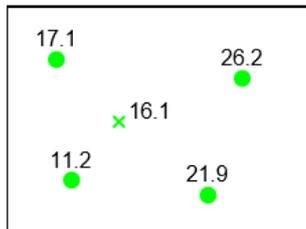


Fig.83. Paso C. Obtención del valor de la variable. (Fuente: Cassiraga, 2015)

D. Asignar el valor generado al nodo siendo simulado y en adelante este valor será tratado como un dato para la simulación de los nodos restantes.

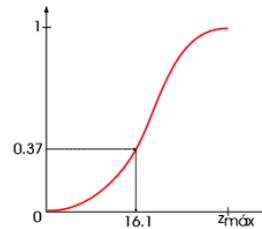


Fig.84. Paso D. Asignar el valor al nodo simulado. (Fuente: Cassiraga, 2015)

E. Repetir la secuencia anterior hasta que todos los nodos han sido simulados.

2.7.2.3. Distribuciones condicionales

La función de densidad de probabilidad condicional de una sola variable dado un grupo de variables aleatorias condicionantes puede ser construida para algunos modelos multivariados:

- Modelo multigausiano o paramétrico que da origen a la técnica de simulación gaussiana secuencial:
 - **Simulación de un sólo atributo.**
 1. Comprobar que la elección de un modelo multigausiano es apropiada, lo cual supone primero calcular la transformada normal de los datos y segundo verificar la normalidad bivariada de los datos transformados.
 2. Aplicar el algoritmo de simulación secuencial a los datos transformados teniendo en cuenta que en cada nodo los parámetros de la función de distribución acumulada condicionada son determinados realizando un krigado simple y asimilando su media y su varianza a los de una distribución normal
 3. Una vez que todos los nodos han sido simulados es necesario realizar la transformación inversa a la normal lo cual nos devuelve al espacio de los datos originales.
 4. Una nueva simulación se puede conseguir repitiendo los pasos 2 y 3 pero siguiendo un camino aleatorio diferente.
 - **Simulación de un atributo incorporando información secundaria.**
 1. Comprobar que la elección de un modelo multigausiano es apropiada, lo cual supone primero calcular la transformada normal de los datos de cada variable y segundo verificar la normalidad bivariada de cada uno de estos conjuntos de datos transformados.
 2. Aplicar el algoritmo de simulación secuencial a los datos transformados teniendo en cuenta que en cada nodo los parámetros de la función de distribución acumulada condicionada son determinados realizando un cokrigado simple y asimilando su media y su varianza a los de una distribución normal.

3. Una vez que todos los nodos han sido simulados es necesario realizar la transformación inversa a la normal lo cual nos devuelve al espacio de los datos originales.
 4. Una nueva simulación se puede conseguir repitiendo los pasos 2 y 3 pero siguiendo un camino aleatorio diferente
- **Simulación conjunta de varias variables.**
 1. Comprobar que la elección de un modelo multigaussiano es apropiada, lo cual supone primero calcular la transformada normal de los datos de cada variable y segundo verificar la normalidad bivariada de cada uno de estos conjuntos de datos transformados
 2. Aplicar el algoritmo de simulación secuencial a los datos transformados de idéntica manera que para simular una variable, sólo que en lugar de simular un valor en cada nodo, se simula un vector con un número de componentes igual al número de variables principales. Cada nodo simulado se condiciona a los datos disponibles más cercanos de la variable a simular, de las demás variables principales, de las variables secundarias, y a los valores previamente simulados.
 3. Concluida la simulación es necesario realizar la transformación inversa a la normal lo cual nos devuelve al espacio de los datos originales.
 4. Una nueva simulación se puede conseguir repitiendo los pasos 2 y 3 pero siguiendo un camino aleatorio diferente.
 - Modelo no paramétrico que da origen a la técnica de simulación indicadora secuencial:
 - **Simulación de un sólo atributo.**

La simulación indicadora secuencial de un atributo continuo z en N nodos de una malla u' condicionada a los datos $\{z(u_\alpha), \alpha = 1, \dots, n\}$ procede como sigue:

1. Discretizar el rango de variación de z en $(K + 1)$ clases a través de K valores umbrales z_k .
2. Transformar cada dato $z(u_\alpha)$ en un vector de indicadores duros (ceros y unos) definidos a partir de:

$$i(\mathbf{u}; z_k) = \begin{cases} 1 & \text{si } z(\mathbf{u}_\alpha) \leq z_k \\ 0 & \text{si } z(\mathbf{u}_\alpha) > z_k \end{cases}$$

3. Definir un camino aleatorio que visite cada nodo de la malla sólo una vez.
4. En cada nodo u' :
 - a. Determinar los K valores de la función de distribución acumulada condicionada haciendo un krigeado simple de los indicadores para cada uno de los K umbrales.
 - b. Corregir las posibles desviaciones en las relaciones de orden e interpolar y extrapolar para determinar un modelo completo de función de distribución acumulada condicionada.
 - c. Generar un valor simulado a partir de la función de distribución acumulada condicionada.
 - d. Agregar el valor simulado al conjunto de datos condicionantes.
 - e. Pasar al siguiente nodo según el camino aleatorio establecido y repetir los pasos 1 a 4

Para generar otra realización repetimos el algoritmo con un camino aleatorio diferente a partir de C.

2.7.2.4. Consideraciones acerca de la simulación

2.7.2.4.1. Reproducción de los estadísticos del modelo

Las realizaciones generadas vía simulación reproducen los estadísticos del modelo en valor esperado.

Las discrepancias entre los estadísticos de cada realización y los impuestos a través de un modelo se denominan fluctuaciones ergódicas. Éstas están influenciadas por:

- El algoritmo utilizado para generar el conjunto de realizaciones: exceptuando el recocido simulado, las técnicas de simulación reproducen el modelo de variograma impuesto en valor esperado, es decir, en media sobre un gran número de realizaciones.
- La densidad de los datos condicionantes: las fluctuaciones ergódicas disminuyen si el número de datos condicionantes aumenta y los estadísticos impuestos fueron modelizados a partir de esos datos.
- Los parámetros del variograma y el tamaño de la malla simulada: si el rango del variograma impuesto es grande con respecto al tamaño de la malla simulada las fluctuaciones ergódicas son mayores.

2.7.2.4.2. Visualización de la incertidumbre espacial

El conjunto de realizaciones generadas por simulación estocástica provee una medida de incertidumbre acerca de la distribución espacial de los valores del atributo estudiado.

La incertidumbre espacial puede visualizarse a través de:

- Animaciones: visualización de las realizaciones en una rápida sucesión.
- Mapas de probabilidad: en cada nodo simulado la probabilidad de exceder un determinado umbral se calcula como la proporción de valores que están por encima de ese umbral considerando el conjunto de realizaciones.
- Mapas de cuantiles: en lugar de mostrar la probabilidad de excedencia se muestran los cuantiles para un determinado valor de probabilidad.
- Mapas de dispersión: representación gráfica de alguna medida de dispersión (varianza, rango intercuartil, etc.) para cada localización a través del conjunto de realizaciones.

Un ejemplo se puede observar en la figura 85.

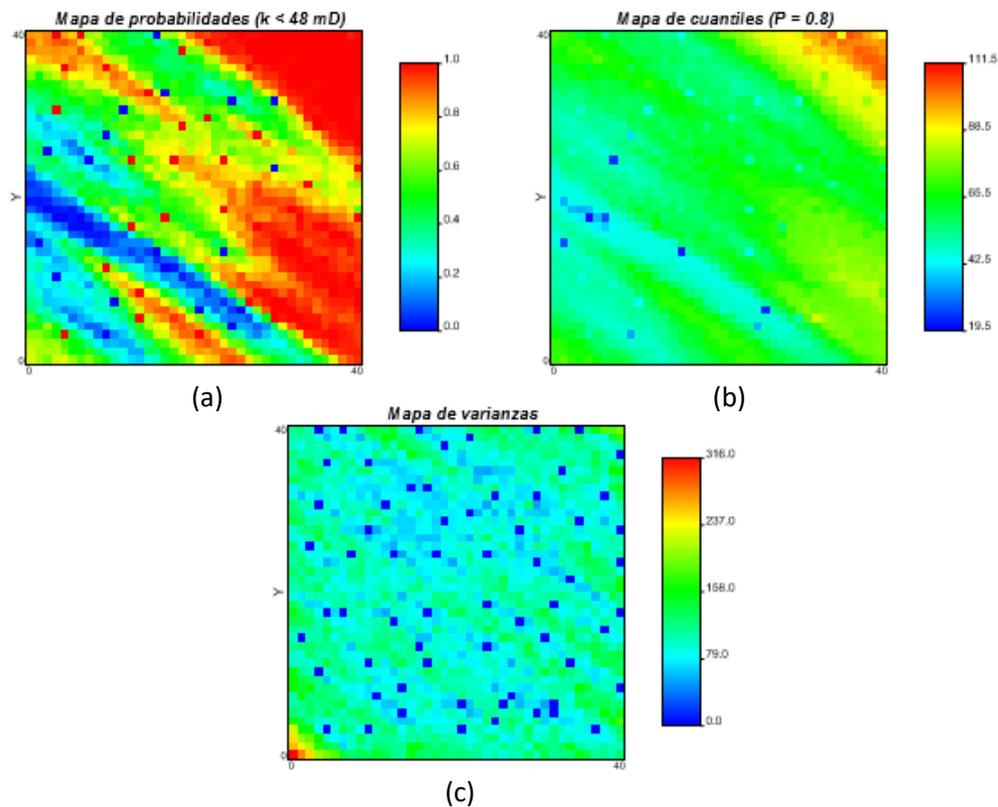


Fig.85. Mapa de probabilidades (a), mapa de cuantiles (b) y mapa de dispersión (varianza) (c).
(Fuente: Cassiraga, 2015)

2.7.2.4.3. Eligiendo un algoritmo de simulación

No existe un algoritmo que sea el “mejor” para resolver todos los problemas prácticos o, en otras palabras, el “mejor” algoritmo de representación estocástica depende del problema al que nos enfrentamos.

En muchos casos una combinación de algoritmos resulta ser el mejor procedimiento a seguir.

Algunos factores que condicionan la elección de un algoritmo de simulación son:

1. Los objetivos del estudio.
2. El tiempo humano y de CPU requerido para la generación de realizaciones.
3. La cantidad de información relevante a tener en cuenta.
4. La precisión y la exactitud de las predicciones probabilísticas realizadas.

El algoritmo de simulación por campos de probabilidad y los algoritmos gaussianos demandan menos esfuerzo por parte del modelizador y de los ordenadores con respecto a los indicadores. En contrapartida, los algoritmos indicadores son más flexibles a la hora de incorporar información.

La precisión y la exactitud se refiere al comportamiento de los algoritmos de simulación en relación a la distribución de las variables de respuestas de los modelos matemáticos en las que las simulaciones son utilizadas como entradas.

Una distribución de respuestas se considera exacta cuando la respuesta verdadera está comprendida en algún intervalo de probabilidad de interés.

Una distribución de respuestas se considera precisa en función de la dispersión de sus valores

En cuanto a la precisión, la precisión de un algoritmo de simulación es una medida frecuentemente llamada espacio de incertidumbre.

Una mayor cantidad de información condicionante no necesariamente reduce el espacio de incertidumbre: si agregamos datos que entran en conflicto con los que había, la incertidumbre en las respuestas puede crecer.

2.8. Ejercicio ejemplo de aplicación

Como ejemplo aplicación de la teoría anteriormente expuesta, se trabajará con un conjunto de datos de datos Berea (Giordano et al., 1985) que consiste en 1600 medidas de permeabilidad distribuidas regularmente en una malla de 40 por 40 celdas sobre una sección de arenisca de Berea de 61 cm por 61 cm. Este conjunto de datos ha sido aumentado con un conjunto de 64 medidas de porosidad generadas sintéticamente, asegurando la existencia de cierta correlación entre ambas variables.

Además, de los 1600 datos de permeabilidad se han seleccionado 64 datos por muestreo aleatorio estratificado. Tras dividir el área de trabajo en 64 paneles de igual tamaño, se ha elegido una muestra al azar en cada uno de los paneles. En cada uno de los 64 puntos se ha muestreado permeabilidad y porosidad.

Para su resolución se utilizará el software S-GeMS (Stanford Geostatistical Modeling Software), versión 2.1.

Los pasos a seguir en dicha resolución consistirán en los ya indicados anteriormente:

1. **Análisis de los datos** para lo cual se realiza tanto un análisis univariado como bivariado de los datos para así poder tanto representar gráficamente los valores y las localizaciones de los datos y verificar si existe algún valor fuera del rango como explicar la relación entre dos variables en juego.
2. **Análisis estructural** para encontrar un modelo de continuidad espacial y así poder utilizarlo en los siguientes apartados. Para ello se obtendrá el variograma al cual se le ajustará un modelo.
3. **Estimación.** Una vez realizado el apartado anterior, se puede realizar la estimación de la variable de interés en aquellas localizaciones donde no la conocemos. Para ello se utilizarán diferentes técnicas de estimación que contiene el software y distintos parámetros de continuidad espacial, para así poder comparar los diferentes resultados.
4. **Simulación.** Utilizando el algoritmo de simulación gaussiana secuencial, que dispone S_GeMS, se generarán 100 campos de permeabilidad condicionados a los 64 muestreados. A parte de ello, con la opción "Postsim" se procesarán los resultados obtenidos.

2.8.1. Análisis de los datos

2.8.1.1. Análisis Univariado

Tanto en la figura 86 como en la figura 87 se observan los valores de la porosidad y permeabilidad respectivamente.

En ambos se comprueba que los picos de valores se encuentran en el NE de la zona de estudio. Sin embargo, mientras que la porosidad también dispone en el SE otra zona con valores altos, la permeabilidad muestra un único dato con un valor muy superior con respecto al resto, por lo que se puede suponer que es un valor erróneo ya sea por un error de medición, de toma de datos...

Los rangos de valores mas bajos, en contraste, se encuentran en la zona opuesta en el suroeste para ambas variables Teniendo un cambio gradual de los valores más bajos para la porosidad de 4.68 hasta los valores mas altos en el noreste de 21 y de 15.32 hasta 101 para la permeabilidad.

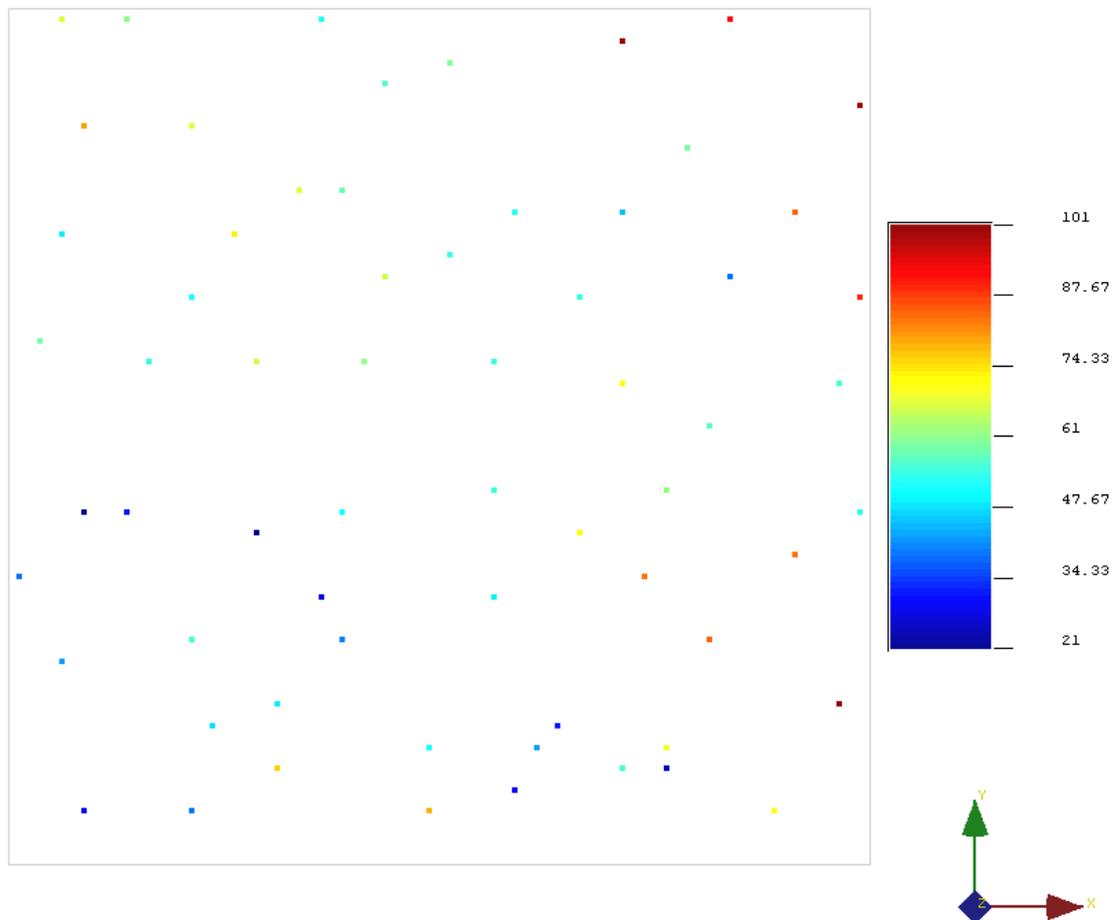
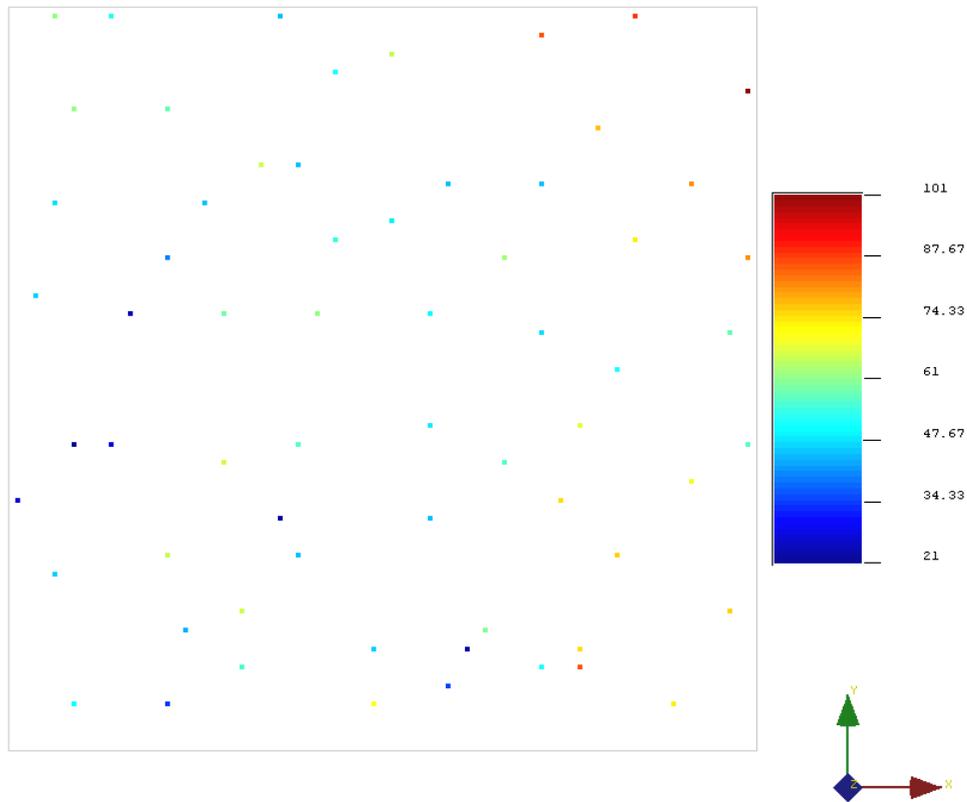


Fig.86. Ejemplo aplicación. Datos porosidad disponibles.
(Fuente: Elaboración propia)



*Fig.87. Ejemplo aplicación. Datos permeabilidad disponibles.
(Fuente: Elaboración propia)*

Es por ello, que a simple vista se suponga que la dirección de máxima y mínima variación sea entorno a N45-35E y N45-55W respectivamente.

A continuación, se estudian los histogramas con sus correspondientes funciones de probabilidad acumulada y sus estadísticos para los 64 datos que se muestran en las figuras 88 y 89.

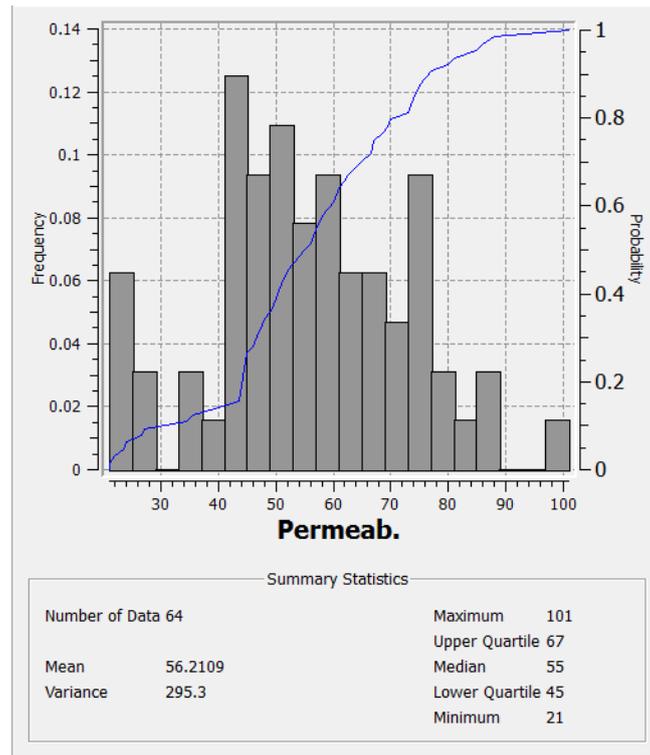


Fig.88. Ejemplo aplicación. Histograma permeabilidad.
(Fuente: Elaboración propia)

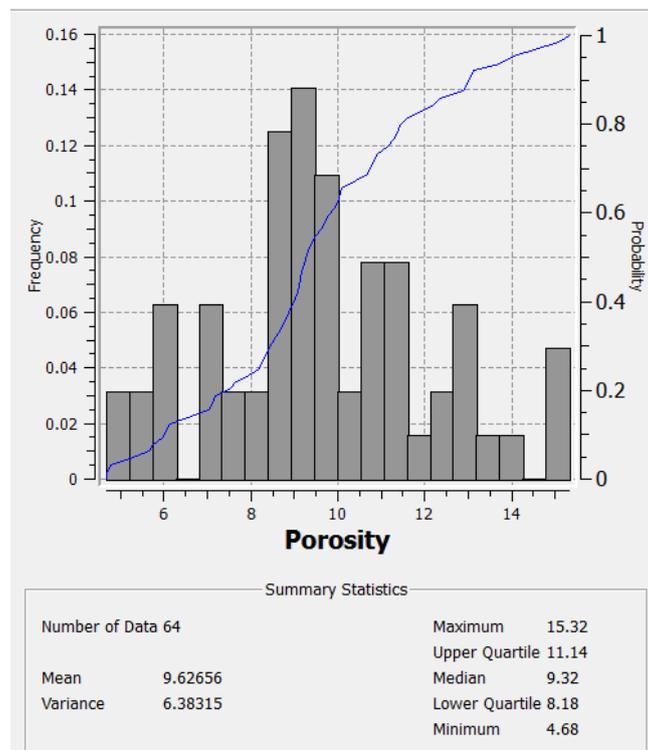


Fig.89. Ejemplo aplicación. Histograma porosidad.
(Fuente: Elaboración propia)

En el caso de la porosidad, las medidas de posición (media y mediana) indican el valor central de la distribución, y en este caso aproximadamente coinciden los dos estadísticos en el valor 9. Esto significa que la distribución es bastante simétrica y que la valoración media de la porosidad no sería del todo incorrecta.

En el caso de la permeabilidad se obtiene una media de 56,21 con un valor muy similar a la mediana (55). Siendo en este caso lo que destaca el coeficiente de variación de la media de 295.3, lo que muestra la heterogeneidad de los valores de la variable mientras que para la porosidad este representa un valor bajo (6.38), siendo este por ello más homogéneo.

2.8.1.2. Análisis Bivariado

Para el análisis bivariado se hará uso tanto del diagrama de dispersión para las 64 parejas de valores de la permeabilidad y porosidad como del diagrama de cuantiles que se muestran en las figuras 90 y 91.

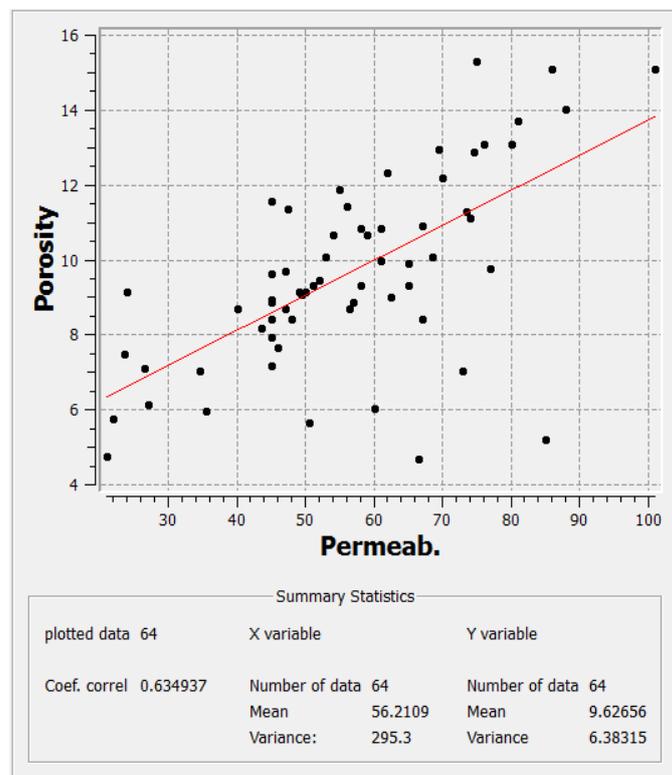


Fig.90. Ejemplo aplicación. Diagrama de dispersión. (Fuente: Elaboración propia)

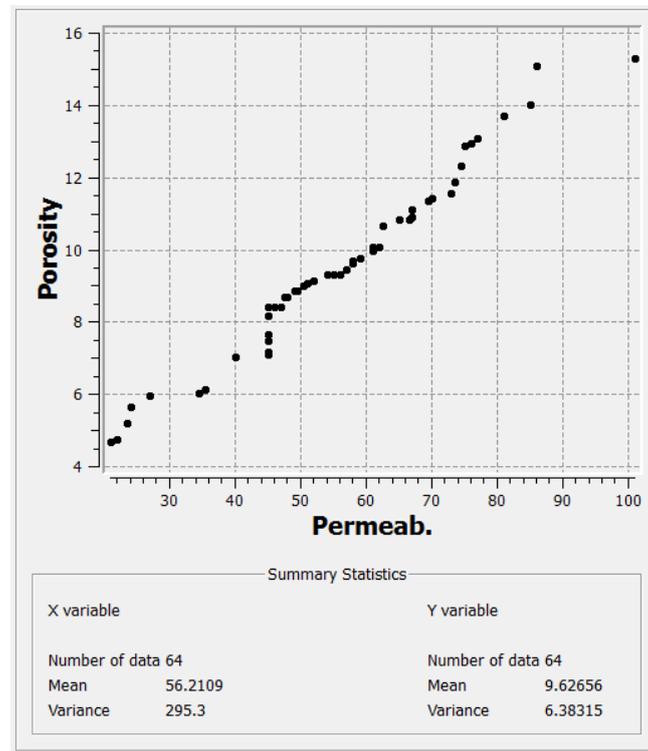


Fig.91. Ejemplo aplicación. Diagrama de cuantiles. (Fuente: Elaboración propia)

En el diagrama de dispersión, cuyo objetivo es poder comparar la relación que existe entre dos variables, nos muestra si hay alguna muestra que esté fuera del rango, el cual este extrañamente alejado de la nube de puntos. En este caso, en la figura 91, se puede observar que la nube de puntos sigue una correlación positiva, es decir, para valores altos de una variable, le corresponden valores altos de la segunda variable, sin embargo, en la zona inferior, se pueden observar varios puntos que no siguen esa “norma” y por tanto se les puede considerar como erratas, ya sea por errores en la medición, en la transcripción, ...

En cuanto al diagrama de cuantiles de las 64 parejas de valores, en el cual los cuantiles de dos distribuciones son representados versus el otro, se puede observar una clara tendencia lineal entre ambas, aunque al tener escalas diferentes, el hecho de que se concentren en torno a la recta de 45º no tiene un significado especial.

2.8.2. Análisis estructural

Como se ha dicho anteriormente en el libro *An introduction to Applied Geostatistics* (1989); en este apartado se busca obtener un modelo de continuidad para la variable permeabilidad. Para ello, se experimentará con el cálculo de variogramas y el ajuste de modelos a dichos variogramas.

Para ello primeramente se obtiene el variograma omnidireccional de la permeabilidad con el cual se puede obtener el alcance y el vector separación, el cual nos permitirá construir diferentes variogramas direccionales con los cuales poder evaluar la existencia o no de anisotropía y las direcciones de máxima y mínima continuidad espacial.

Además, una vez terminado el análisis de los variogramas experimentales, se ajustará el modelo correspondiente, pudiendo así escribir la ecuación del modelo ajustado.

2.8.2.1. Variograma omnidireccional.

Para la obtención de este, se necesita caracterizar el vector separación, para ello es necesario establecer el número de intervalos a calcular y el módulo de dicho vector.

Para ello se tomará como criterio:

- En la estimación del variograma omnidireccional los pares de las observaciones se agrupan según la distancia dentro de un intervalo $h = h$ con una tolerancia $\pm \Delta h$ y dentro de una dirección θ con una tolerancia $\pm \Delta \theta$. El variograma omnidireccional así estimado es considerado suavizado o regularizado.
- Además, debe ser considerado solamente para pequeñas distancias por lo que generalmente, se estima para valores de h menores que la mitad de la distancia máxima ($|h| < \frac{d_{max}}{2}$).
- También se realizará de tal forma que la meseta del variograma no exceda el valor de la varianza de los 64 datos de permeabilidad (295.3), para así poder obtener un valor inicial del valor de la meseta en los variogramas direccionales.
- La elección del número de intervalos es arbitraria. No obstante, se considera que un número máximo de 25 intervalos es suficiente para cualquier propósito, y un mínimo de 10 debe ser usado para determinar con precisión el rango y la meseta del dicho variograma
- El largo de los intervalos debe ser elegido de forma tal que el número de pares en cada intervalo sea lo suficientemente grande para que el valor estimado del variograma omnidireccional sea relativamente estable. Se considera que entre 30 y 50 pares satisfacen este requerimiento.

Con estos criterios iniciales, y tras diversas pruebas realizadas, se ha obtenido como valores óptimos:

Número de intervalos	12
Vector separación	2
Tolerancia	1

Tabla 8. Ejemplo aplicación. Numero de intervalos, vector separación y tolerancia óptimos variograma omnidireccional. (Fuente: Elaboración propia)

En la figura 92 se observa que aun no cumpliendo en uno de los puntos el número de pares de valores suficiente para obtener un estimador fiable y estable (11 pares de valores en vez del mínimo 30), cumple el resto de recomendaciones establecidas anteriormente.

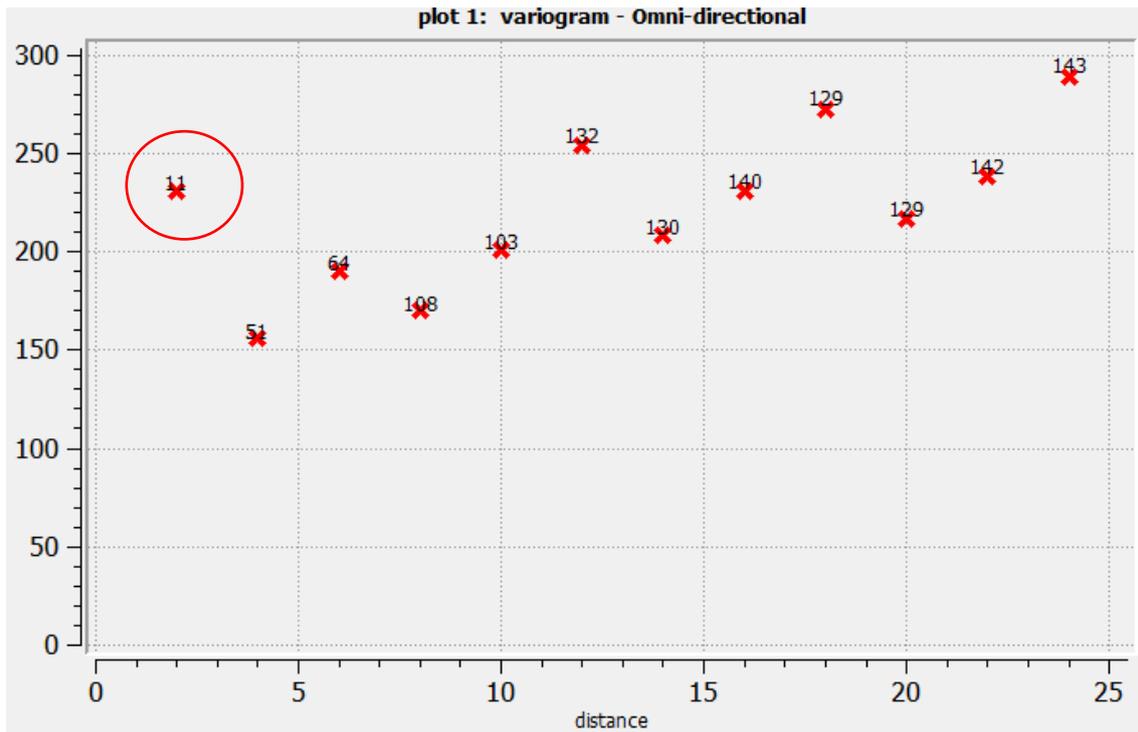


Fig.92. Ejemplo aplicación. Variograma omnidireccional. (Fuente: Elaboración propia)

A este se le ajusta un modelo para así poder obtener parámetros necesarios para la obtención de las direcciones de máxima y mínima continuidad del apartado siguiente, como puede ser el efecto pepita y la meseta de 249.1. Obteniendo el siguiente variograma ajustado de la figura 93.

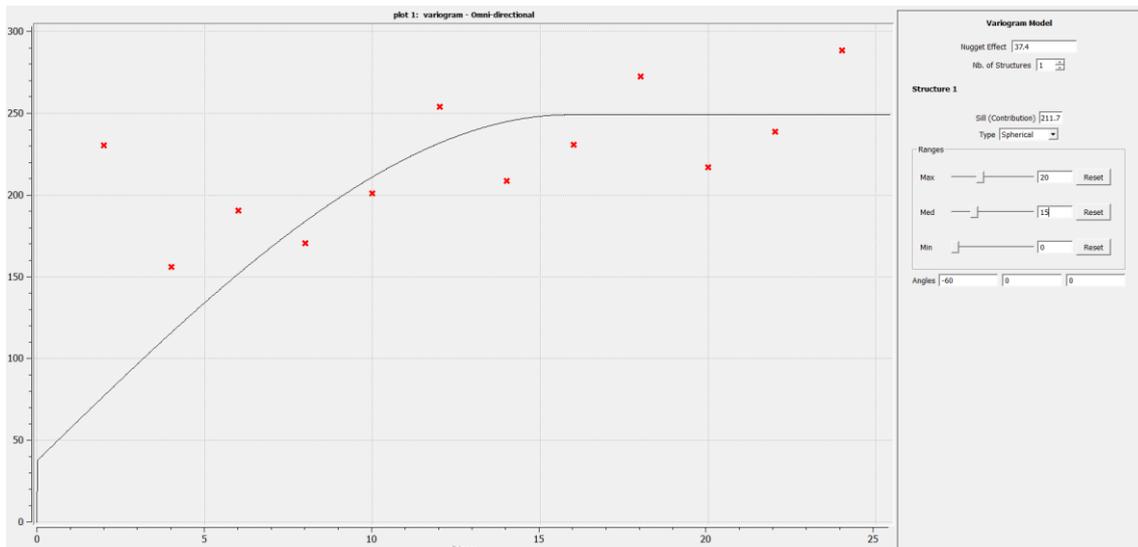


Fig.93. Ejemplo aplicación. Variograma omnidireccional ajustado. (Fuente: Elaboración propia)

2.8.2.2. Variograma direccional.

Una vez obtenido el vector separación y el número de intervalos óptimos, se realiza el proceso anterior, pero para diferentes direcciones. Con ello se podrá observar si el fenómeno a estudiar es isótropo, es decir, si se comportará siempre igual en cualquier dirección, tendrán el mismo alcance, punto a partir del cual se pierde la correlación espacial, no hay continuidad espacial.

Para ello se calculan los variogramas direccionales en varias direcciones típicas y se compararán los alcances de unos con otros. Si todos tienen un valor muy similar el fenómeno se considerará isótropo, sin embargo, si varían de una dirección a otra, el fenómeno será anisótropo, y se deberá determinar las direcciones de máxima y mínima continuidad a través de la elipse de anisotropía.

Para ello se toma una tolerancia angular de 30º y un ancho de banda de 50, además de los valores obtenidos en el apartado anterior. A continuación, en la Tabla 9, se muestra las distintas direcciones tomadas y los alcances correspondientes a cada dirección.

Direcciones	Alcance
N0W	15
N30W	17.5
N60W	19
N90W	18
N120W	14
N150W	14
N180W	15

Tabla 9. Ejemplo aplicación. Direcciones tomadas y alcances obtenidos. (Fuente: Elaboración propia)

En la figura 94 se muestran los variogramas obtenidos para cada una de las direcciones.

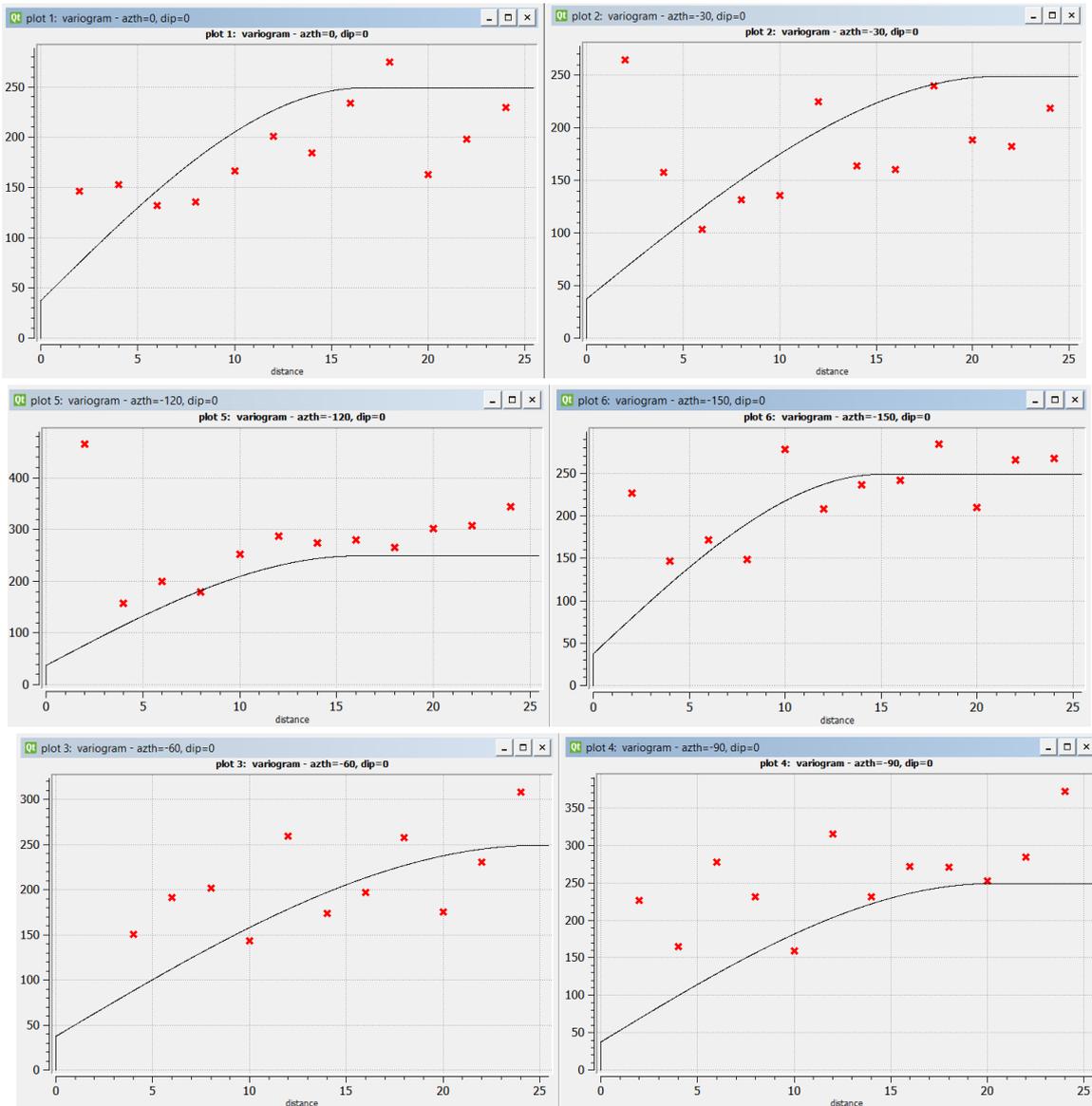


Fig.94. Ejemplo aplicación. Variogramas direccionales. (Fuente: Elaboración propia)

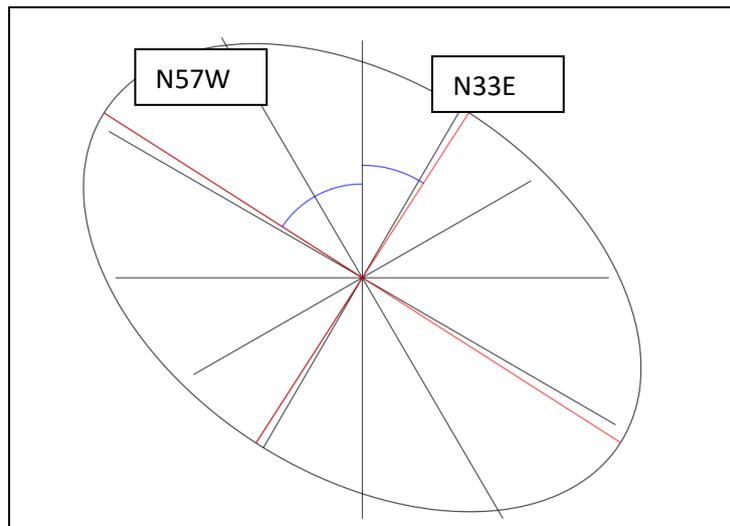


Fig.95. Ejemplo aplicación. Elipsoide de anisotropía. (Fuente: Elaboración propia)

Se muestran los variogramas correspondientes a las direcciones de máxima (figura 96) y mínima continuidad (figura 97) ajustados al modelo.

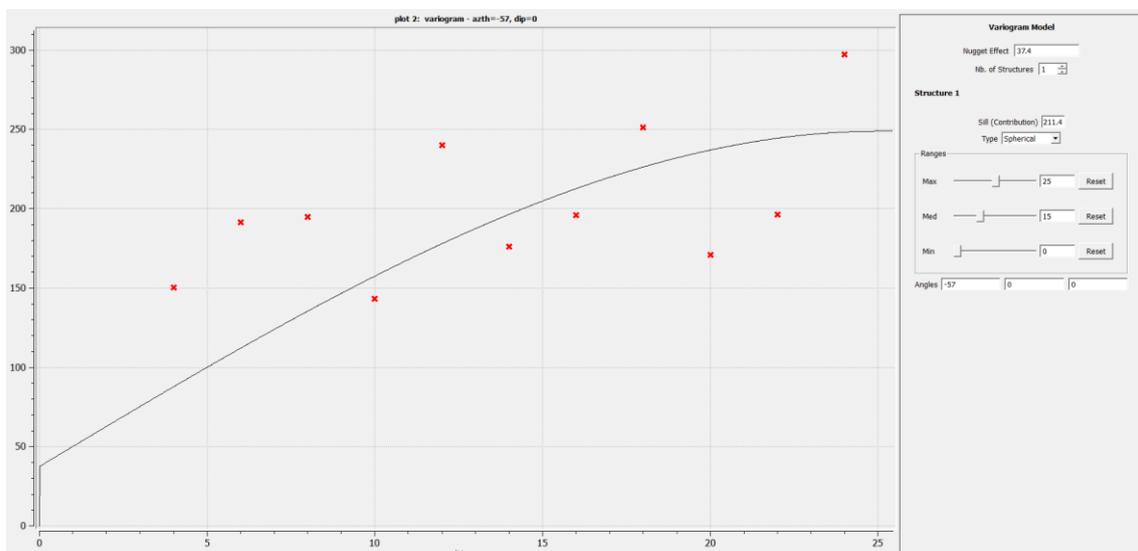


Fig.96. Ejemplo aplicación. Elipsoide de anisotropía. Variograma direccional máxima continuidad ajustada. (Fuente: Elaboración propia)

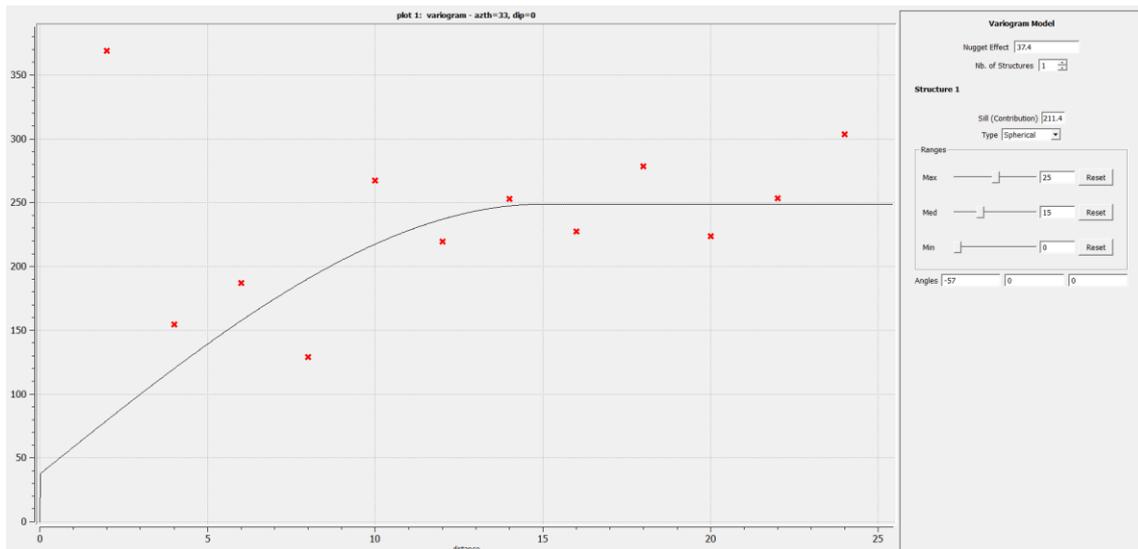


Fig.97. Ejemplo aplicación. Variograma direccional mínima continuidad ajustada. (Fuente: Elaboración propia)

Se le ajusta al variograma un modelo esférico de ecuación:

$$\gamma(h) = 37.4 + 211.7 Sph(h)[a_{max} = 25; a_{min} = 15; \alpha = -57^\circ]$$

Donde:

- Efecto pepita \rightarrow 37.4
- Meseta \rightarrow 249.1 (Contribución 211.7)
- Alcance de máxima continuidad \rightarrow 25
- Alcance de mínima continuidad \rightarrow 15
- Dirección de máxima continuidad \rightarrow -57°

2.8.3. Krigeado

Una vez analizados los datos y construido el modelo de continuidad espacial, SGeMS permite realizar estimaciones de una variable espacialmente distribuida en aquellas localizaciones donde no la conocemos, siendo el objetivo, la obtención de un mapa de permeabilidad utilizando como datos condicionantes los 64 datos de esa variable y el modelo de continuidad espacial que para ella ha sido construido.

A continuación, se realizan diferentes estimaciones mediante Krigeado ordinario en las cuales se irán variando los parámetros para así ver la influencia que cada uno de ellos tiene en dichas estimaciones.

2.8.3.1. Estimación campo permeabilidad mediante krigeado ordinario.

Para la estimación del campo de permeabilidad, se ha empleado un krigeado ordinario con un vecindario de búsqueda circular de 50 (Lo suficientemente grande para que al menos contenga la elipse de anisotropía)

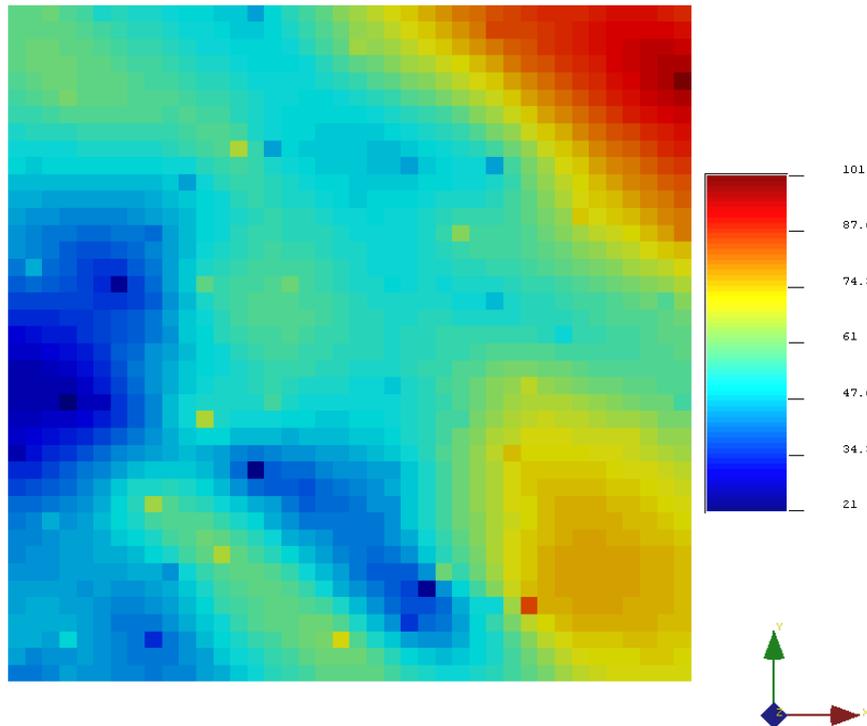


Fig.98. Ejemplo aplicación. Campo permeabilidad mediante Krigeado ordinario.
(Fuente: Elaboración propia)

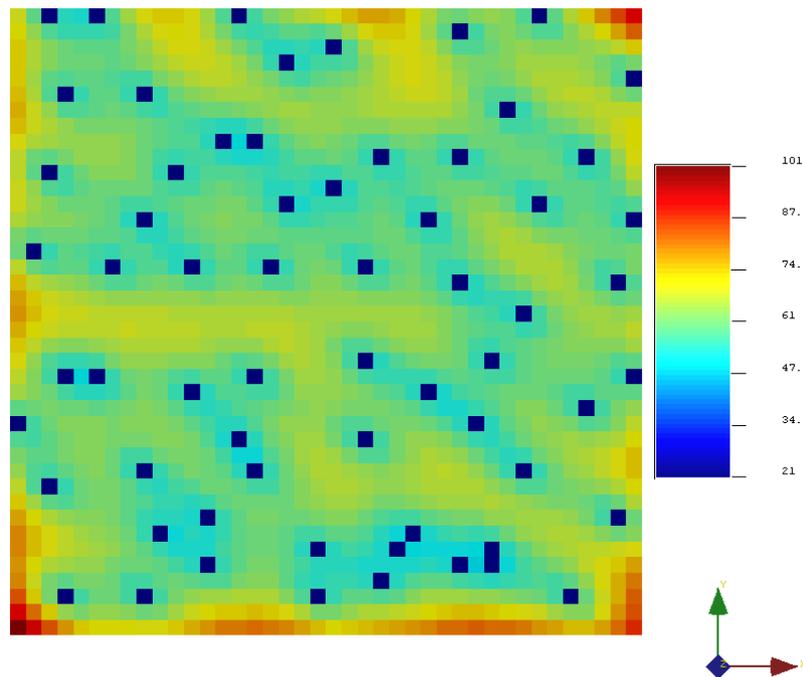


Fig.99. Ejemplo aplicación. Varianza campo de permeabilidad estimado mediante Krigeado ordinario.
(Fuente: Elaboración propia)

Como cabía esperar, se puede observar en la figura 98 que en la dirección de máxima continuidad N57W la variación de los valores de permeabilidad es mínima, mientras que en su perpendicular N33E es máxima, coincidente con la dirección de mínima continuidad.

En cuanto a la varianza de los datos observamos en la figura 99 que los mayores valores se encuentran en los extremos superior derecho e inferior izquierdo, coincidentes con la dirección de mínima continuidad, lo que tiene sentido a la vista de la definición de varianza la cual es definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

2.8.3.2. Estimación campo permeabilidad mediante Krigeado ordinario. Efecto pepita

En este apartado se quiere observar la influencia del efecto pepita en el modelo de continuidad espacial a la hora de hacer las estimaciones. En este caso, se ha aumentado considerablemente el efecto pepita hasta 200, de forma que si no se varía la meseta, la contribución será de 49.1, quedando por lo tanto como ecuación del modelo esférico:

$$\gamma(h) = 200 + 49.1 Sph(h)[a_{max} = 25; a_{min} = 15; \alpha = -57^\circ]$$

Tras su estimación se ha podido comprobar que los valores estimados tienen una variación más suave que la inicial ya que casi toda la continuidad se pierde a pequeñas distancias. Es por ello que la correlación entre los datos es menor.

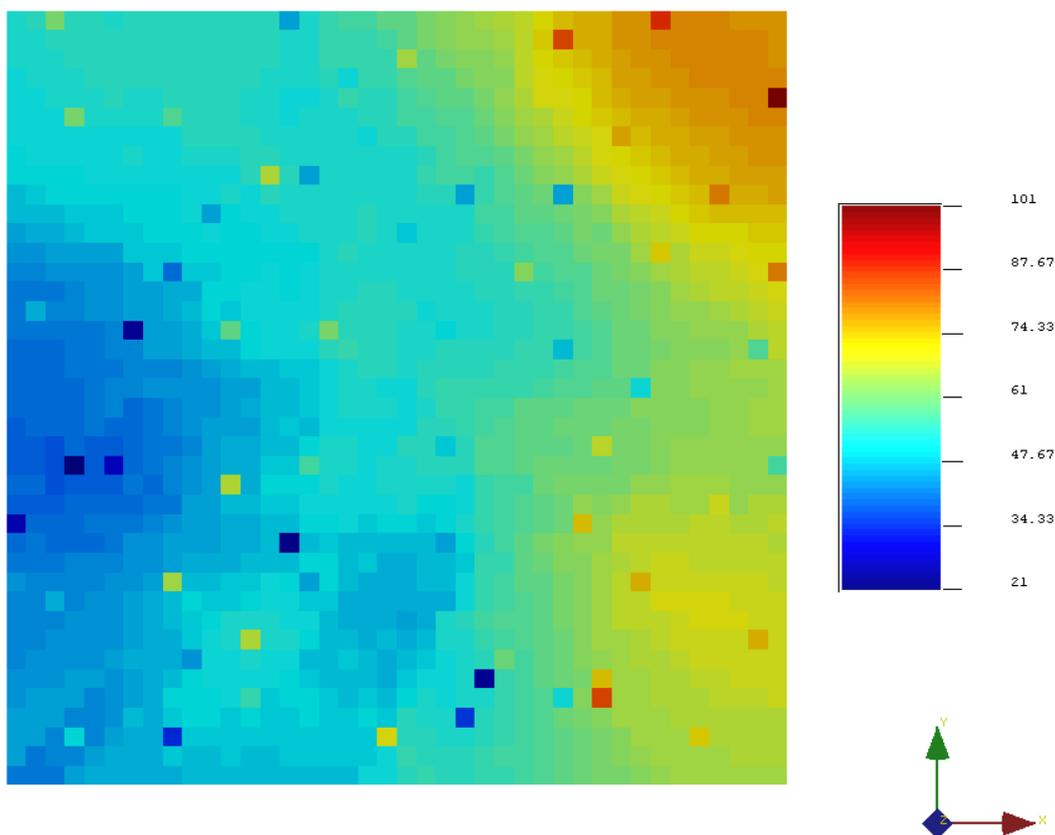


Fig.99. Ejemplo aplicación. Efecto pepita en el campo de permeabilidad estimado. (Fuente: Elaboración propia)

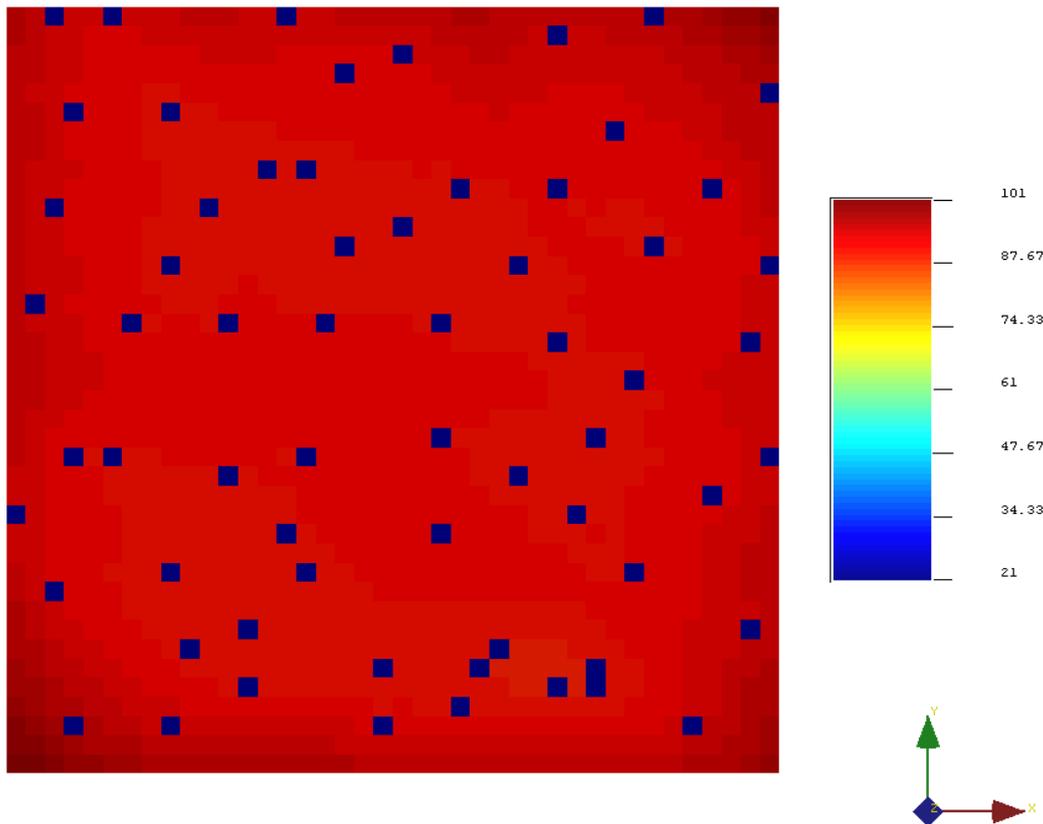


Fig.100. Ejemplo aplicación. Efecto pepita en la varianza del campo de permeabilidad estimado.
(Fuente: Elaboración propia)

Si la continuidad se pierde muy rápidamente, se puede comprobar que la varianza da valores muy altos.

2.8.3.3. Estimación campo permeabilidad mediante Krigeado ordinario. Efecto pepita puro

Si el caso anterior se lleva al extremo, siendo el efecto pepita puro, hay una pérdida de continuidad espacial, como se puede ver en la figura 101.

En este caso, los puntos muestreados solo dan información de dicho punto, pero en cuanto se desplaza algo, no hay continuidad, estructuralmente todos estarán igual de lejos, ninguno recibirá más peso que otro. Por ello el campo de estimación en principio debería presentar el mismo color en todo el campo. Pero como el programa SGeMS no permite realizar estimaciones con el efecto pepita puro, se ha dado un valor muy alto de él y por ello se ve que la variación es muy suave, con un rango de valores mínimo.

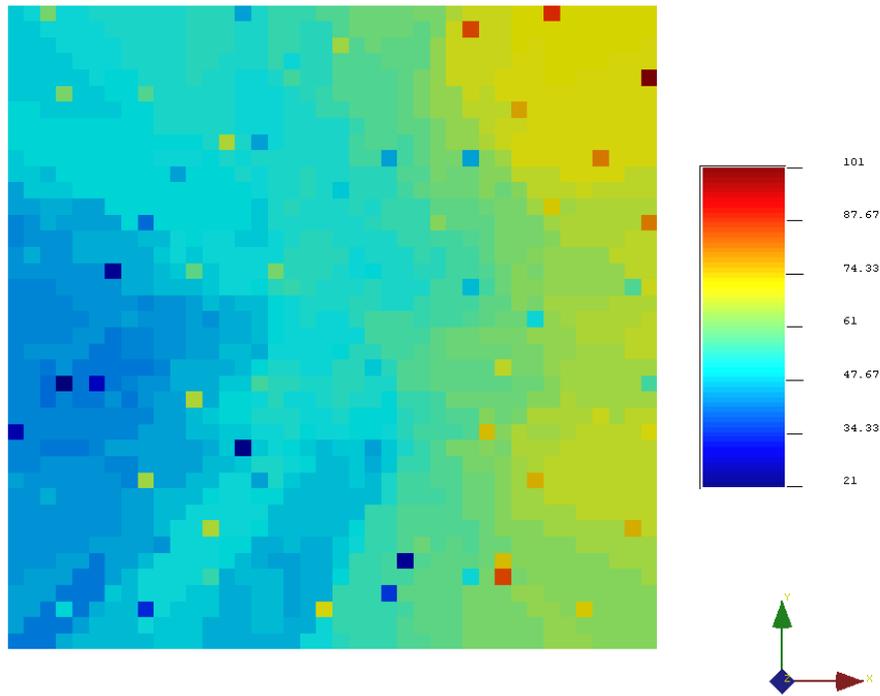


Fig.101. Ejemplo aplicación. Efecto pepita puro en el campo de permeabilidades estimado.
(Fuente: Elaboración propia)

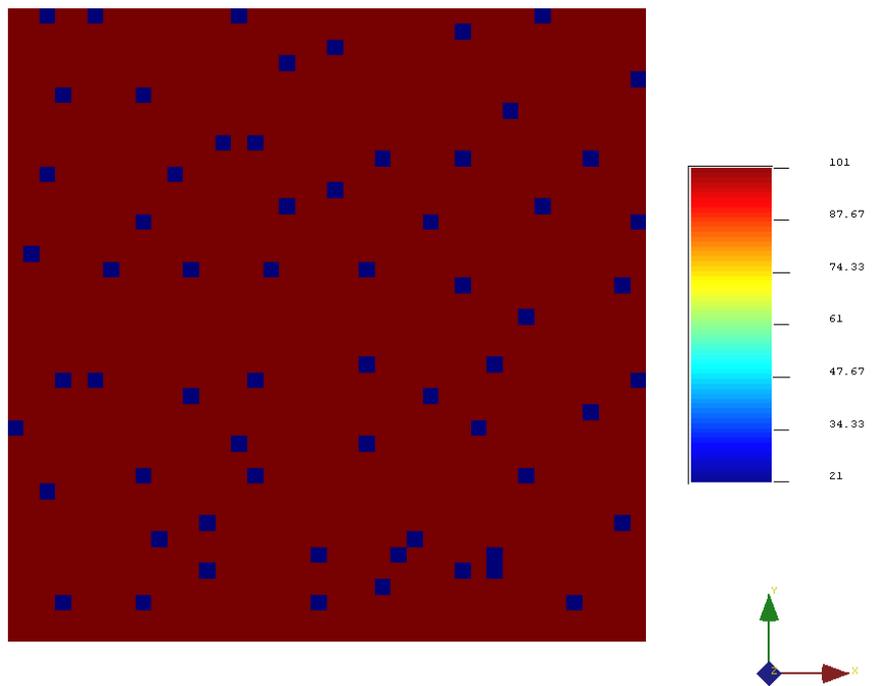


Fig.102. Ejemplo aplicación. Efecto pepita puro en la varianza del campo de permeabilidad estimado.
(Fuente: Elaboración propia)

En cuanto a la varianza se observa en la figura 102 que es máxima, como cabía esperar.

2.8.3.4. Estimación campo permeabilidad mediante Krigeado ordinario. Cambio escala variograma

Este apartado se centrará en estudiar el efecto que produce un cambio de escala en el variograma calculado inicialmente. Para ello, se dobla el valor del variograma inicial:

$$\gamma(h) = 74.8 + 423.4 Sph(h)[a_{max} = 25; a_{min} = 15; \alpha = -57^\circ]$$

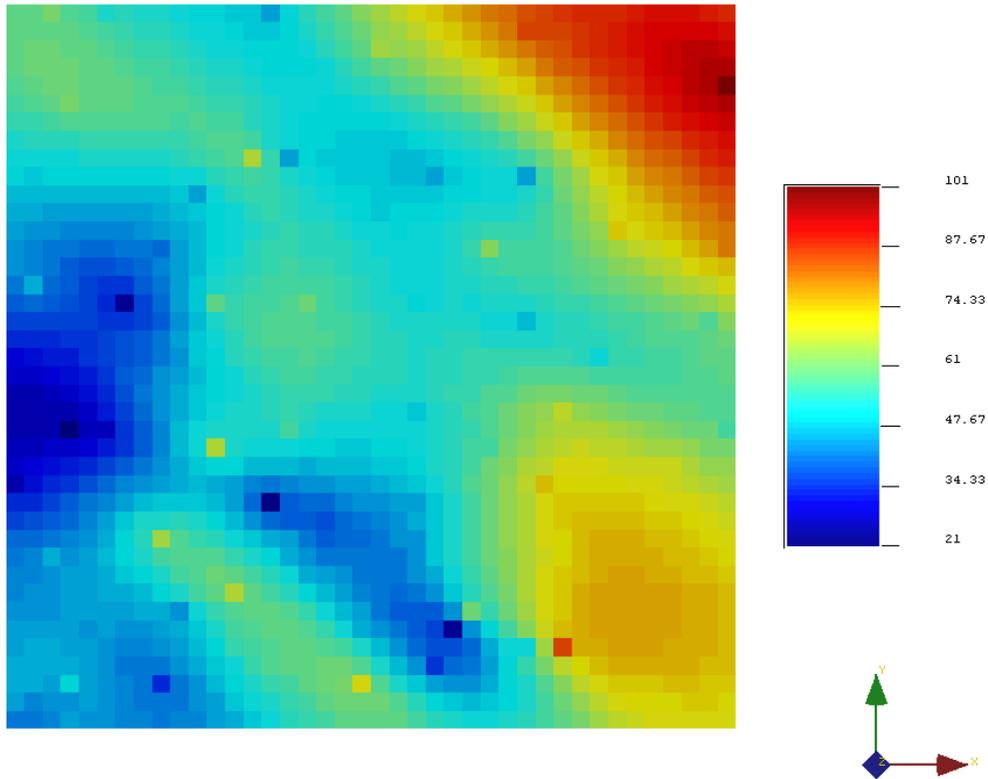


Fig.103. Ejemplo aplicación. Efecto de doblar el variograma en la estimación del campo de permeabilidades. (Fuente: Elaboración propia)

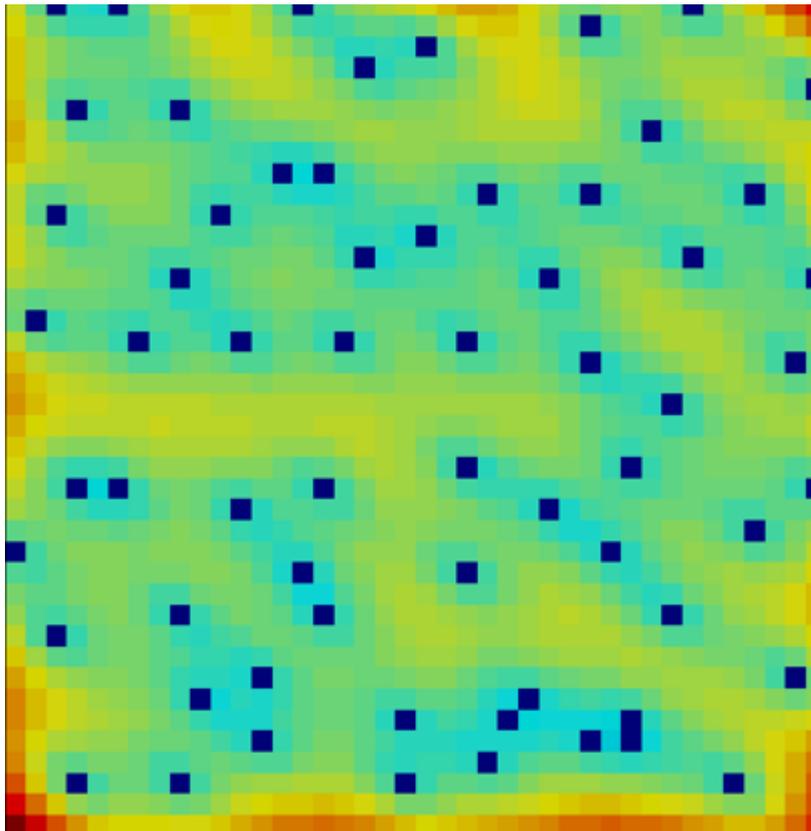


Fig.104. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al doblar el variograma.
(Fuente: Elaboración propia)

A la vista de los resultados (Figura 103), el doblar el variograma no afecta de manera perceptible al caso inicial (Figura 98), al igual que ocurre con la varianza

2.8.3.5. Estimación campo permeabilidad mediante Krigeado ordinario. Efecto cambio modelo continuidad espacial

En cuanto al modelo de continuidad seleccionado, se eligió un modelo esférico ya que éste proporcionaba un mejor ajuste. Sin embargo, a continuación, se van a estudiar los otros dos tipos (Exponencial y Gaussiano) que el programa SGeMS permite utilizar.

A continuación, en las siguientes figuras se muestran los diferentes variogramas (esférico, exponencial y gaussiana) con sus correspondientes estimaciones.

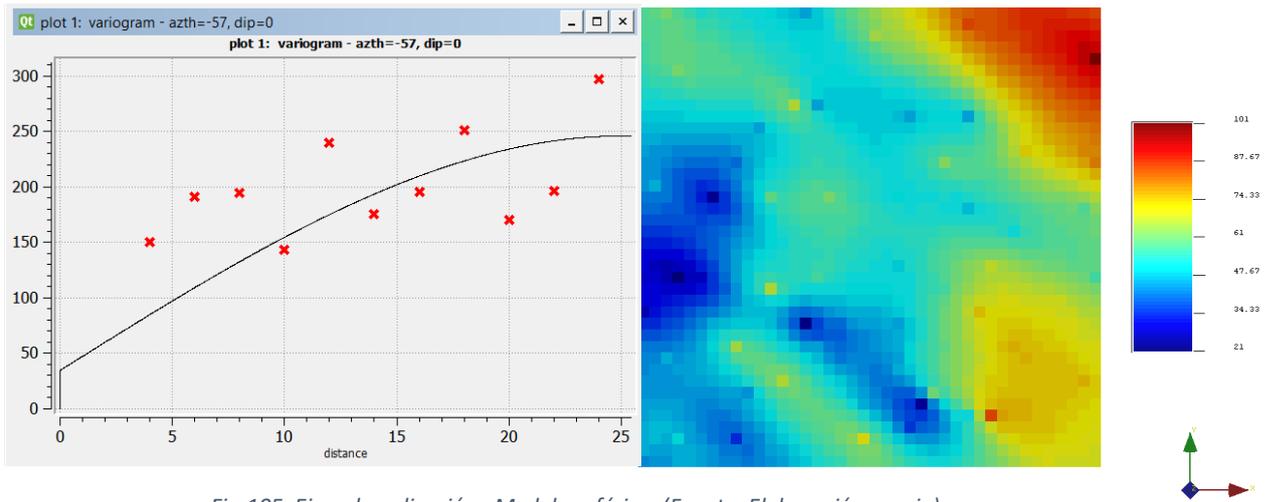


Fig.105. Ejemplo aplicación. Modelo esférico. (Fuente: Elaboración propia)

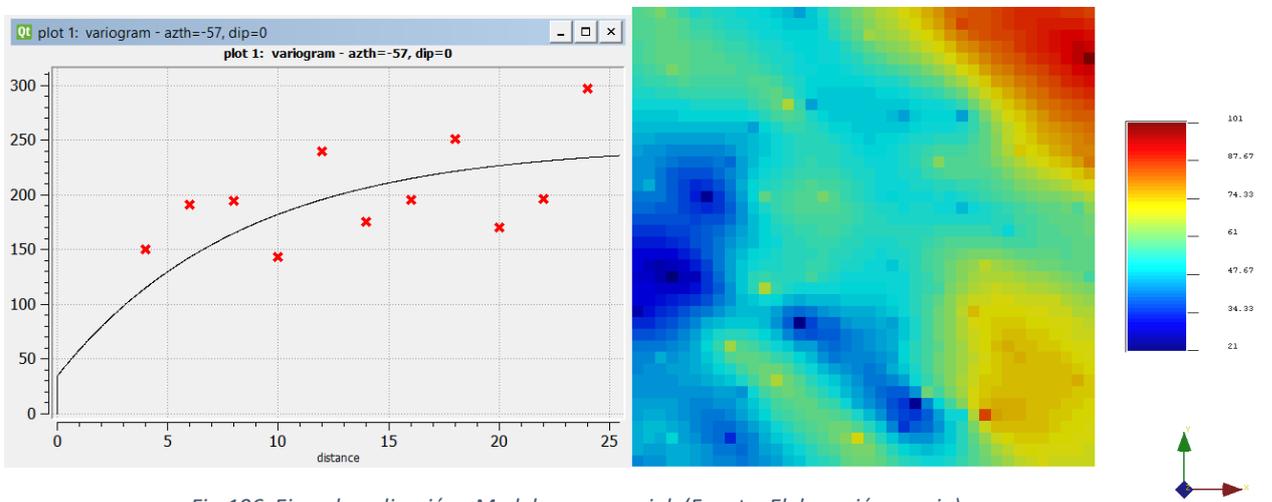


Fig.106. Ejemplo aplicación. Modelo exponencial. (Fuente: Elaboración propia)

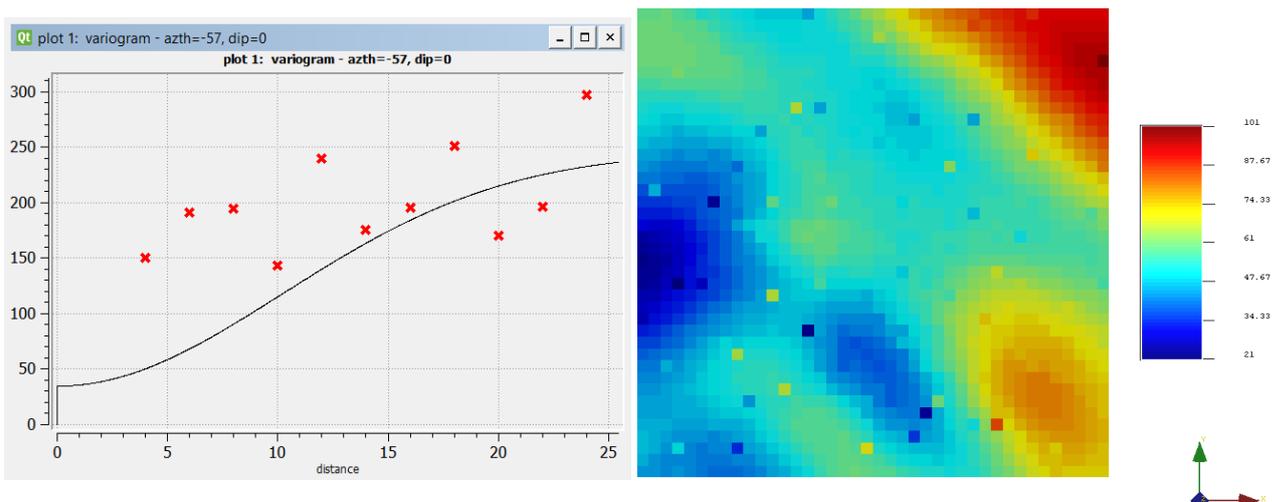


Fig.107. Ejemplo aplicación. Modelo gaussiano. (Fuente: Elaboración propia)

Ambos modelos (figura 105 y 106) dan soluciones aproximadas. Sin embargo, en el modelo gaussiano (figura 107), la continuidad se pierde mas tarde ya que, como se puede observar en el variograma, este modelo comienza con tangente nula.

2.8.3.6. Estimación campo permeabilidad mediante Krigeado ordinario. Efecto del alcance

En este apartado y en los dos siguientes se llevarán a cabo estimaciones con distintos alcances estudiando su distinto efecto.

Como bien se ha comentado, en este

En este caso se verá el efecto de doblar el alcance en la dirección de mayor continuidad, pasando de 25 a 50, manteniendo el resto de parámetros como el inicial:

$$\gamma(h) = 37.4 + 211.7 Sph(h)[a_{max} = 50; a_{min} = 15; \alpha = -57^\circ]$$

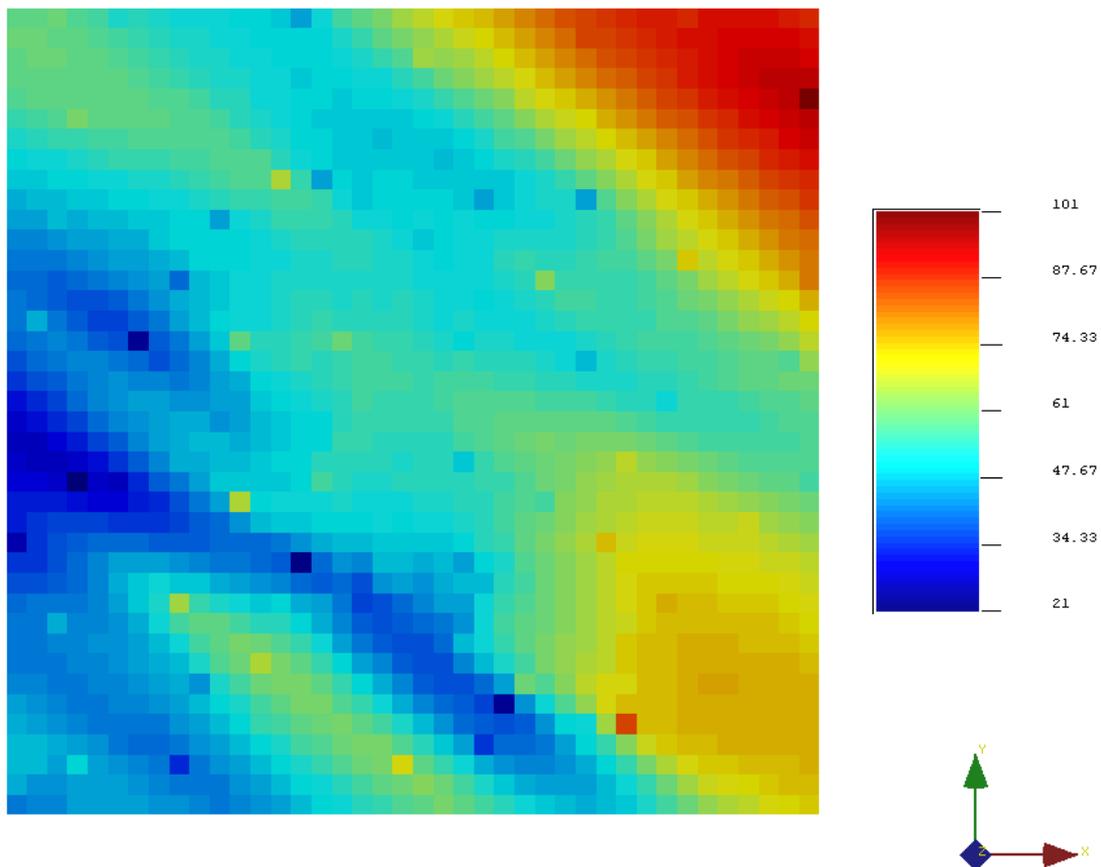


Fig.108. Ejemplo aplicación. Efecto de doblar el alcance en la estimación del campo de permeabilidades. (Fuente: Elaboración propia)

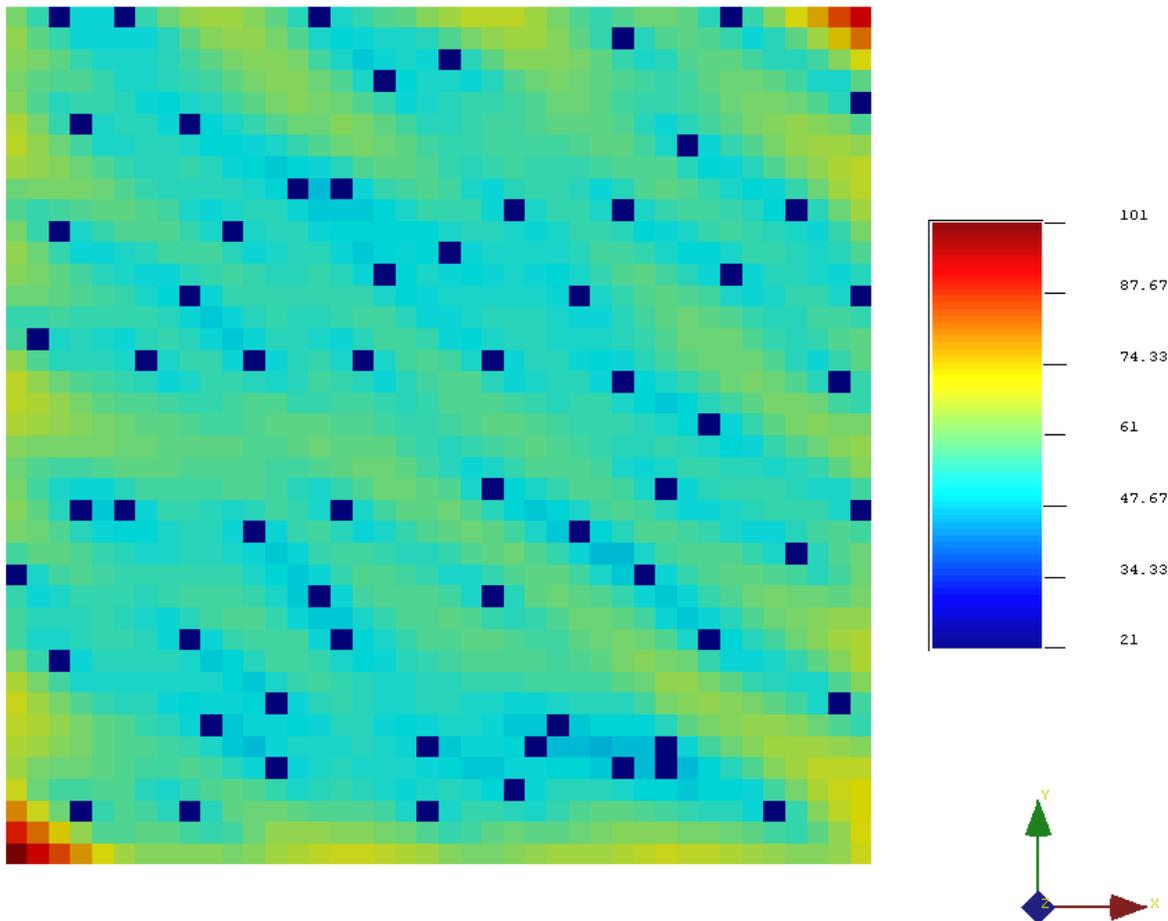


Fig.109. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al doblar el alcance.
(Fuente: Elaboración propia)

Al doblar el alcance, cada dato tiene una mayor influencia en la estimación de los datos entre los puntos que están próximos. Es por ello, que se observa una mayor continuidad en la dirección de máxima continuidad.

Lo mismo se puede comprobar con el campo de varianzas de la figura 109, que disminuye su valor con respecto al campo inicial de la figura 99.

2.8.3.7. Estimación campo permeabilidad mediante Krigeado ordinario. Alcances dirección de máxima continuidad igual a dirección mínima continuidad.

En este caso, se tomará como alcance para la dirección de máxima continuidad el mismo valor que para la dirección de mínima continuidad, por lo que el campo de continuidad espacial se asume un círculo de radio 15.

$$\gamma(h) = 37.4 + 211.7 Sph(h)[a_{max} = 15; a_{min} = 15; \alpha = -57^\circ]$$

Comparando con el campo inicial calculado en la figura 98, las variaciones son mínimas, ello se debe a que la anisotropía sigue estando presente.

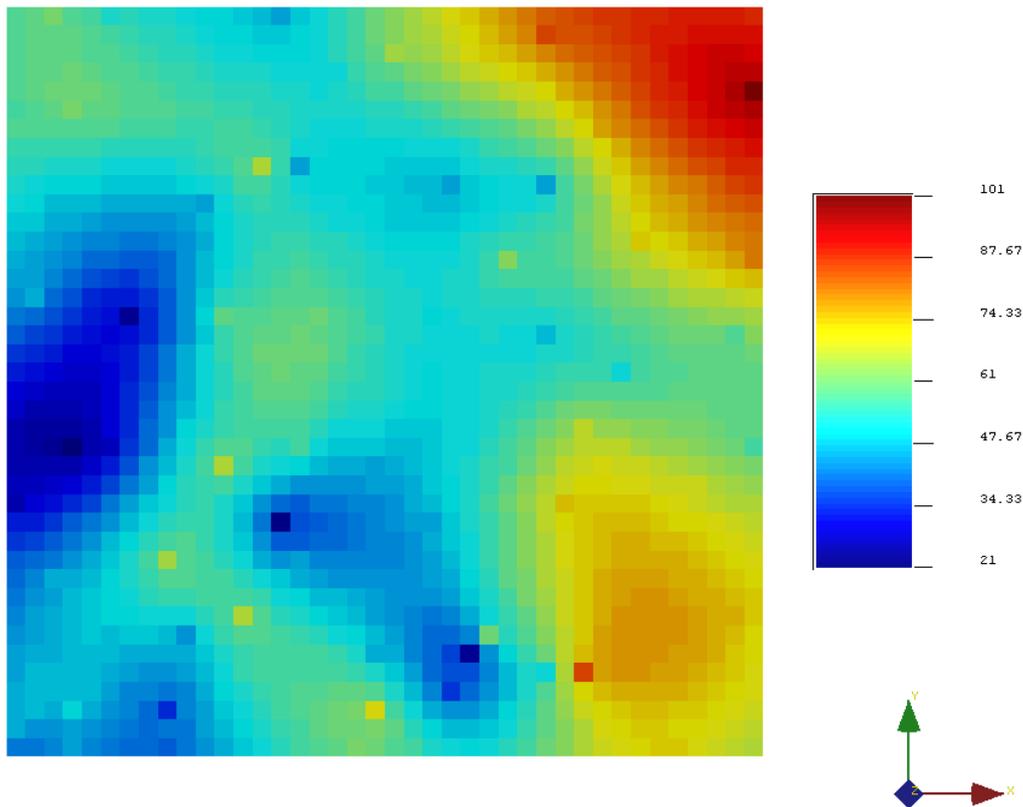


Fig.110. Ejemplo aplicación. Efecto de igualar el alcance de máxima continuidad con el de mínima continuidad en la estimación del campo de permeabilidades. (Fuente: Elaboración propia)

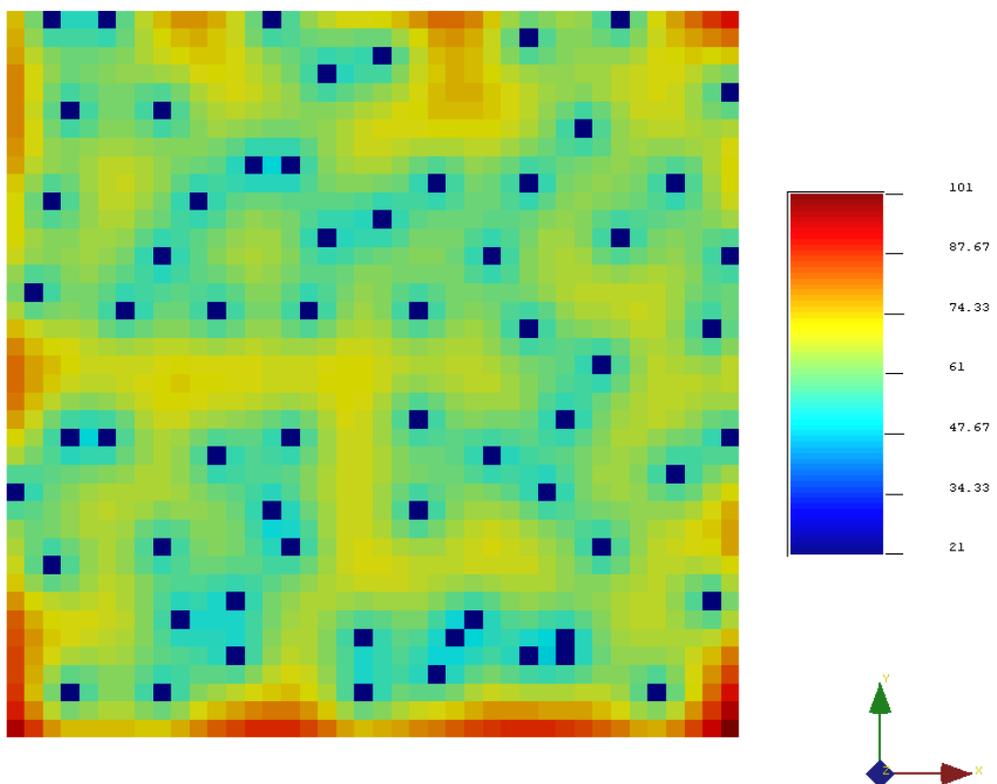


Fig.111. Ejemplo aplicación. Varianza del campo de permeabilidades estimado al igualar el alcance de máxima continuidad con el de mínima continuidad. (Fuente: Elaboración propia)

2.8.3.8. Estimación campo permeabilidad mediante Krigeado ordinario. Inversión alcances direcciones de máxima y mínima continuidad.

En este caso se ha seguido el siguiente modelo de continuidad espacial:

$$\gamma(h) = 37.4 + 211.7 Sph(h)[a_{max} = 15; a_{min} = 25; \alpha = -57^\circ]$$

En este caso, se observa en la figura 112 que al imponer los valores invertidos, la dirección de máxima y mínima continuidad también se han invertido, siendo en este caso la dirección de máxima continuidad la que correspondería realmente a la de mínima, siendo por ello muy sensible a errores.

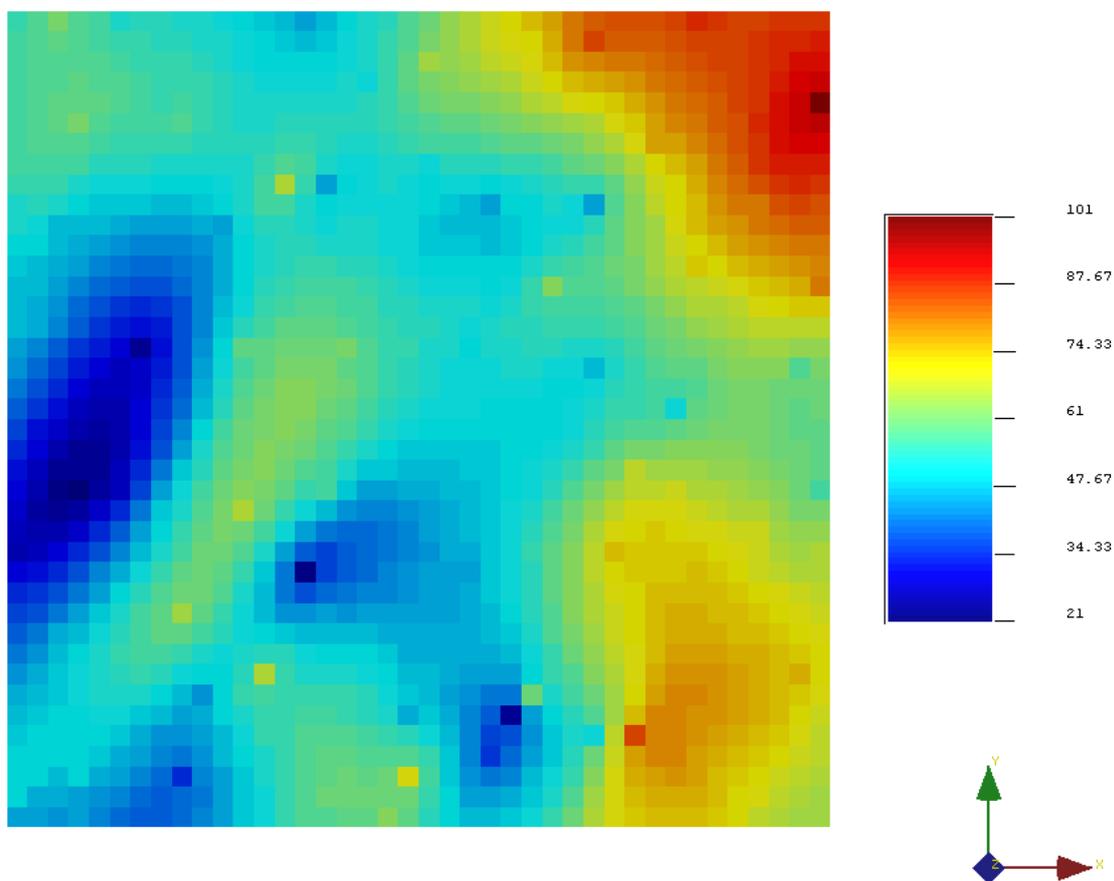


Fig.112. Ejemplo aplicación. Efecto de invertir los alcances de las direcciones de máxima y mínima continuidad en la estimación del campo de permeabilidades. (Fuente: Elaboración propia)

2.8.4. Cokrigado

En este apartado se pretende utilizar como técnica de estimación el cokrigado y así poder comparar el impacto que conlleva el incorporar en el proceso de estimación de una variable información secundaria que no es exhaustivamente conocida, en este caso la porosidad.

Para ello se han empleado los variogramas de la permeabilidad y porosidad y el variograma cruzado de ambas siguientes:

$$\gamma_k(h) = 37.4 + 211.7 Sph(h)[a_{max} = 70; a_{min} = 12; \alpha = -57^\circ(h)]$$

$$\gamma_p(h) = 2.3 + 17.8 Sph(h)[a_{max} = 70; a_{min} = 12; \alpha = -57^\circ(h)]$$

$$\gamma_{kp}(h) = 7.6 + 17.8 Sph(h)[a_{max} = 70; a_{min} = 12; \alpha = -57^\circ(h)]$$

Se comienza realizando el mapa de permeabilidad con 16 datos por Krigado ordinario (figura 113) para después considerar además los 64 datos de porosidad como variable secundaria mediante cokrigado (figura 114).

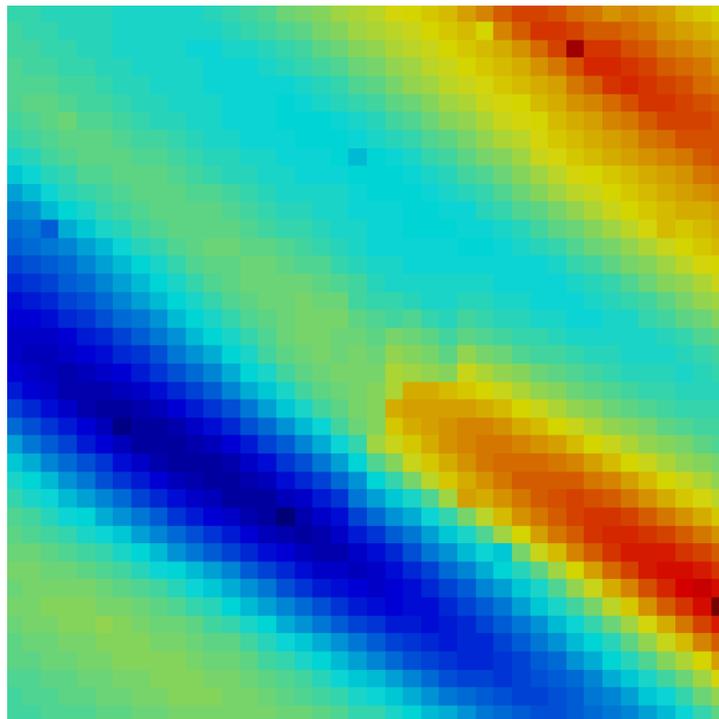


Fig.113. Ejemplo aplicación. Campo permeabilidades por krigado ordinario mediante 16 datos permeabilidad. (Fuente: Elaboración propia)

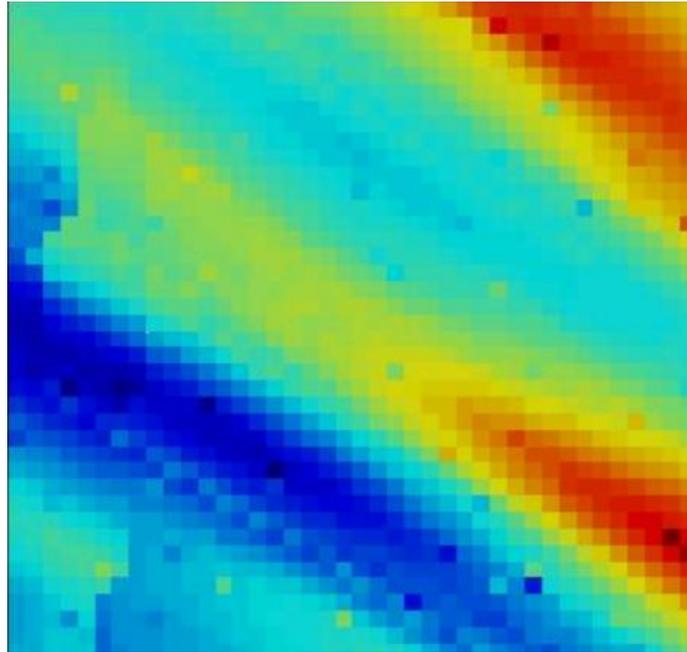


Fig.114. Ejemplo aplicación. Campo permeabilidades por cokrigado mediante 16 datos permeabilidad y 64 porosidad. (Fuente: Elaboración propia)

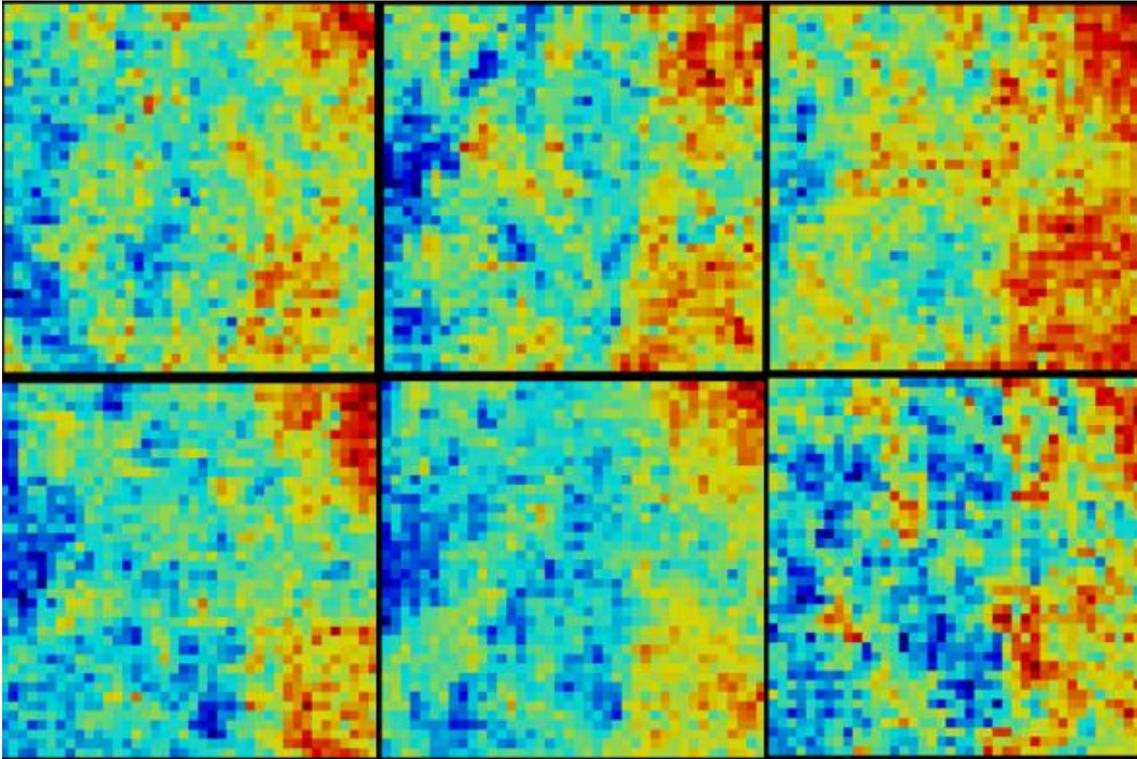
A la vista de los resultados se puede observar que utilizando la técnica del cokrigado los resultados obtenidos son más precisos, en los cuales destaca la anisotropía, alargando las zonas con mayor continuidad.

2.8.5. Simulación gaussiana secuencial

La simulación reproduce la variabilidad espacial de los valores reales mientras que en las estimaciones se produce un suavizado de las dispersiones reales, además, estas últimas no reproducen las propiedades estadísticas de la función aleatoria, mientras que las simulaciones sí.

Por ello, y utilizando el algoritmo de simulación gaussiana secuencial se van a generar 100 campos de permeabilidad condicionados a los 64 datos para después procesar los resultados mediante la opción Postsim de SGeMS y así poder obtener distintos estadísticos.

A continuación, se muestran 6 de las 100 simulaciones obtenidas a partir del variograma hallado en el apartado 2.2 (figura 115).



*Fig.115. Ejemplo aplicación. Simulaciones obtenidas del campo de permeabilidades (6/100).
(Fuente: Elaboración propia)*

A simple vista se puede observar que los resultados varían ampliamente, aunque la tendencia de anisotropía expuesta a lo largo de la práctica se conserva.

Una vez obtenidas las simulaciones correspondientes, se puede obtener los diferentes estadísticos a través de la opción PostSim.

Primeramente, se estudiará el valor esperado punto a punto de las realizaciones, dando los resultados obtenidos en la figura 111. En ella se puede ver que la media de los resultados se parece a las simulaciones obtenidas. Por ello se puede concluir que este valor esperado se puede utilizar como una buena aproximación del campo de permeabilidades.

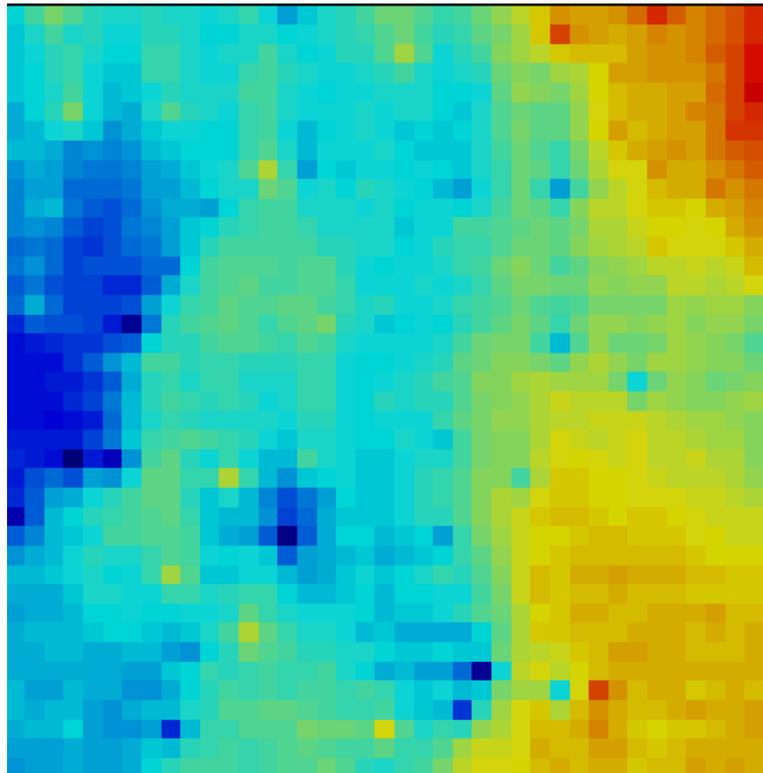


Fig.116. Ejemplo aplicación. Valor esperado estimaciones. (Fuente: Elaboración propia)

A continuación, se muestra la probabilidad de que la permeabilidad esté por debajo del umbral 45,55 y 65. Se puede observar que tiene sentido los resultados obtenidos ya que las zonas con valores altos de probabilidad corresponden justamente con aquellas que tenían los valores más bajos en la zona oeste.

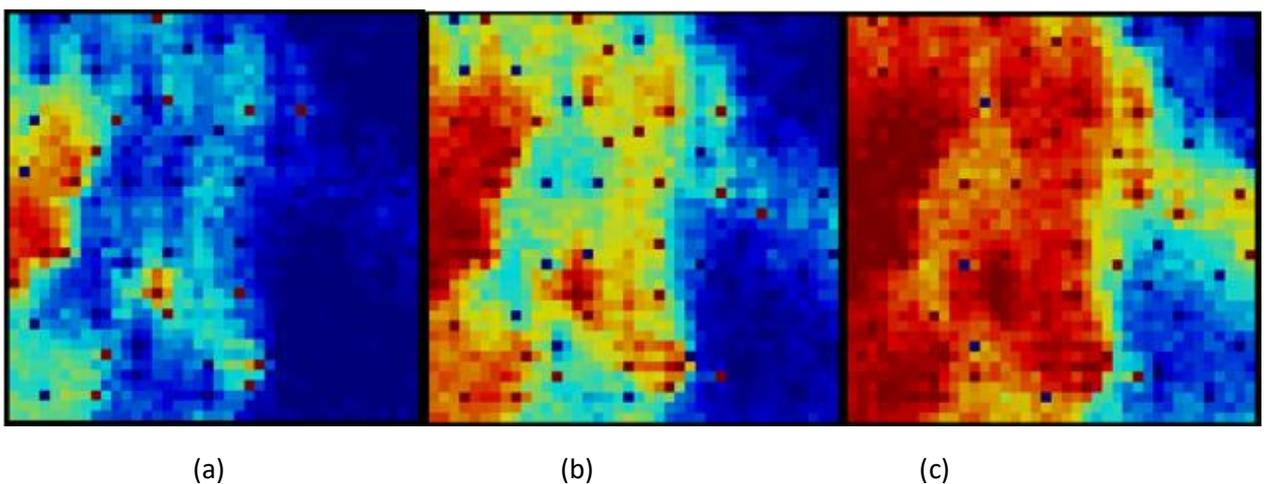


Fig.117. Ejemplo aplicación. Probabilidad permeabilidad tenga un valor menor de 45 (a), 50 (b), 55 (c). (Fuente: Elaboración propia)

En cuanto a los cuantiles (figura 118), la figura nos indica la baja probabilidad de que se den valores altos en la zona de estudio, excepto en aquellas en las que desde un primer momento se vieron que tenían los puntos más altos (Zona NE y SE)

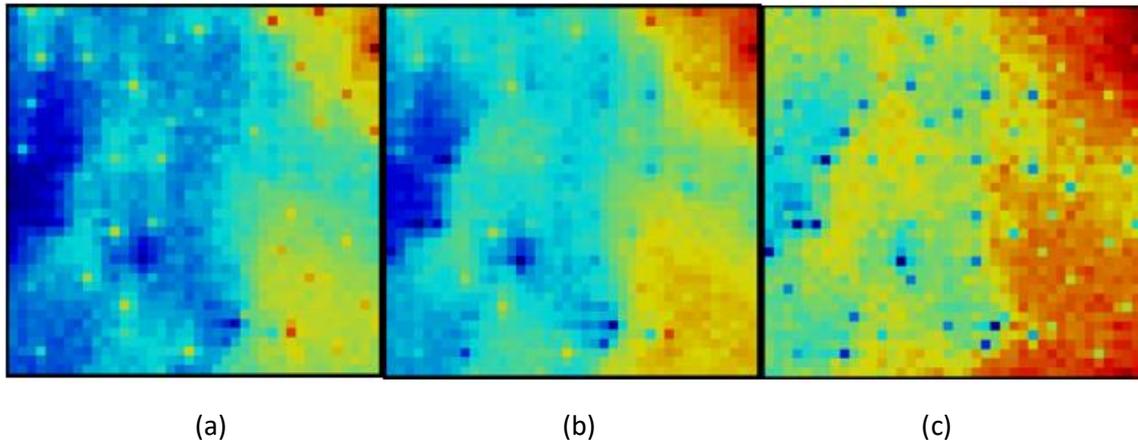


Fig.118. Ejemplo aplicación. Cuantiles correspondientes a las probabilidades 0.2(a), 0.46(b) y 0.95(c).
(Fuente:Elaboración propia)

Para finalizar, se obtienen las varianzas condicionales punto a punto de las 100 realizaciones (figura 119) donde se puede observar que la varianza toma valores muy bajos en la mayor parte de la zona de estudio.

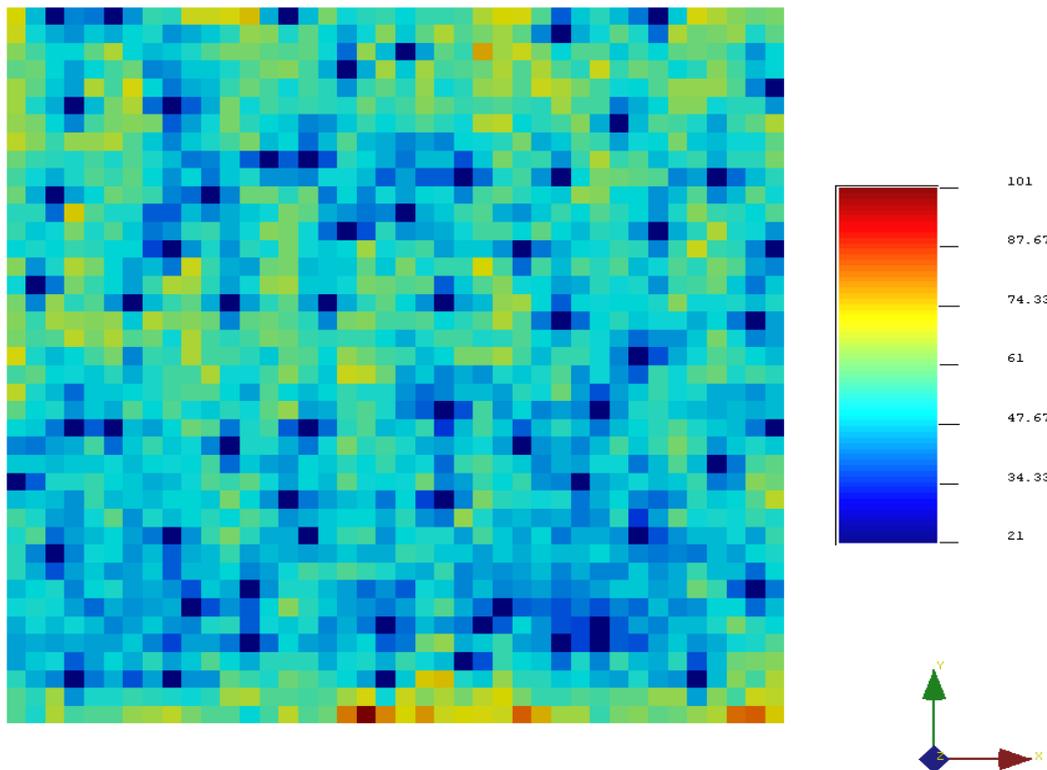


Fig.119. Ejemplo aplicación. Varianzas condicionales. (Fuente: Elaboración propia)

3

La elaboración de mapas de riesgo

Los mapas de probabilidad de superar el umbral de espesores establecido en el apartado 1.2 del presente trabajo, permite delimitar las áreas que no presentan ningún riesgo o un riesgo mínimo en relación con este parámetro. Las técnicas geoestadísticas constituyen una herramienta de uso común para elaborar estos mapas, aunque lamentablemente no hay un acuerdo sobre qué técnica es la más adecuada.

La generación de estos mapas de probabilidad no es una tarea sencilla. Por un lado, se debe de disponer de una red de muestras con una amplia cobertura espacial. Por otra parte, el algoritmo de representación estocástica elegido para estimar un parámetro en los puntos no medidos influye en el mapa resultante. A los problemas mencionados, hay que añadir que todo valor simulado tiene una incertidumbre asociada.

En geoestadística, el enfoque habitual para la modelación de la incertidumbre espacial se lleva a cabo a través de la generación de múltiples realizaciones de la distribución espacial del atributo de interés, proceso que se denomina representación estocástica o simplemente simulación. Las representaciones generadas se utilizan después como entradas a funciones de transferencia, estudiando la distribución de los valores de respuesta.

El conjunto de múltiples imágenes de la realidad desconocida generadas por simulación estocástica provee una medida visual y cuantitativa de la incertidumbre espacial del atributo modelizado.

Dicha incertidumbre puede visualizarse a través de los mapas de probabilidad en el cual en cada nodo simulado, la probabilidad de no llegar a un determinado umbral se calcula como la proporción de valores que están por encima o debajo de ese umbral considerando el conjunto de realizaciones, obteniendo así los mapas de riesgo deseados, y como resultado, la probabilidad de que en un área determinada los espesores de las capas de arcilla, en el caso a tratar a continuación, u otro parámetro no sea inferior a 10m por ejemplo .

Los diferentes algoritmos de simulación y su base teórica han sido ya explicados en el apartado 2.7. "Simulación secuencial" del presente estudio.

4

Aplicación práctica a un vertedero real

4.1. Introducción

El objetivo de este apartado es mostrar una aplicación de las técnicas geoestadísticas para la definición de un área en la cual un nuevo vertedero RSU, en la región de Valencia, pueda ser construido, de tal manera que no suponga ninguna amenaza y cumpla los requisitos establecidos en el Real Decreto 1481/2001, de 27 de diciembre, indicados en el apartado 1.2 del presente trabajo.

Para ello, diferentes técnicas de estimación y simulación han sido utilizadas sobre un conjunto de datos obtenidos a través de estudios de campo y laboratorio, pudiendo así proponer unos nuevos límites, donde las probabilidades de superar un valor determinado de una cierta característica, sea lo suficientemente elevada.

El procesamiento geoestadístico, como bien se ha indicado en puntos anteriores, se ha elaborado mediante el programa SGeMS (Stanford Geostatistical Modelling Software).

4.2. Estudio geológico de la zona de ampliación del vertedero

Con el objeto de determinar la estratigrafía y estructura del potencial emplazamiento ha sido realizada una campaña de 32 sondeos a testigo continuo, con una profundidad media de 40 m (La situación de los sondeos se representa en el plano nº1). Con ello, los materiales atravesados han sido identificados y caracterizados geológica y litológicamente. En dichos sondeos, se ha determinado un total de 6 formaciones geológicas:

- Formación 1 (Fm.1): Margas, arcillas, arenas y bolos calizos.
- Formación 2(Fm.2): Arcillas rojas.
- Formación 3(Fm.3): Conglomerado arcilloso y paleocanales intercalados.
- Formación 4(Fm.4): Conglomerante carbonatado.
- Formación 5(Fm.5): Calizas y dolomías.
- Formación 9(Fm.9): Arcillas, gravas y arenas.

Además de los sondeos, se ha determinado la permeabilidad mediante ensayos "in situ" tipo Lefranc de carga variable para cada sondeo y ensayos de laboratorio de 59 muestras representativas de las arcillas de la formación 2 y algunos tramos de la formación 3.

De los resultados de estos ensayos se puede concluir lo siguiente:

- La formación Fm. 2, presenta valores de permeabilidad comprendidos entre 10^{-9} y 10^{-13} m/s.
- Los valores calculados de permeabilidad en la formación Fm. 3 ocupan un rango entre 10^{-8} y 10^{-12} m/s. Estimándose los valores más altos, 10^{-8} m/s, en el borde NE de la zona de estudio (sondeos ST-1, ST-2, ST-20, ST-25, ST-26, ST-27 y ST-28).
- La formación Fm. 4 presenta valores de permeabilidad entre 10^{-8} y 10^{-12} m/s, presentando los valores más altos también en el borde NE de la zona de estudio (sondeos ST-1, ST-2, ST-18, ST-23 y ST-24).
- La formación Fm. 5, calizas del Cretácico, presenta una permeabilidad en el ST-3 de $1,26 \times 10^{-7}$ m/s.

De ello se obtiene que para los fines citados, los niveles arcillosos de la Fm.2 son los únicos en conjunto y de forma general, que cumplen ampliamente la normativa vigente para la ejecución de un vertedero de estas características, ya que ésta prescribe un mínimo de 1m de la barrera geológica si ésta dispone de una permeabilidad igual o inferior a 10^{-9} m/sg

Además, en dicho estudio, se recomienda que el vaso del vertedero se apoye sobre un mínimo de 10m de las arcillas de la formación 2, cuya permeabilidad está comprendida entre 10^{-9} a 10^{-13} m/s.

A continuación, en la tabla 10 se muestra los 32 sondeos realizados con sus coordenadas y los diferentes espesores que cumplen las características anteriormente expuestas y que serán utilizados para la delimitación del emplazamiento propuesto. Se puede observar en color gris aquellos valores que no cumplen los valores de permeabilidad y espesor necesarios. De la misma manera aquellos valores no conocidos se les pone de valor -999.99, sin embargo, los datos no conocidos de permeabilidad, cuya formación sea la 2, se les ha supuesto un valor de $1E^{-9}$ debido a que este es el valor mayor que se le ha dado en el estudio a esta formación geológica.

SONDEO	COORD. X	COORD. Y	FORMACIÓN	ESPESOR	PERMEABILIDAD
ST-1	683823	4351088	2	10.8	7.47E-10
ST-2	683892	4351111	2	10.4	1.69E-09
ST-3	683952	4351134	5	0	1.26E-07
ST-4	683958	4350934	5	0	-999.99
ST-5	683442	4350970	2	62.8	2.23E-12
ST-6	683503	4351039	2	73.4	3.47E-13
ST-7	683584	4351093	2	43.1	5.62E-12
ST-8	683622	4351168	2	52.9	1.94E-11
ST-9	683716	4351210	2	37.3	1.24E-11
ST-10	683604	4350914	2+3	53.5	1.1622E-10
ST-11	683805	4350914	3+4	13.4	6.91E+11
ST-12	683805	4351215	2+3	39.1	9.05E-11
ST-13	683687	4350882	3	13.4	1.21E-11
ST-14	683705	4351312	2+3	48	1.48E-11
ST-15	683601	4351276	2+3	27.8	2.005E-11
ST-16	683557	4350901	4	45.5	6.27E-12
ST-17	683784	4350894	2	6.1	1.00E-09
ST-18	683713	4351306	4	0	3.92E-08
ST-19	683685	4351112	2	26	3.17E-10
ST-20	683695	4351357	3	0	1.02E-08
ST-21	683580	4350988	2	31.8	1.59E-10
ST-22	683785	4351157	2	16.5	1.65E-09
ST-23	683564	4351344	2	6.6	4.41E-11
ST-24	683654	4351248	4	0	1.03E-08
ST-25	683756	4351199	2	4.6	8.86E-10
ST-26	683833	4351236	2	6	1.00E+09
ST-27	683808	4351115	2	10	1.16E-11
ST-28	683811	4351051	2	17.7	9.62E-11
ST-29	683763	4350961	2	22	6.54E-11
ST-30	683444	4351216	2	17.3	9.64E-10
ST-31	683516	4351081	1	0	-999.99
ST-32	683522	4351219	2+3	11.5	5.845E-10

Tabla.10. Ejercicio aplicación. Datos de los sondeos

En círculos rojos se muestran aquellos puntos donde el espesor del estrato con permeabilidad superior a $1 \cdot 10^{-9}$ m/sg no cumplen los requisitos establecidos por el Real Decreto (espesor mayor de 1 m). Y en verde, aquellos que aun cumpliendo la normativa (espesor >1m) no llegan a la recomendada (espesor >10m). Con ello, se va tomando una idea de la forma que puede tomar la delimitación final.

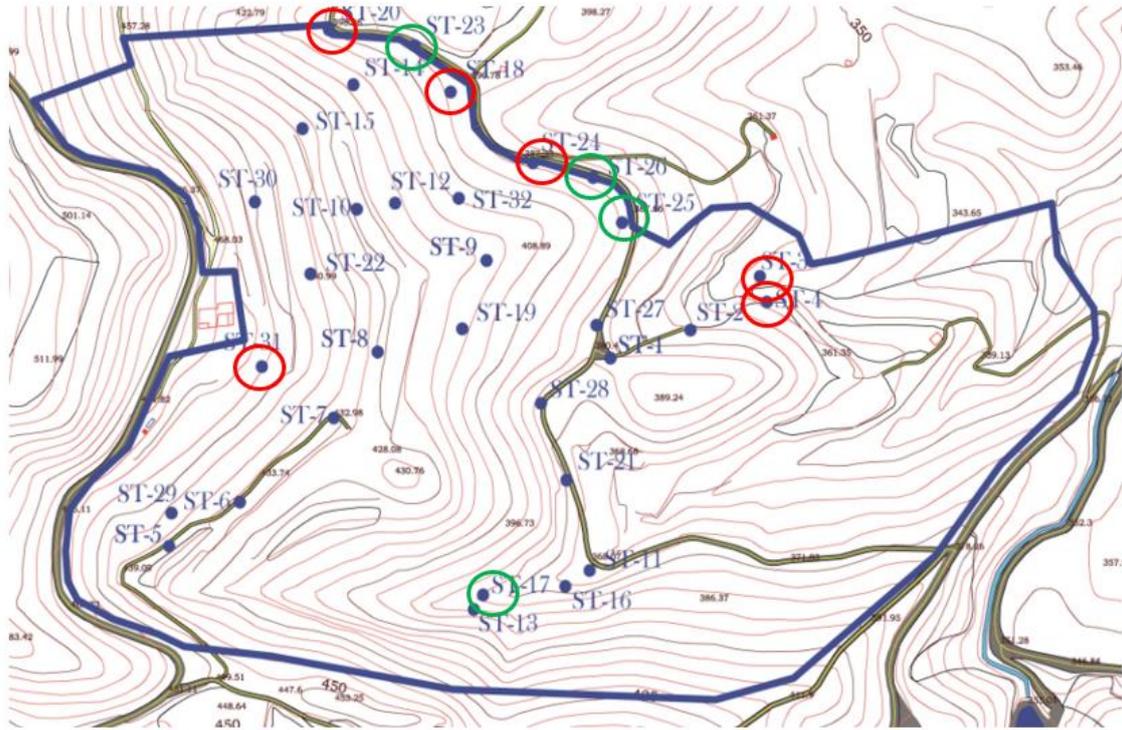


Fig.120. Ejemplo aplicación. Señalización de sondeos no aptos según normativa y según las recomendaciones. (Fuente: Elaboración propia)

Las Figuras 121 y 122 muestran los valores y la distribución de las dos variables anteriores, espesor y permeabilidad.

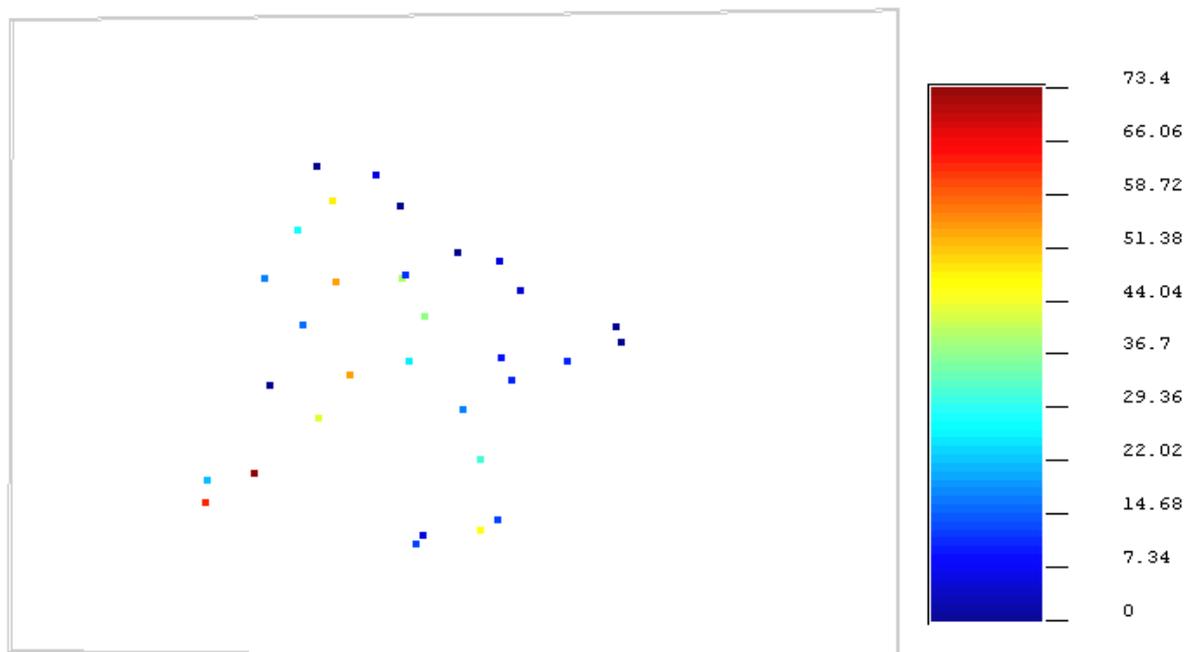


Fig. 121. Ejercicio Aplicación. Datos espesores disponibles. (Fuente: Elaboración propia)

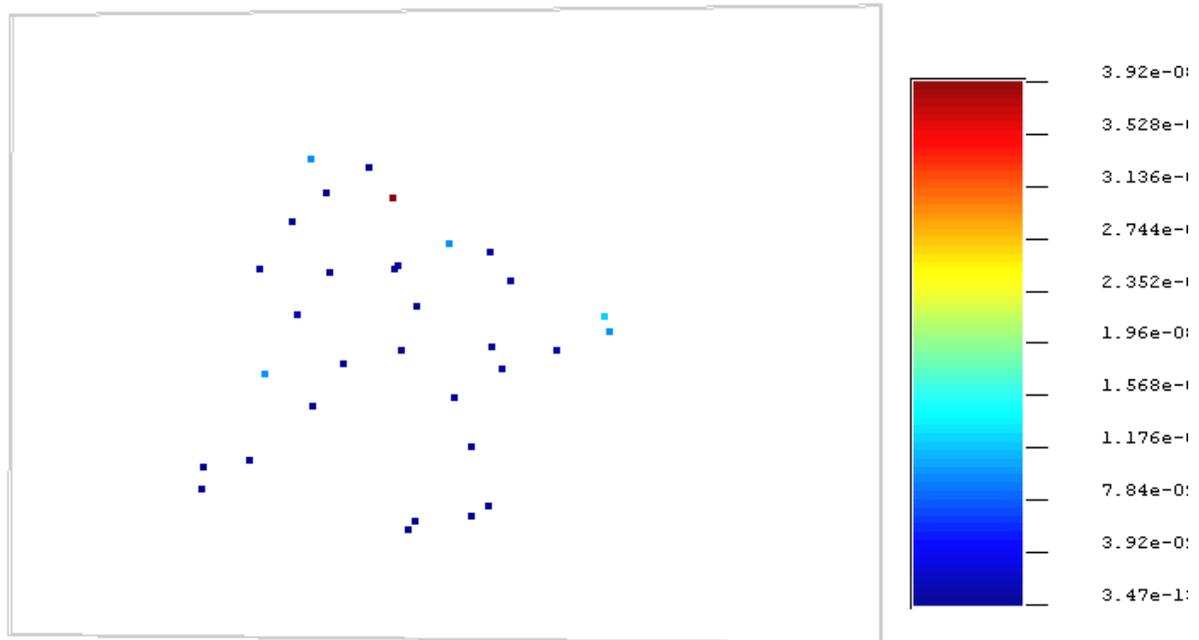


Fig. 122. Ejercicio Aplicación. Datos permeabilidad disponibles. (Fuente: Elaboración propia)

4.3. Estudio geoestadístico

4.3.1. Análisis Univariado

En la figura 121 Y 122 se pueden observar el valor del espesor y permeabilidad de los firmes obtenidos en cada sondeo que cumplen las condiciones establecidas para la ubicación de un vertedero.

Se puede observar que los picos de valores, para la variable espesor, se encuentran en el SW y la zona central de la zona de estudio, teniendo un cambio gradual de 73.4 hasta 0. Sin embargo, para la permeabilidad no se observa ninguna zona que destaque por tener valores muy superiores al resto, aunque se puede observar que la zona norte de los sondeos es donde se concentra una mayor variación de espesores.

Es por ello, que a simple vista se suponga que la dirección de máxima y mínima variación sea entorno a N146W y N56W respectivamente para la porosidad, y para la permeabilidad N58E y N148E.

A continuación, se estudia el histograma con su correspondiente función de probabilidad acumulada y sus estadísticos para los 32 datos. (Figuras 123 y 124)

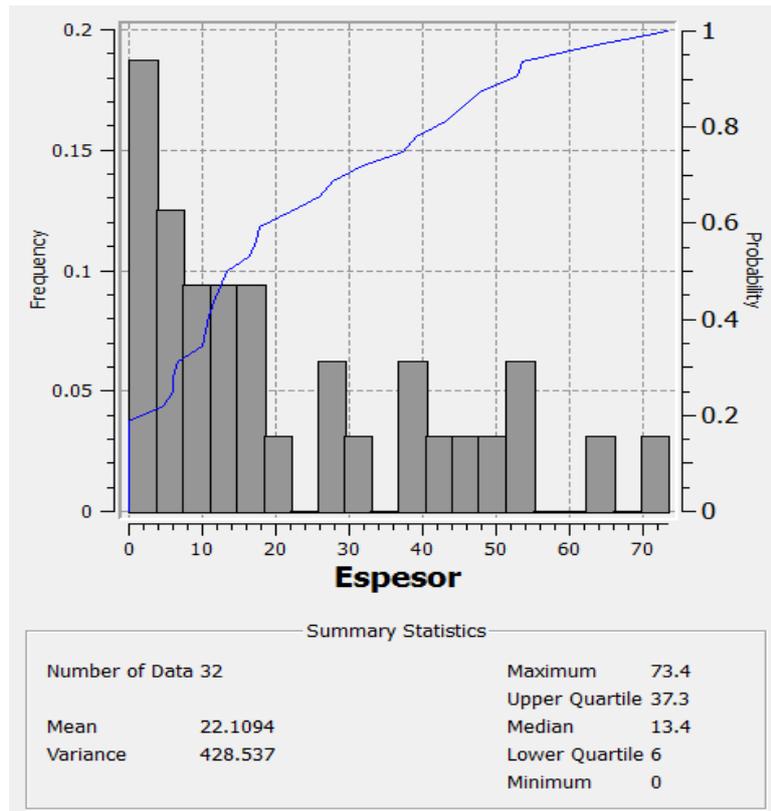


Fig. 123. Ejercicio Aplicación. Histograma variable espesor. (Fuente: Elaboración propia)

Comenzando con la variable espesor, las medidas de posición (media y mediana) indican el valor central de la distribución, y en este caso se puede observar una diferencia notable entre ambas, correspondiéndole a la media un valor de 19.66 y 10.8 para la mediana. Esto significa que la distribución no es muy simétrica y por lo tanto, la valoración media de la porosidad no sería correcta. Además, el coeficiente de variación de la media es de 454.604, lo que muestra la heterogeneidad de los valores de la variable.

En cuanto al histograma de la permeabilidad (figura 124), se ha estudiado en un diagrama lognormal (figura 125) ya que los valores a escala normal no se pueden apreciar de forma correcta, estando la mayoría de los valores concentrados entorno a los valores más bajos de ésta (80%). En cuanto a la media ($3,17 \cdot 10^{-9}$) y la mediana ($9,62 \cdot 10^{-11}$), aunque se habla de diferencias con valores muy bajos, es muy relativo al tratarse con un rango de valores con dichas magnitudes, considerando la variación bastante notable. Lo mismo sucede con el coeficiente de variación de $5,78 \cdot 10^{-17}$, lo que nos muestra a simple vista homogeneidad en los datos obtenidos, pero al fijarse más detenidamente en la magnitud de los datos tratados, quizás no sean tan homogéneos.

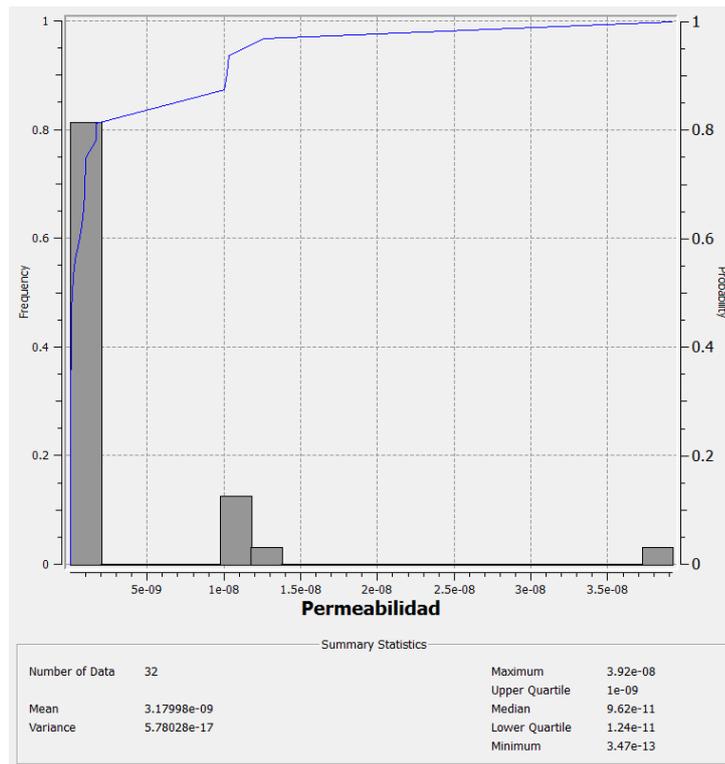


Fig.124. Ejercicio Aplicación. Histograma variable espesor. Diagrama normal (Fuente :Elaboración propia)

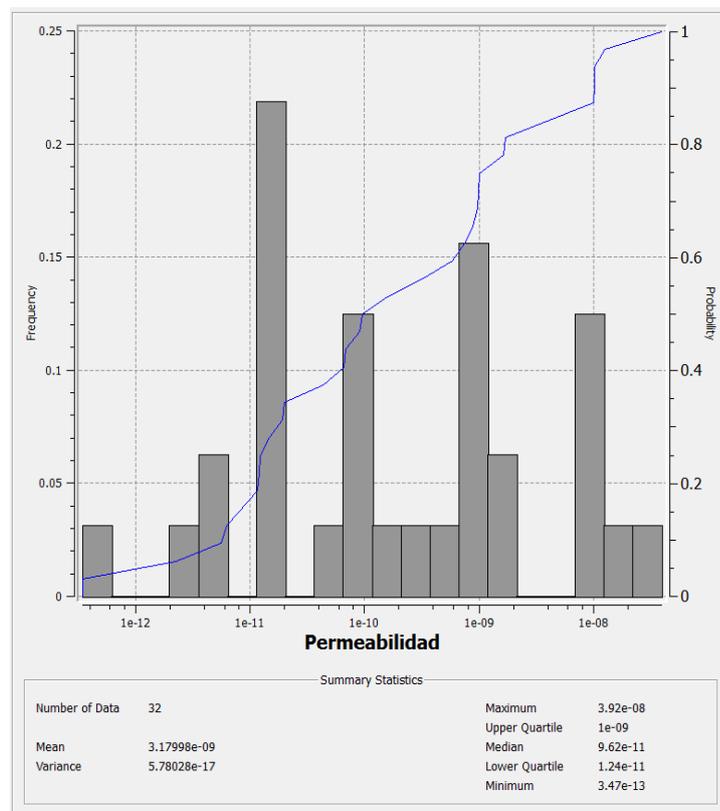
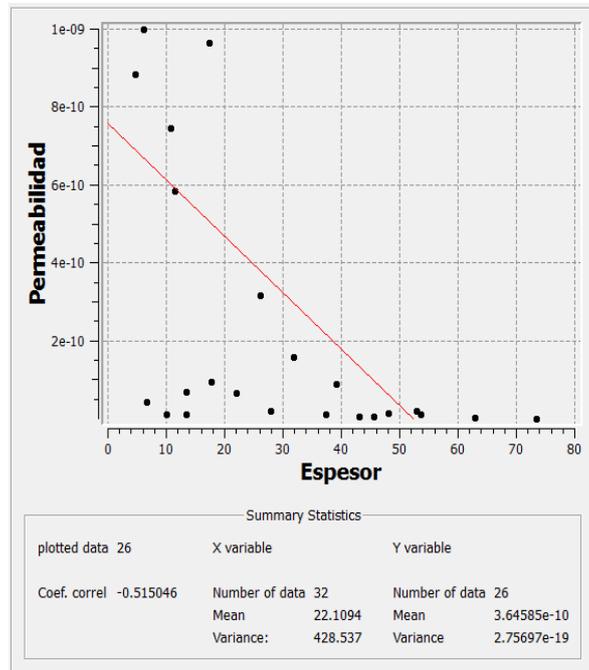


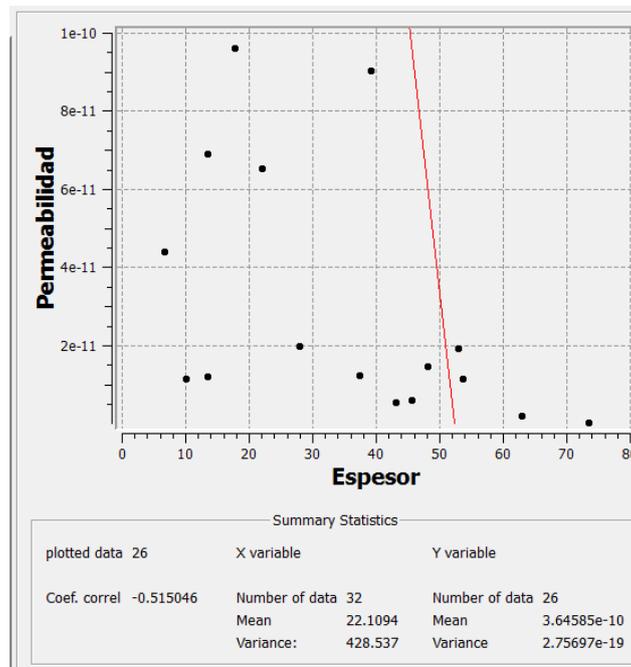
Fig. 125. Ejercicio Aplicación. Histograma variable espesor. Diagrama lognormal (Fuente: Elaboración propia).

4.3.2. Análisis bivariado

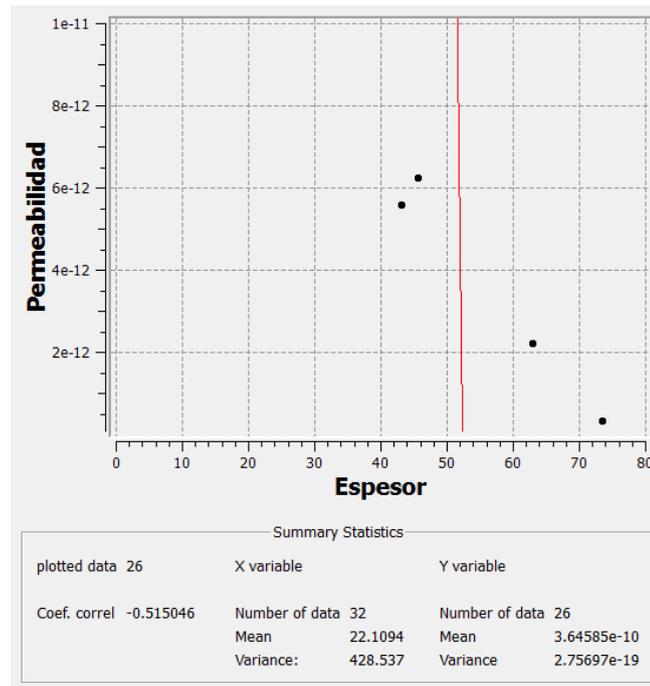
Para cuantificar las relaciones existentes entre las variables de permeabilidad y espesor, se realiza un análisis bivariado de la información disponible. En la Figura 126.a se muestra el diagrama de dispersión para las 32 parejas de datos. A primera vista se puede observar que la correlación entre ambas variables no existe, aunque según los estadísticos proporcionados, el coeficiente de correlación de -0.51.



(a)



(b)



(c)

Fig. 126. Ejercicio Aplicación. Diagrama de dispersión permeabilidad-espesor. Valores permeabilidad (a) $(10^{-8}-10^{-9})$, (b) $(10^{-10}-10^{-11})$ y $(10^{-11}-10^{-12})$. (Fuente: Elaboración propia)

Debido a que la mayoría de los valores se encuentran en un rango inferior, se van disminuyendo para observar si se encuentra algún grado más de correlación entre ambas variables 126.b y 126.c.

A la vista de los resultados, se concluye que no existe correlación entre ambas variables, por lo tanto, el resto del estudio se realizará a partir de una variable, en este caso el espesor, ya que los datos de los que se dispone son más precisos y no ha sido necesario la realización de ninguna media ni extrapolación de resultados.

A pesar de ello, se obtiene el diagrama de cuantiles (Figura 120) para las parejas de valores de la permeabilidad y espesores. En este diagrama los cuantiles de dos distribuciones son representados uno frente el otro, obteniendo parejas de valores con la misma probabilidad experimental de ocurrencia. Como las distribuciones de probabilidad de la permeabilidad y espesores no son idénticas, el diagrama de cuantiles no presenta una distribución lineal.

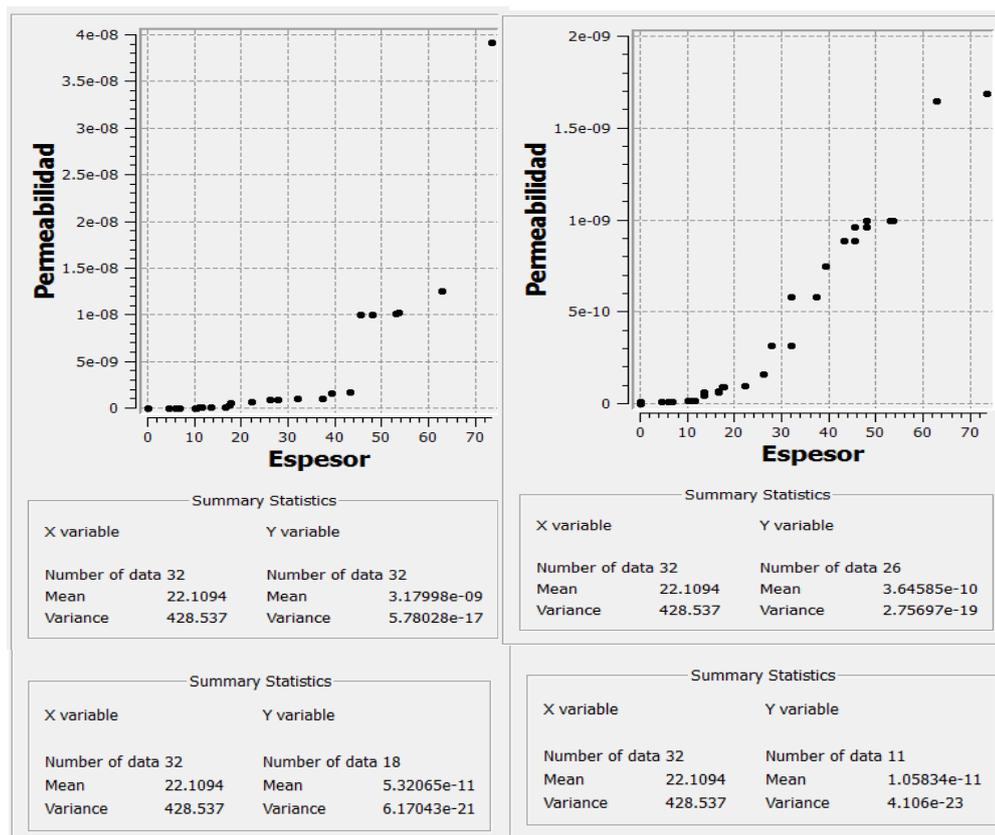


Figura 127. Ejercicio Aplicación. Diagrama de cuantiles. (Fuente: Elaboración propia)

4.3.3. Análisis estructural

En este apartado se busca obtener un modelo de continuidad para la variable espesor.

Para ello, primeramente, se obtiene el variograma omnidireccional del espesor con el cual se busca obtener el alcance y el vector separación. Con ello se construirán diferentes variogramas direccionales y a partir de él, se evaluará la existencia o no de anisotropía y entonces obtener las direcciones de máxima y mínima continuidad espacial.

Una vez terminado el análisis de los variogramas experimentales, se ajustará el modelo correspondiente, pudiendo así escribir la ecuación del modelo ajustado.

4.3.3.1. Variograma omnidireccional.

Para la obtención del variograma omnidireccional, se necesita caracterizar el vector separación, para ello es necesario establecer el número de intervalos a calcular y el módulo de dicho vector.

Siguiendo los criterios indicados en el apartado 2.8.2.1. "Variograma omnidireccional" del presente estudio, y tras realizar diversas pruebas, se ha obtenido como valores óptimos:

Número de intervalos	16
Vector separación	25
Tolerancia	12.5

Tabla 11. Ejercicio Aplicación. Valores óptimos variograma omnidireccional. (Fuente: Elaboración propia)

Que aun no cumpliendo todos los puntos con uno de los criterios, el que hace referencia al número de pares encontrados para obtener un estimador fiable y estable, cumple el resto de recomendaciones establecidas anteriormente. Se he supuesto que al disponer de un número de puntos poco elevado, aumenta la dificultad en alcanzar dicho rango de valores.

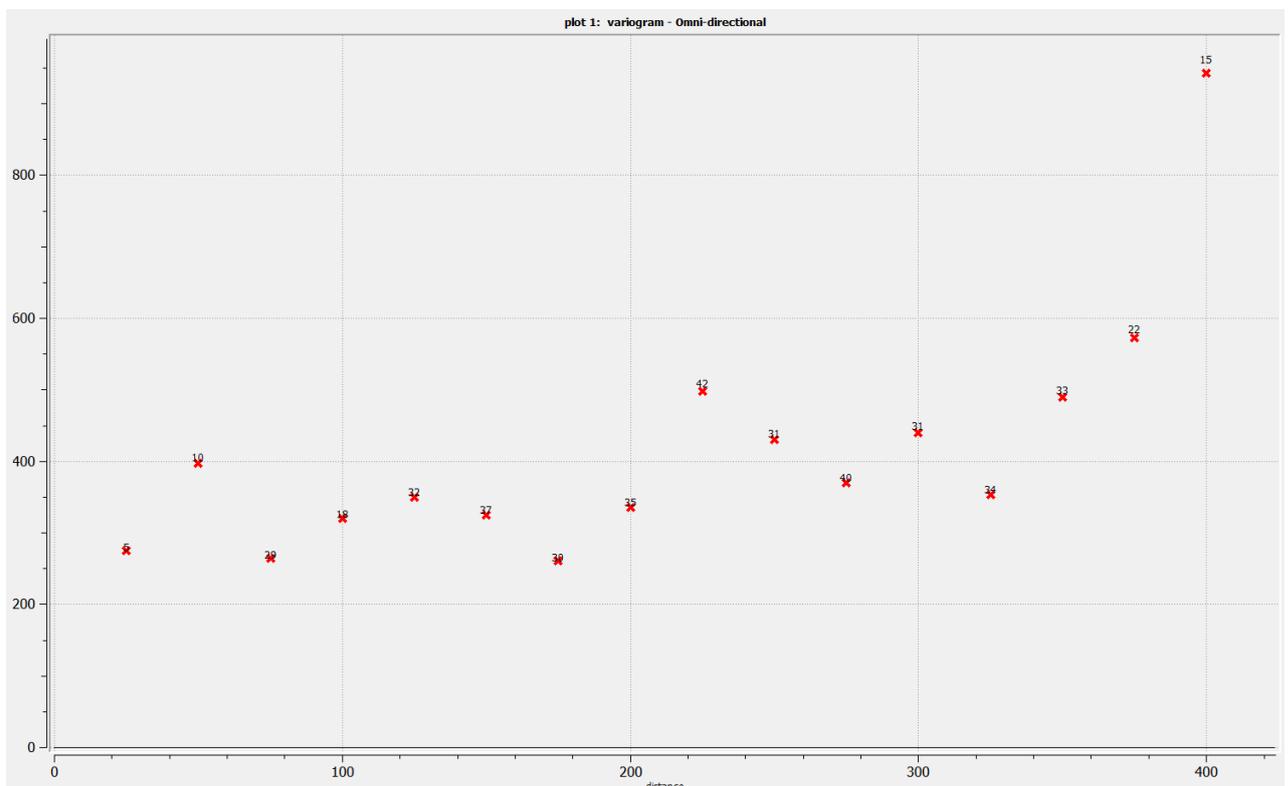


Fig.128. Ejercicio Aplicación. Variograma omnidireccional variable espesor. (Fuente: Elaboración propia)

A continuación, se le ajusta un modelo para poder obtener los parámetros necesarios para la obtención de las direcciones de máxima y mínima continuidad del apartado siguiente, como puede ser el efecto pepita y la meseta (350). Obteniendo el siguiente variograma ajustado de la figura 129.

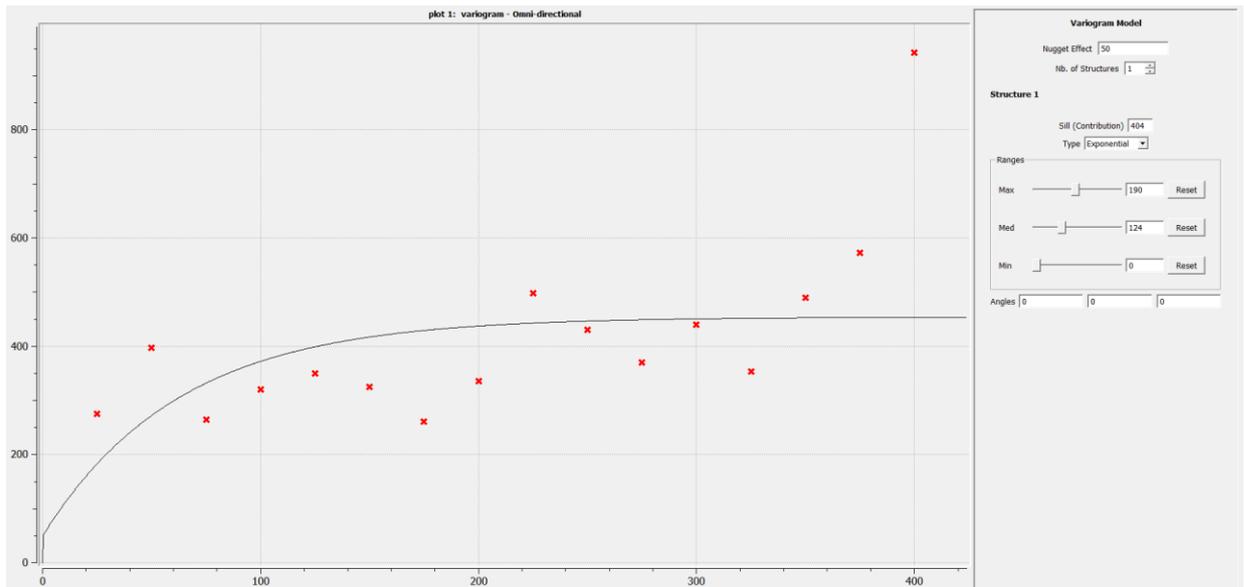


Fig.129. Ejercicio Aplicación. Variograma omnidireccional ajustado. (Fuente: Elaboración propia)

4.3.3.2. Variograma direccional.

A continuación se estudiará si el fenómeno a estudiar es isótropo. Esto se realiza a través del variograma direccional, el cual se obtiene realizando el paso anterior, pero para distintas direcciones y observando si el alcance para el cual se pierde la correlación espacial coincide para todas las direcciones. En caso contrario, se dirá que el fenómeno es anisótropo y se deberá determinar las direcciones de máxima y mínima continuidad a través de la elipse de anisotropía.

Para ello se toma una tolerancia angular de 30° y un ancho de banda de 50, además de los valores obtenidos en el apartado anterior. A continuación, en la tabla 11, se muestra las distintas direcciones tomadas y los alcances correspondientes a cada dirección. Como se puede observar, los alcances toman diferentes valores, luego el fenómeno de estudio es anisótropo, por lo tanto es necesario la construcción de la elipsoide de anisotropía para poder obtener la dirección de máxima y mínima continuidad.

Direcciones	Alcance
N0W	244
N30W	320
N60W	350
N90W	295
N120W	270
N150W	205
N180W	292

Tabla.12. Ejercicio Aplicación. Direcciones y alcances obtenidos variogramas direccionales. (Fuente:Elaboración propia)

En la figura 130 se muestran los variogramas obtenidos para cada una de las direcciones.

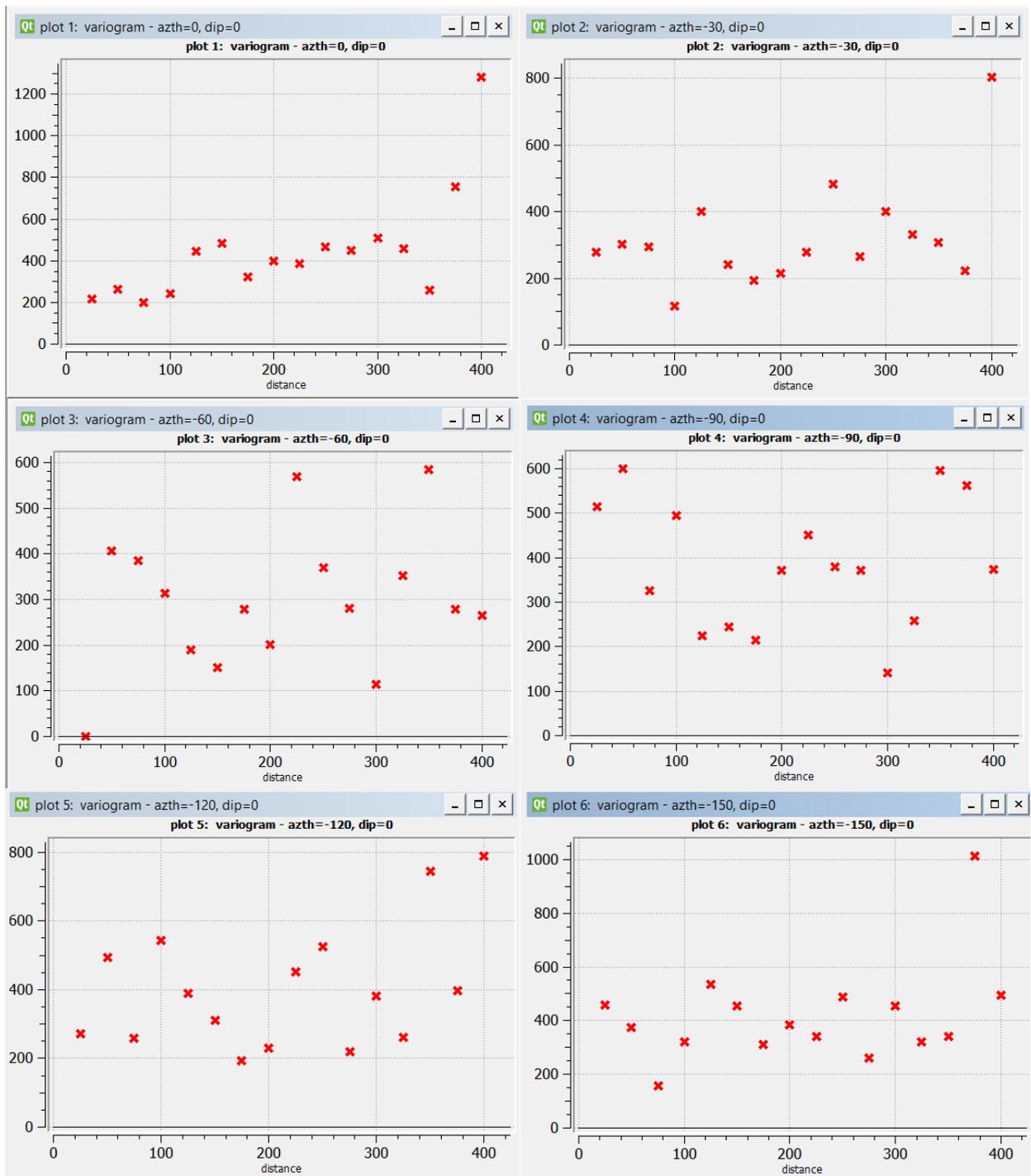


Fig. 130. Ejercicio Aplicación. Variogramas direccionales. (Fuente: Elaboración propia)

A partir de los valores del alcance y sus respectivas direcciones se realiza el elipsoide de anisotropía (figura 131) de donde se obtienen las direcciones de máxima y mínima continuidad, N57W y N33E respectivamente.

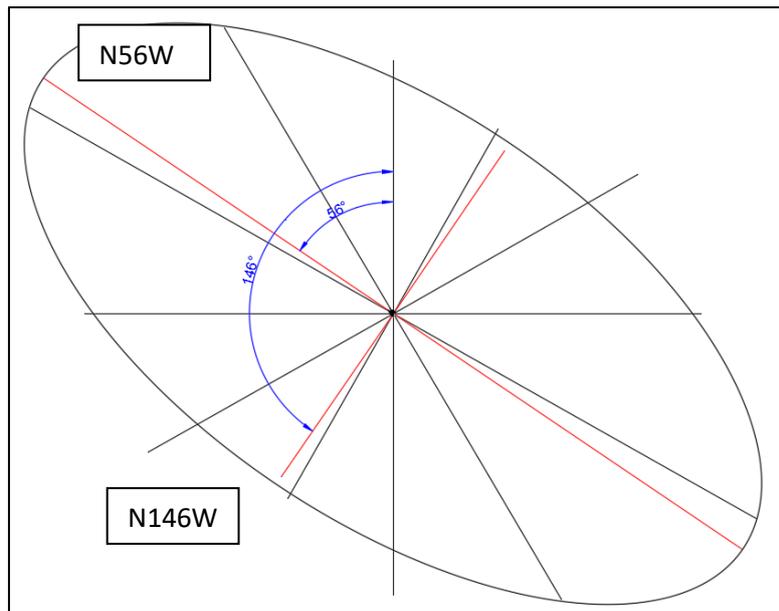


Fig. 131. Ejercicio Aplicación. Elipsoide de anisotropía. (Fuente: Elaboración propia)

A continuación, se muestran los variogramas correspondientes a las direcciones de máxima y mínima continuidad ajustados al modelo.

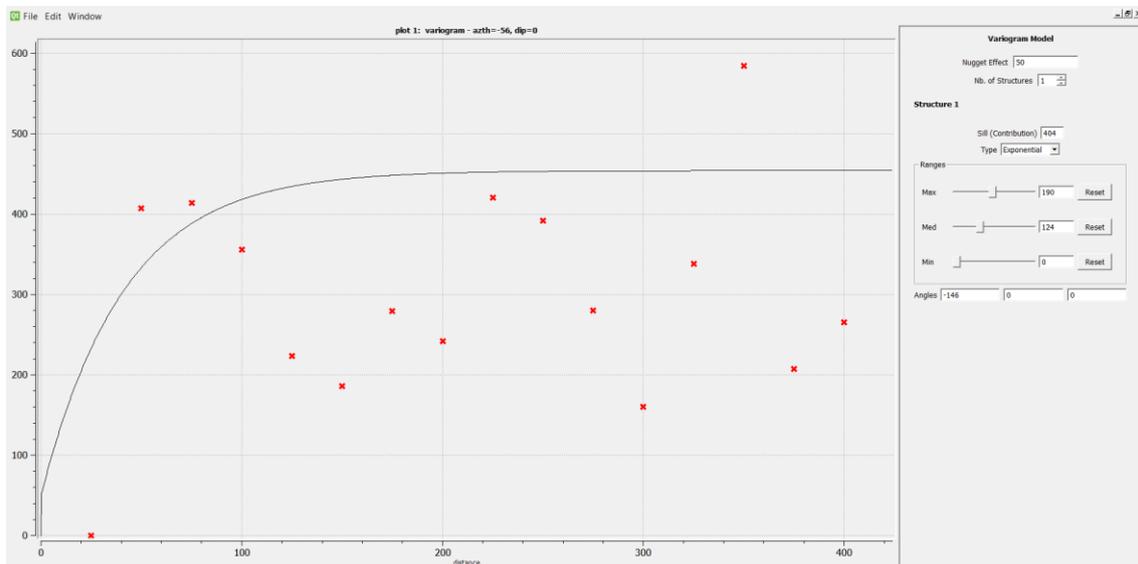


Fig. 132. Ejercicio Aplicación. Variograma direccional mínima continuidad ajustada. (Fuente: Elaboración propia)

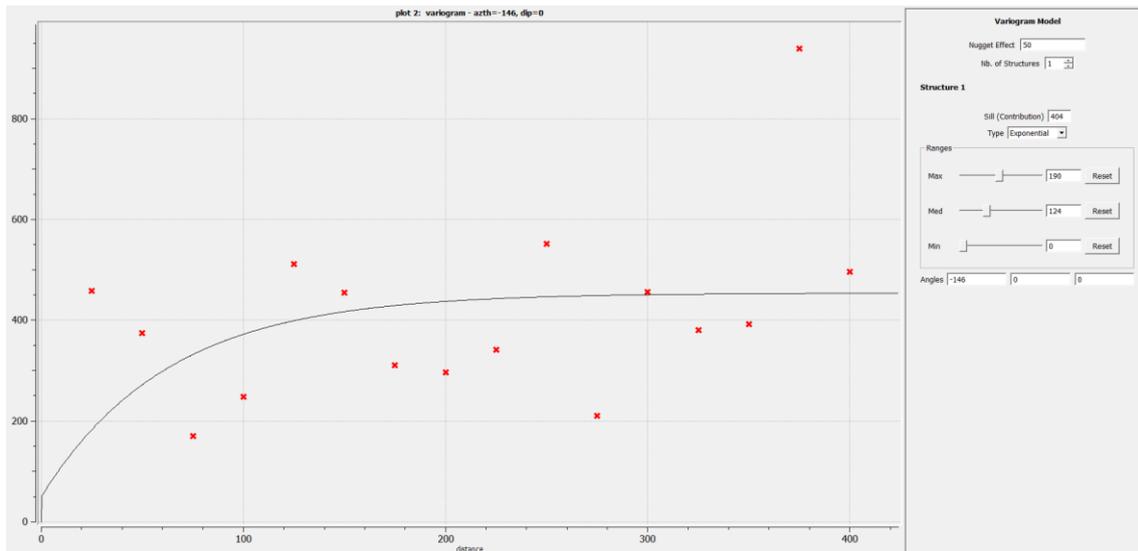


Fig. 133.Ejercicio Aplicación. Variograma direccional máxima continuidad ajustada. (Fuente: Elaboración propia)

Se le ajusta al variograma un modelo exponencial de ecuación:

$$\gamma(h) = 50 + 404 \text{Exp}(h)[a_{max} = 190; a_{min} = 124; \alpha = -146^\circ]$$

Donde:

- Efecto pepita $\rightarrow 50$
- Meseta $\rightarrow 454$ (Contribución 404)
- Alcance de máxima continuidad $\rightarrow 190$
- Alcance de mínima continuidad $\rightarrow 124$
- Dirección de máxima continuidad $\rightarrow -146^\circ$

4.3.4. Krigeado

Una vez analizados los datos y construido el modelo de continuidad espacial, SGeMS permite realizar estimaciones de una variable espacialmente distribuida en aquellas localizaciones donde no la conocemos, siendo el objetivo, la obtención de un mapa de espesores utilizando como datos condicionantes los 32 datos de la variable espesor y el modelo de continuidad espacial que para ella ha sido construido.

A continuación, se realizan diferentes estimaciones mediante Krigeado ordinario para la estimación del campo de espesor. Se ha empleado un krigeado ordinario con un vecindario de búsqueda circular de 50.

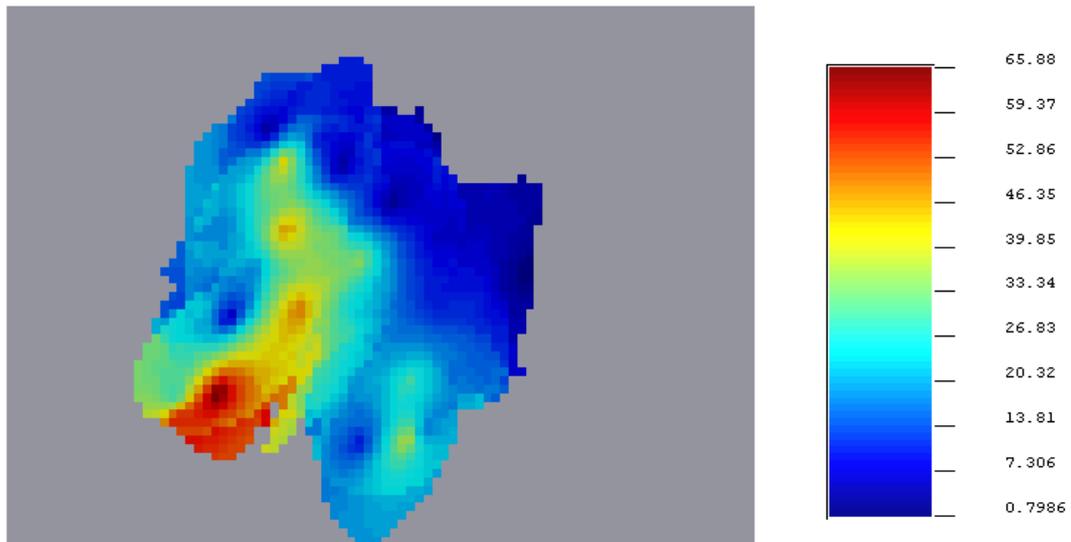


Fig. 134. Ejercicio Aplicación. Campo permeabilidad estimado mediante Krigeado ordinario.
(Fuente: Elaboración propia)

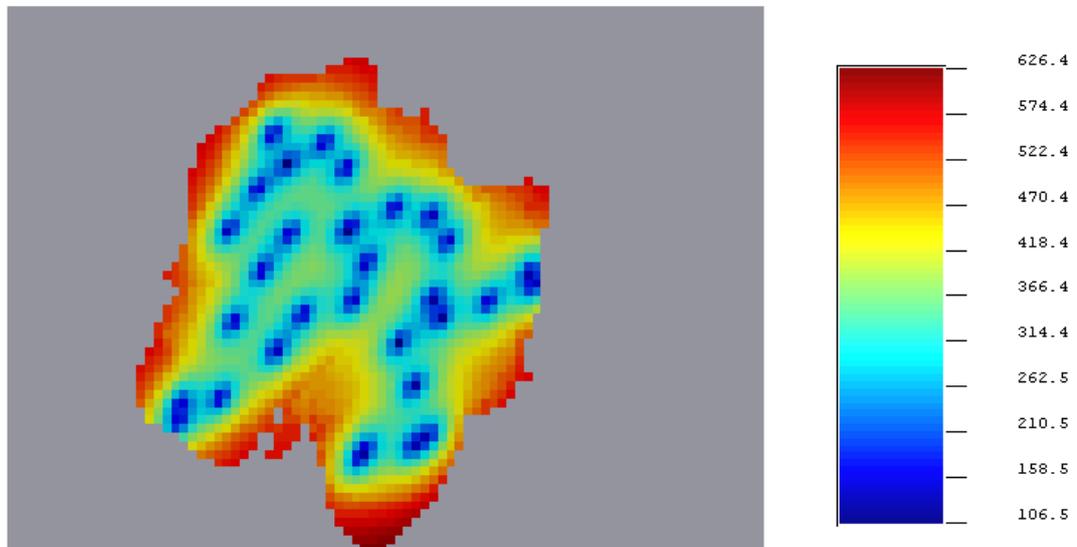


Fig. 135. Ejercicio Aplicación. Varianza campo de permeabilidad estimado mediante Krigeado ordinario.
(Fuente: Elaboración propia)

Como cabía esperar, se puede observar en la Fig. 134, en la dirección de máxima continuidad (N146W) la variación de los valores de espesores es mínima, mientras que en su perpendicular (N56W) es máxima, coincidente con la dirección de mínima continuidad.

En cuanto a la varianza de los datos (Fig. 135), los mayores valores se encuentran en los extremos, disminuyendo a medida que se acerca a la zona donde se realizaron los sondeos, lo

que tiene sentido a la vista de la definición de varianza como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

Con todo esto y mediante la opción de PostKriging que proporciona S-Gems se obtiene el campo de probabilidades de que el espesor sea superior a 1m.

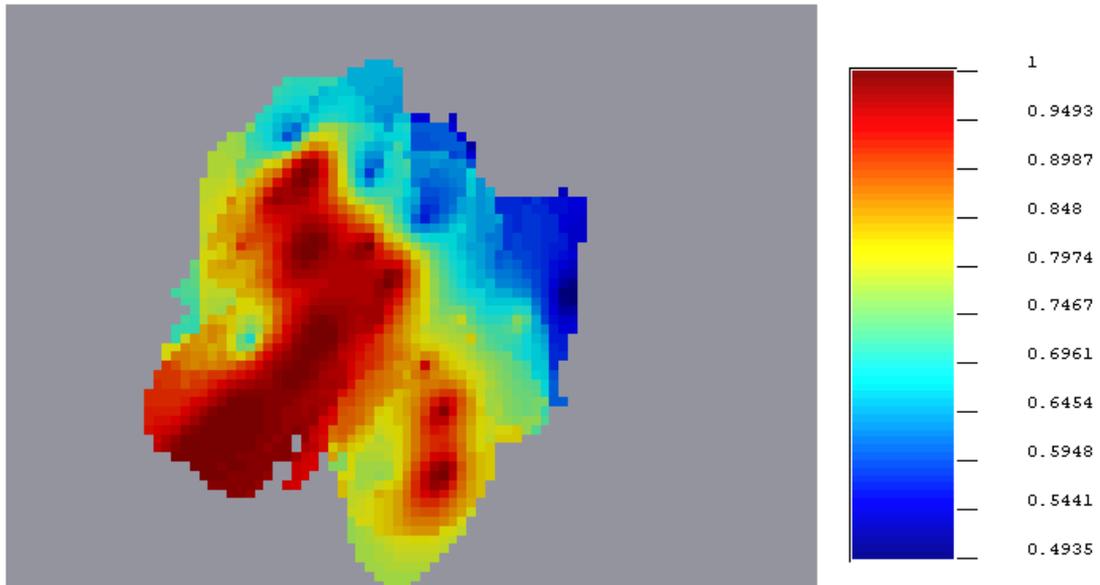


Fig. 136. Ejercicio Aplicación. Probabilidades estimadas campo espesor sea superior a 1 m.
(Fuente: Elaboración propia)

De la misma manera se puede obtener las probabilidades de ocurrencia para los espesores deseados.

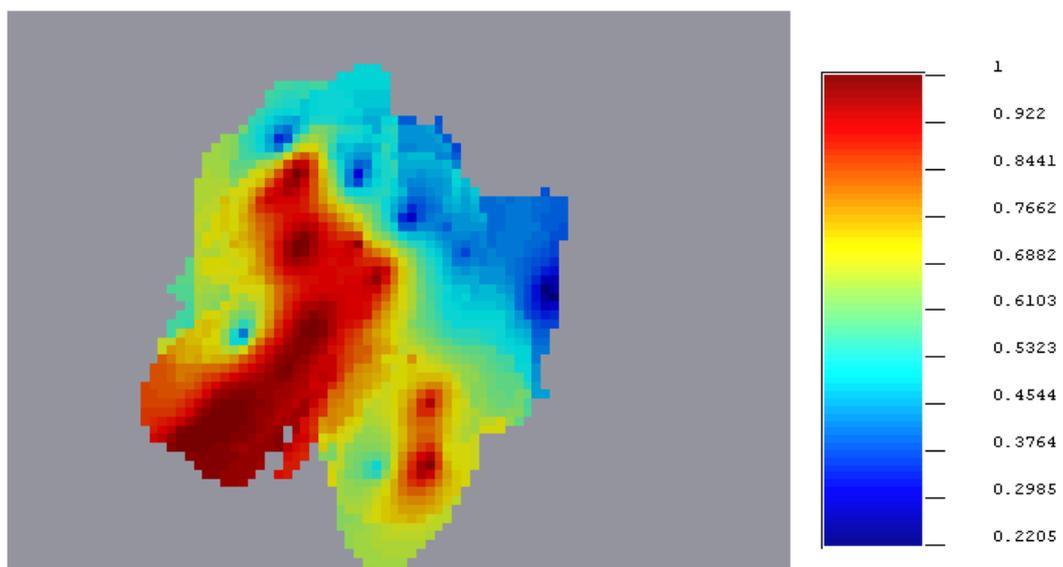


Fig. 137. Ejercicio Aplicación. Probabilidades estimadas campo espesor sea superior a 10 m.
(Fuente: Elaboración propia)

Ubicando los campos de probabilidad en la ubicación correspondiente, se obtiene los mapas de riesgo deseados (Plano nº3 y 4). En este caso a partir de estimación mediante Kriging ordinario. A continuación, se muestran los mapas de riesgo de que el espesor sea mayor de 1 y el mapa de riesgo de que sea mayor de 10 ya que son los que más interesan en este caso, al ser los que cumplen la normativa y/o siguen las recomendaciones establecidas.

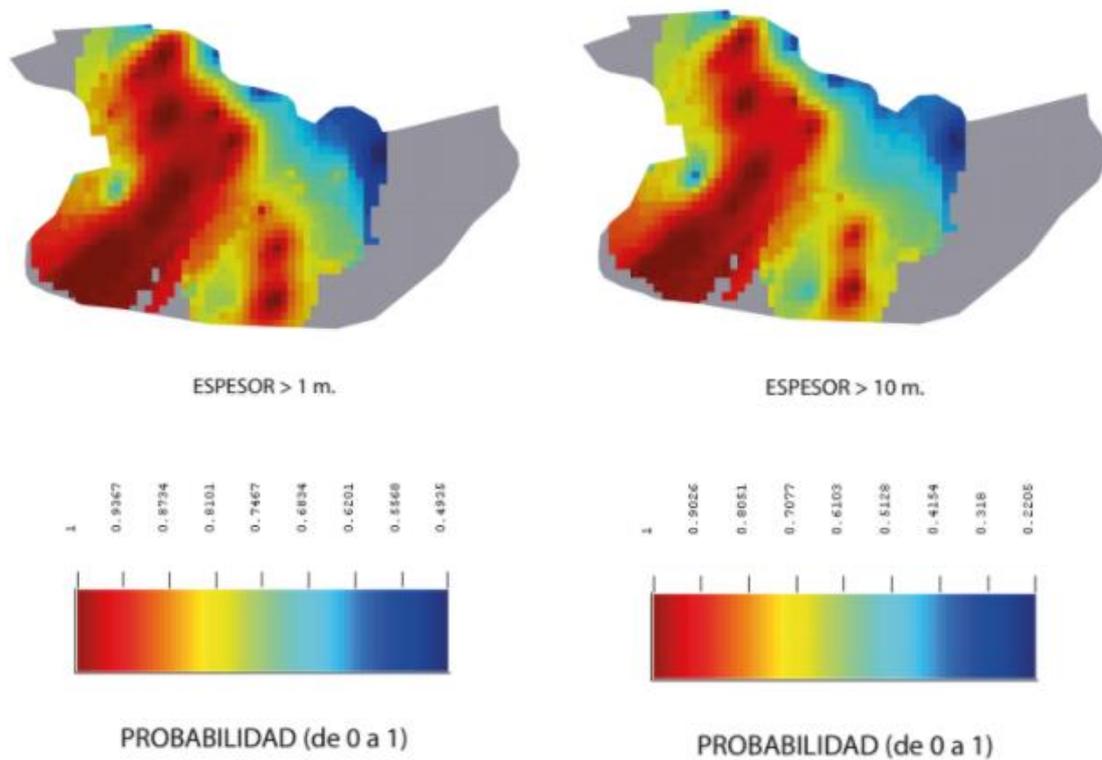


Fig. 138. Ejercicio Aplicación. Mapas de riesgo de superación de 1 y 10m respectivamente.
(Fuente: Elaboración propia)

A la vista de los resultados, uno se puede hacer a la idea del cual es la zona más segura para la ubicación del vertedero, no variando mucho el contorno en el que las probabilidades de superar 1 y 10 m es superior al 90%. Sin embargo, la técnica de estimación utilizada no nos ha llegado a obtener el campo completo ya que, los algoritmos de estimación dan como resultado una única representación del atributo interpolado, dejando construir modelos locales de incertidumbre. En cambio, la simulación estocástica genera múltiples campos de una variable, proveyendo una medida visual y cuantitativa de la incertidumbre espacial. Es por esto por lo que utilizar en nuestro estudio la Simulación Gausiana Secuencial es más óptimo. Aún así, a la hora de tomar la decisión de la delimitación del vertedero, estos casos también se tendrán en cuenta.

4.3.5. Simulación gausiana secuencial

Como bien se ha indicado en el apartado anterior, la simulación estocástica provee una descripción y medida de la incertidumbre de la variabilidad espacial de un fenómeno, en este caso la variable espesor. Se han generado un total de 200 campos de espesor condicionados a los 32 datos utilizando el algoritmo de simulación gausiana secuencial empleando el software S-Gems.

La simulación Gausiana secuencial asume que la variable es Gausiana. Por lo que se comprueba que se cumple la normalidad y binormalidad, asumiendo que el formalismo multigaussiano puede ser utilizado.

A la vista de la función de probabilidad acumulada (figura 139) se puede asumir que sigue una distribución normal ya que se asemeja a una línea recta, por lo que se cumple el criterio de normalidad

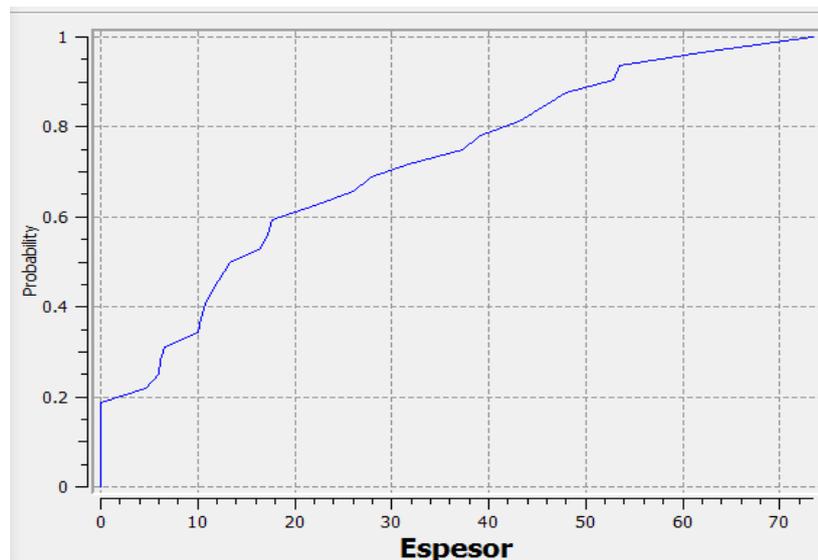


Fig. 139. Ejercicio Aplicación. Función de probabilidad acumulada de la variable espesor. (Fuente: Elaboración propia)

A continuación, se muestran 6 de las 200 simulaciones obtenidas a partir del variograma hallado en el apartado 4.3.3.2 (figura 140)

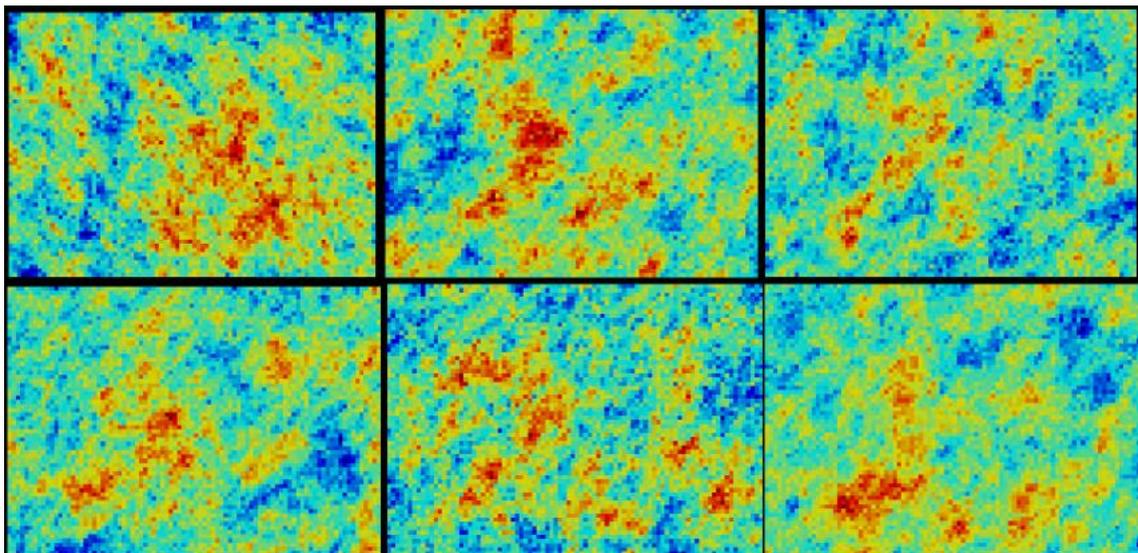


Fig.140. Ejercicio Aplicación. Simulaciones campo de espesores (6/200). (Fuente: Elaboración propia)

A simple vista se puede observar que los resultados varían ampliamente, aunque la tendencia de anisotropía expuesta a lo largo de la práctica se conserva.

En el plano nº 2 se puede observar como una de estas simulaciones queda situada en el terreno disponible.

Tras la generación de las 200 simulaciones, se procesan los resultados mediante la opción Postsim de SGeMS para así poder obtener los diferentes estadísticos y una medida visual de la incertidumbre acerca de la distribución espacial de los datos, el cual finalmente nos dará el resultado buscado: los mapas de probabilidad de riesgo de que el espesor sea superior a diferentes valores, para luego así poder establecer un porcentaje óptimo de riesgo asumible y a partir de él delimitar la zona apta, obviamente disminuyendo dicho área a medida que el porcentaje buscado aumente.

A continuación, en la figura 141, se muestra el mapa de probabilidades de que el espesor sea superior a 1 m, siendo el más restrictivo de las establecidas en las recomendaciones y normativa. En este se puede contemplar una clara ubicación donde las probabilidades de que dichos requisitos se cumplan son bastante elevadas. Además, en el mapa que muestra las probabilidades de ser el espesor mayor a 10m (Fig.142), la ubicación óptima es la misma variando la extensión de esta.

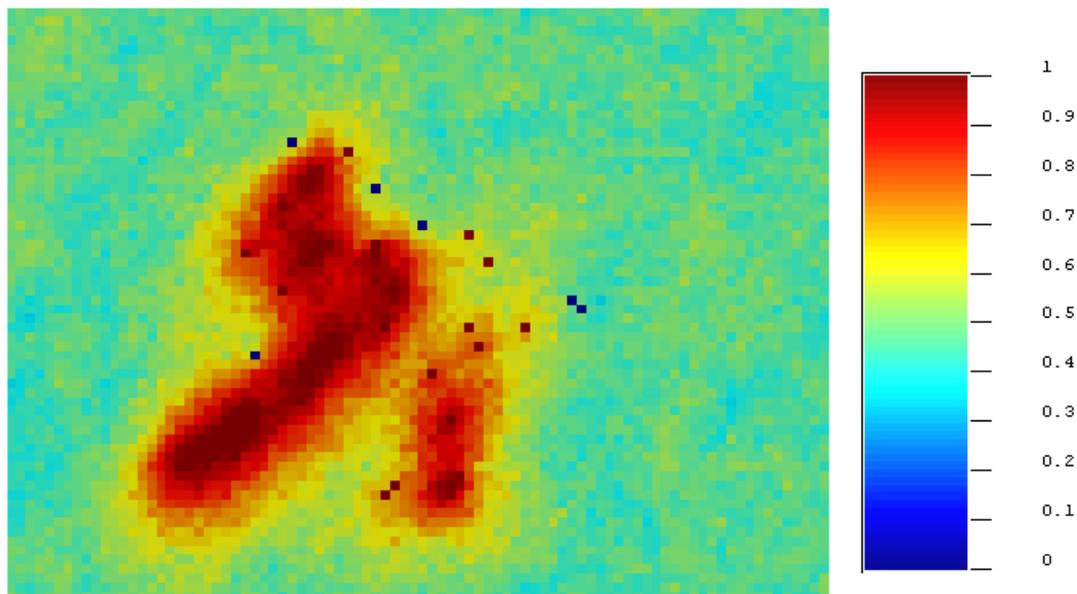


Fig.141. Ejercicio Aplicación. Probabilidad espesores estén por encima del umbral 1m. (Fuente: Elaboración propia)

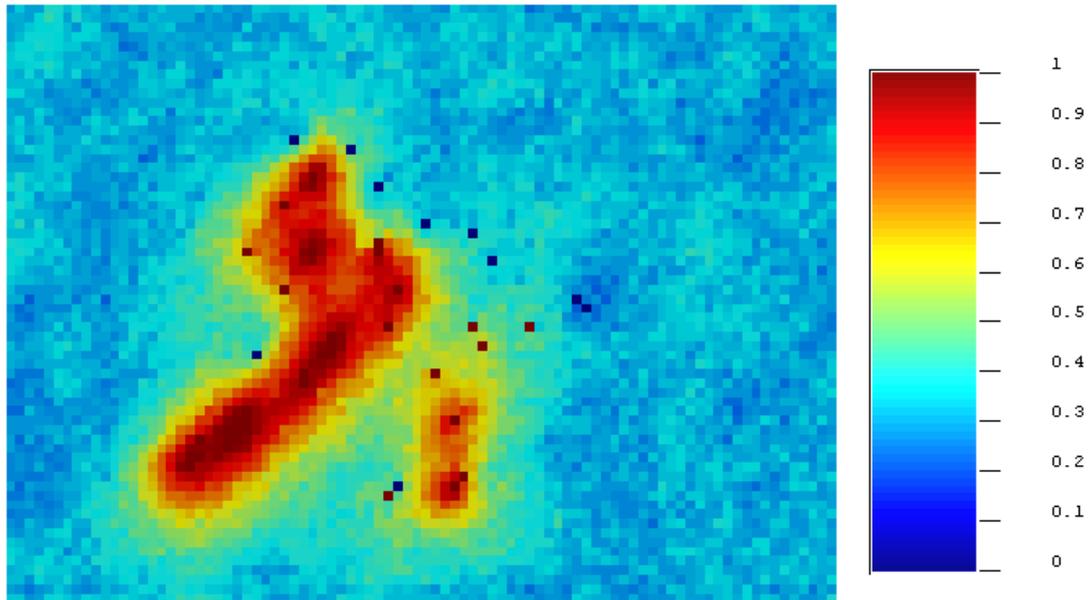


Fig.142.Ejercicio Aplicación. Probabilidad espesores estén por encima del umbral 10m. (Fuente: Elaboración propia)

En el plano nº 3 se puede observar los mapas de riesgo para distintos espesores y delimitados para el terreno disponibles. Con ellos, se ha trazado un contorno de tal manera que se cumplan los requisitos recomendados de obtener un espesor bajo el vaso del vertedero mayor a 10m con una probabilidad superior a 75%, cumpliendo este la normativa en un porcentaje en torno al 90%. Todo esto se puede observar en los planos nº 6 y nº7 respectivamente.

Cabe destacar una zona donde los porcentajes no son tan altos pero se consideran suficientes, ya que con esta zona queda todo uno (Figura 143, zona con círculo verde), simplemente necesitará una mayor cantidad de inspecciones y revisiones.

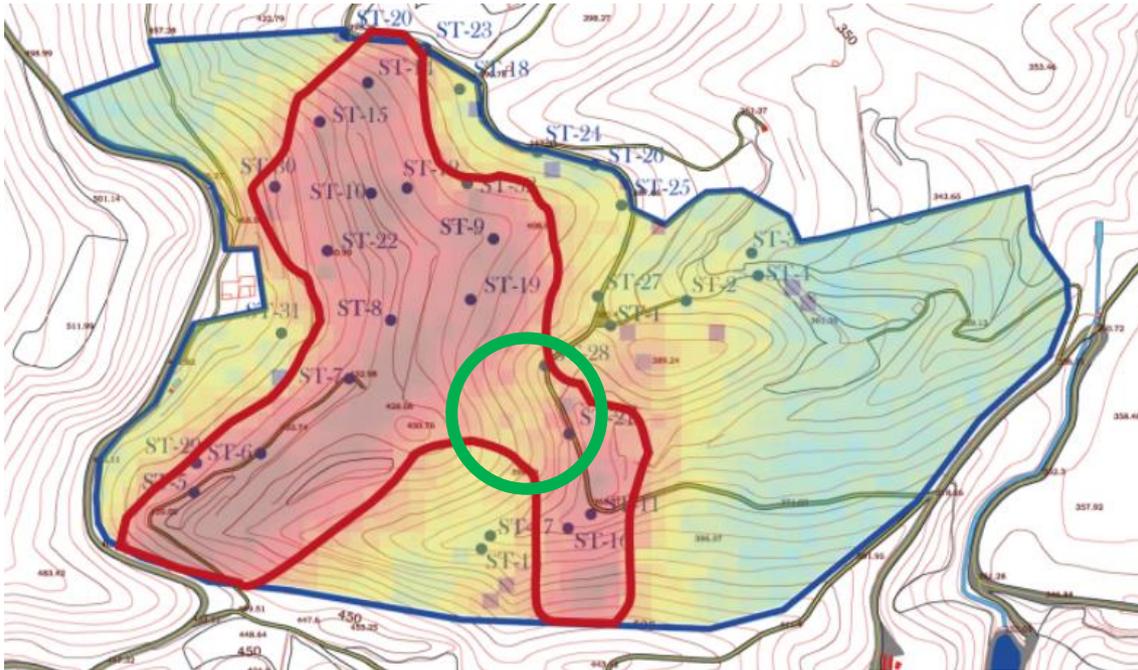


Fig.143. Propuesta delimitación de ubicación del vertedero (Sobre mapa riesgo espesor > 10m.). Zona de probabilidad menor (cículo verde). (Fuente: Elaboración propia)

A continuación, en la figura , se muestra la delimitación final establecida (Plano nº 8).

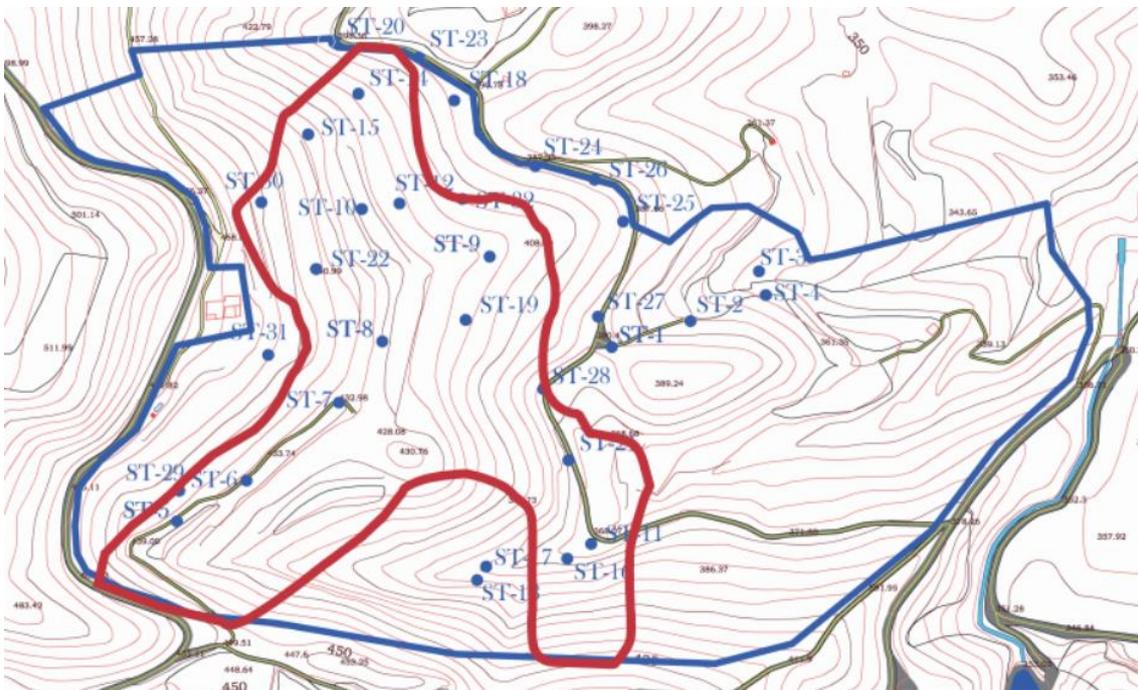


Fig.144. Propuesta final delimitación ubicación del vertedero (Fuente: Elaboración propia)

Además, comparándolo con los datos obtenidos en la estimación mediante Krigeado ordinario del apartado anterior (Figura 136-137), se puede observar que la zona obtenida tampoco se desvía mucho de las propuestas mediante estimación.

A continuación, en la figura 145, se muestran la varianza condicional de la simulación, teniendo valores relativamente bajos cerca de los puntos donde se ha realizado un sondeo y valores altos en los puntos alejados de este.

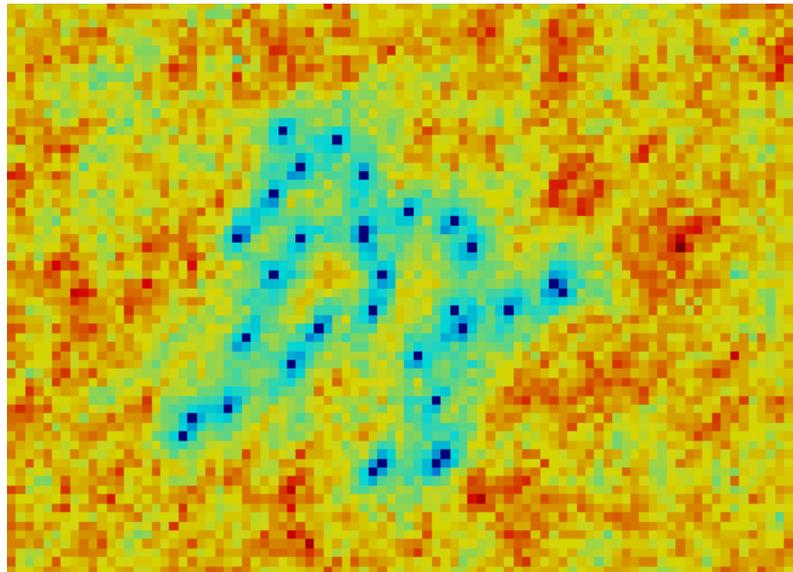


Fig.145. Ejercicio Aplicación. Varianzas condicionales. (Fuente: Elaboración propia).

Por último, se presentan los cuantiles de probabilidad 0.1, 0.25, 0.5, 0.75, 0.85, 0.99 en la Figura 146. Estos campos indican que es muy poco probable que el espesor alcance valores muy altos en zonas alejadas de los sondeos, como bien cabía esperar.

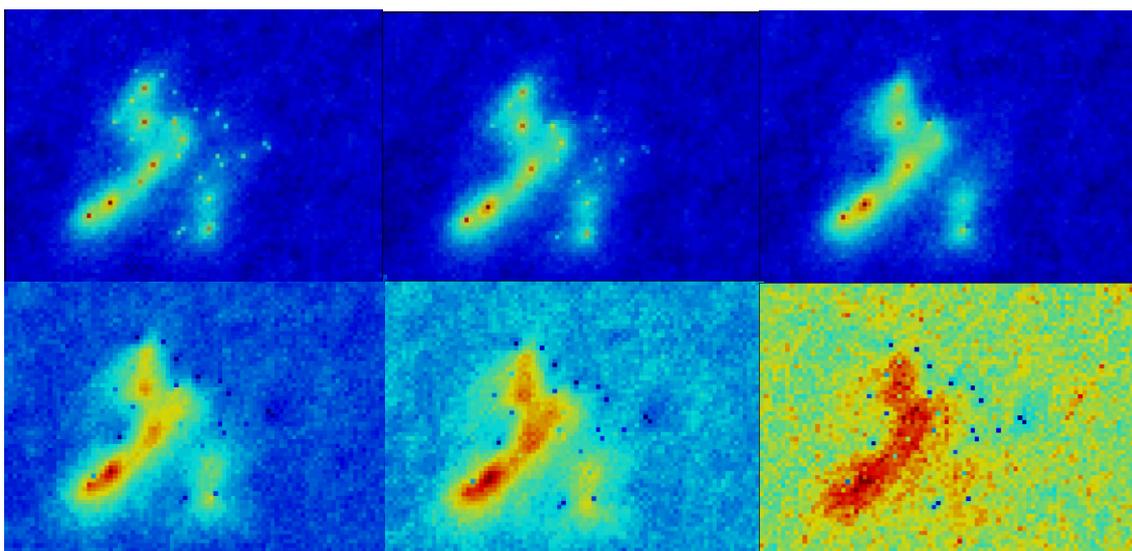


Fig. 146. Cuantiles para una probabilidad acumulada de 0.1 0.25 0.5 0.75 0.85 0.99. (Fuente: Elaboración propia)

5

Conclusiones

En el presente trabajo se ha expuesto una aplicación de técnicas geoestadísticas para la determinación de la zona apta para la implantación de un vertedero de residuos sólidos urbanos. La geoestadística se plantea, pues, como una herramienta de análisis de datos ambientales apta para la aplicación a problemas de ingeniería ambiental y, más concretamente, a la gestión de residuos sólidos.

A partir de datos de sondeos puntuales tomados en el terreno, se ha podido obtener la correlación entre los mismos y a partir de ellos, mediante la aplicación de técnicas de estimación y simulación geoestadística, se han obtenido los mapas de probabilidades correspondientes a los espesores de estratos impermeables.

La determinación de los mapas de riesgo permite delimitar el área óptima para la implantación del vaso, basándose en resultados que garanticen probabilidades relativamente altas, de forma que el vertedero no cause ningún daño ambiental debido a fugas de lixiviados de fondo de vaso.

El proceso realizado se fundamenta en disponer datos reales, ya sea para elegir la técnica óptima de estimación o simulación a utilizar como para establecer el correcto patrón de continuidad espacial, del cual dependen las distintas técnicas ya mencionadas.

Además, gracias a las herramientas informáticas de las que se disponen como el software S-Gems, dicho proceso se facilita disminuyendo además el proceso de realización. A pesar de dicha facilidad se considera complicado ajustar el modelo de continuidad espacial y la obtención de las direcciones de máxima y mínima continuidad. Los datos obtenidos pueden variar en función del grado de experiencia del técnico que lo realice y su grado de precisión.

6

Líneas de futura investigación

Como líneas de futura investigación se sugieren las siguientes:

- Ampliar el trabajo realizando un análisis geoestadístico tridimensional.
- Plantear la posible utilización de técnicas de análisis geofísico que aporten información de variables secundarias correlacionadas con la variable principal de forma que puedan utilizarse técnicas de simulación estocástica que integren la mayor cantidad de información posible, lo cual permitirá reducir la incertidumbre de los resultados obtenidos.



Referencias bibliográficas

Bohling Geoff (2007), *Introduction to Geostatistics*, Idaho

<http://people.ku.edu/~gbohling/BoiseGeostat/IntroToGeostatistics.pdf>

Bohling Geoff (2007), *S-Gems Tutorial Notes*, Idaho

[<http://people.ku.edu/~gbohling/BoiseGeostat/SGeMSTutorial.pdf>](http://people.ku.edu/~gbohling/BoiseGeostat/SGeMSTutorial.pdf)

Cassiraga, Eduardo. *Apuntes de Geoestadística – Curso 2015-2016*. Universidad Politécnica de Valencia.

Clark Isobel (2001), *Practical Geostatistics*, GeostokosLimited, Scotland.

Del pozo Manrique, Marcos (1990), *Se lección de emplazamientos para vertederos controladaso*, Acueductos Romanos en España.

Emery, Xavier (2013). *Geoestadística*. Facultad de ciencias físicas y matemáticas de la Universidad de Chile. Ingeniería de Minas.

[<https://www.ucv-cl/cursos.cl/usuario/f697544fb48474f4153ef2aafbe4c963/mi_blog/r/Geoestadistica.pdf>](https://www.ucv.cl/cursos.cl/usuario/f697544fb48474f4153ef2aafbe4c963/mi_blog/r/Geoestadistica.pdf)

Gallardo A. 2006. *Geoestadística* . Ecosistemas. 2006/3, pag.11

[<http://www.revistaecosistemas.net/articulo.asp?Id=431&Id_Categoria=1&tipo=portada>](http://www.revistaecosistemas.net/articulo.asp?Id=431&Id_Categoria=1&tipo=portada)

Gómez-Hernández, Jaime. *Apuntes S-GeMS (Stanford Geostatistical Modeling Software)*. Universitat Politècnica de Valencia

Goodvaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.

Isaaks, E y Srivastava, R (1989). *An introduction to applied Geostatistics*, Oxford University Press, New York.

Mateu Jorge (2003). *Geoestadística y modelos matemáticos en hidrogeología*, Universidad Jaume I, Castellón de la plana.

Moral García, F.J (2004), *Aplicación de la geoestadística en las ciencias ambientales*, Ecosistemas, pag:78-86.

Real Decreto 1481/2001, de 27 de diciembre, por el que se regula la eliminación de residuos mediante depósito en vertedero. BOE, 29 de enero de 2002, número 25, páginas 3507-3521.

Remy, Nicolas (2004). *Geostatistical earth Modeling Software*.

<http://sgems.sourceforge.net/old/doc/sgems_manual.pdf>

Sampere Calvete, F.Javier, Carrera Ramirez, Jesús (1990), Centro Internacional de Métodos numéricos, Barcelona.

Sanz Salinas, Jorge Gaspar (2005). *Geoestadística con R*.

< <https://es.slideshare.net/xurxosanz/geoestadistica-con-r> >

Spiegel, Murray R. y Stephens, Larry J. (2012), *Estadística*, McGraw Hill

Slideshare (2015). *Variograma*. <<http://slideplayer.es/slide/3179350/>>

Temiño, J, Adarve M.J y Rebollo L.F (1989), Metodología aplicada para la valoración de los aspectos hídricos implicados en la selección de emplazamientos de vertedero sólidos urbanos, Heneares, Rev.Geológica, pág: 37-44.



Trabajo final de master

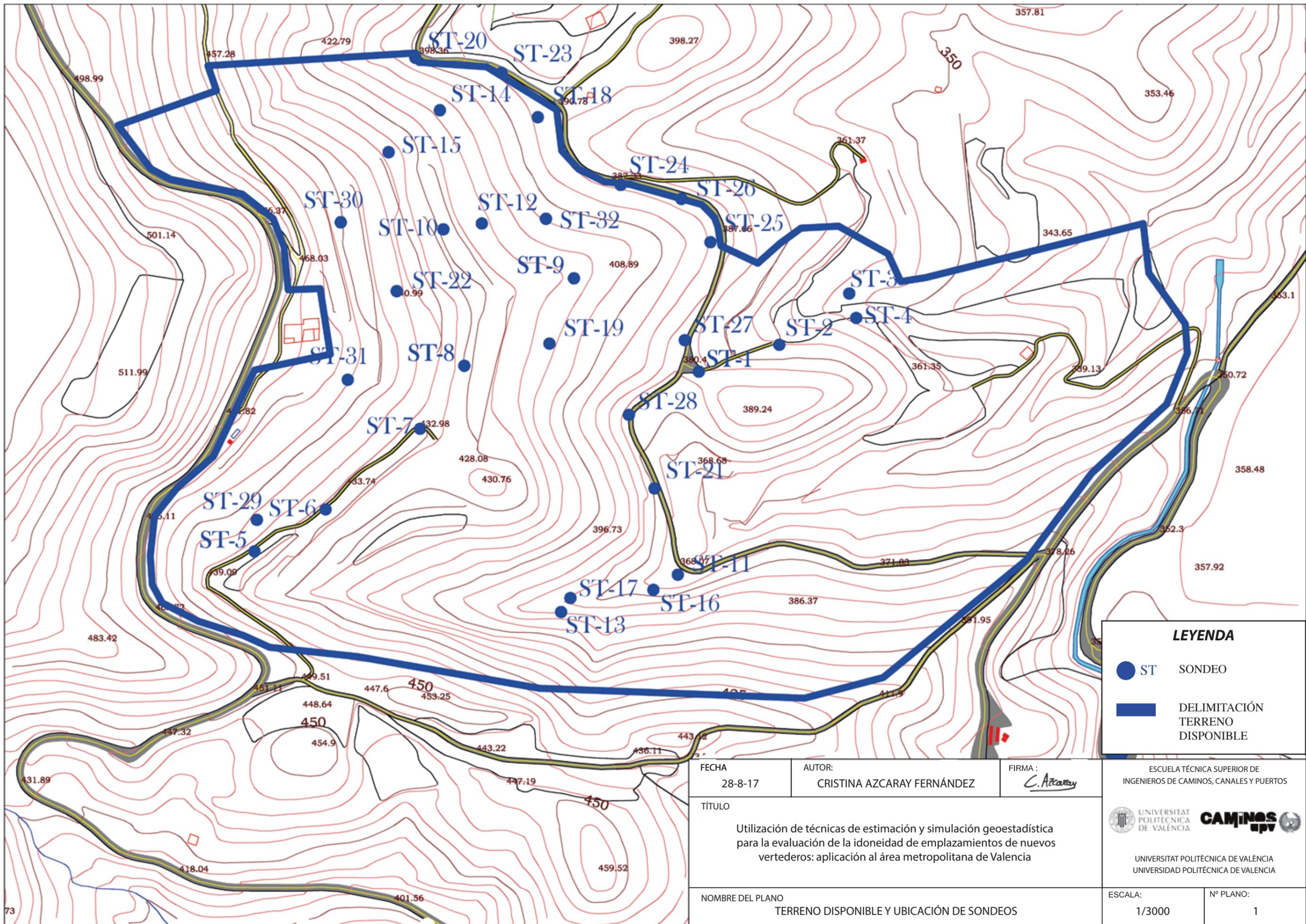
**Utilización de técnicas de estimación y simulación geoestadística para la
evaluación de la idoneidad de emplazamientos de nuevos vertederos:
aplicación al área metropolitana de Valencia**

ANEJO 1

PLANOS TÉCNICOS

ÍNDICE

- Plano 1: Terreno disponible y ubicación de sondeos
- Plano 2: Ejemplo de simulación de espesores
- Plano 3: Mapas de riesgos simulados (Simulación Gausiana Secuencial)
- Plano 4: Mapas de riesgo estimados (Krigado ordinario)
- Plano 5: Mapas de riesgo estimados (Krigado ordinario)
- Plano 6: Propuesta delimitación de ubicación del vertedero (sobre plano e> 1m.)
- Plano 7: Propuesta delimitación de ubicación del vertedero (sobre plano e> 10m.)
- Plano 8: Propuesta final delimitación de ubicación del vertedero.



LEYENDA

- ST SONDEO
- DELIMITACIÓN TERRENO DISPONIBLE

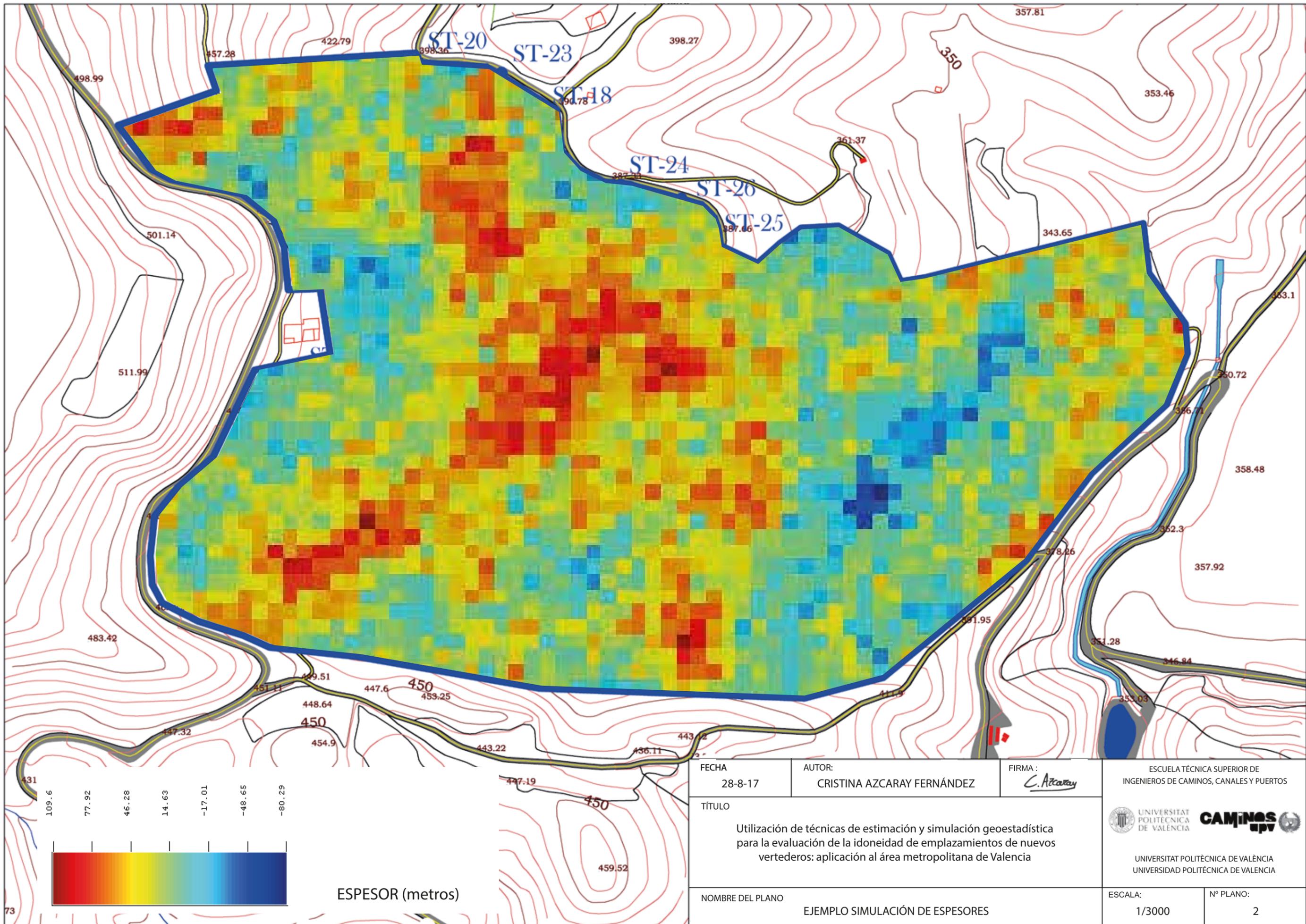
FECHA	AUTOR:	FIRMA:
28-8-17	CRISTINA AZCARAY FERNÁNDEZ	<i>C. Azcaray</i>
TÍTULO		
Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia		
NOMBRE DEL PLANO		
TERRENO DISPONIBLE Y UBICACIÓN DE SONDEOS		

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS

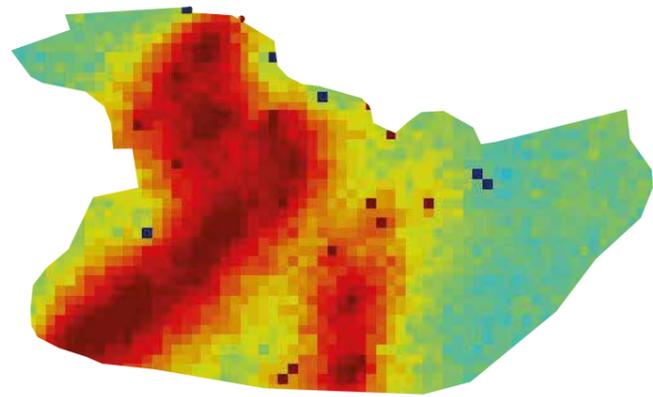
UNIVERSITAT POLITÈCNICA DE VALÈNCIA **CAMINOS UPV**

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
UNIVERSIDAD POLITÉCNICA DE VALENCIA

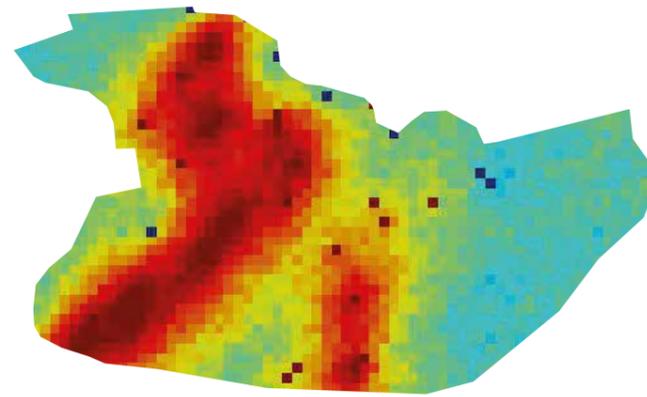
ESCALA:	Nº PLANO:
1/3000	1



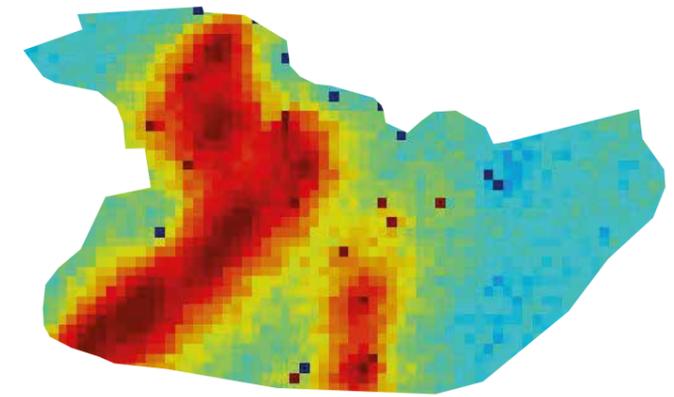
FECHA 28-8-17	AUTOR: CRISTINA AZCARAY FERNÁNDEZ	FIRMA: <i>C. Azcaray</i>	ESCUOLA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
TÍTULO Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia			 UNIVERSITAT POLITÈCNICA DE VALÈNCIA UNIVERSIDAD POLITÉCNICA DE VALENCIA
NOMBRE DEL PLANO EJEMPLO SIMULACIÓN DE ESPESORES		ESCALA: 1/3000	Nº PLANO: 2



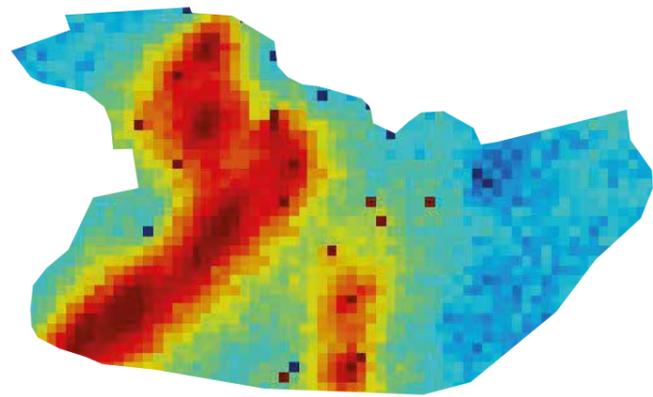
ESPESOR > 1 m.



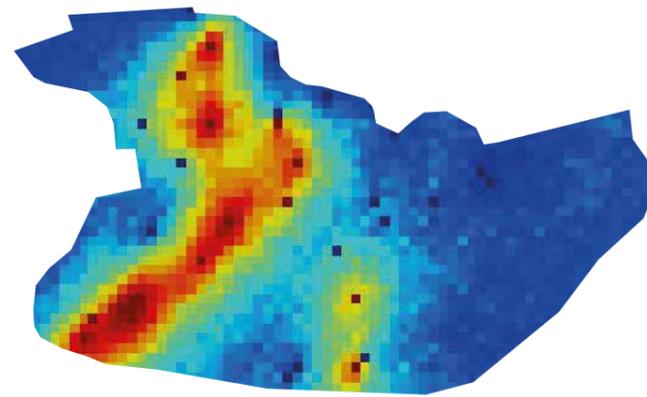
ESPESOR > 4 m.



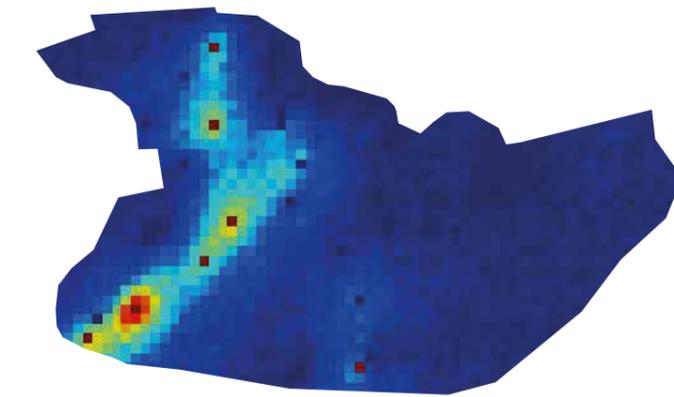
ESPESOR > 7 m.



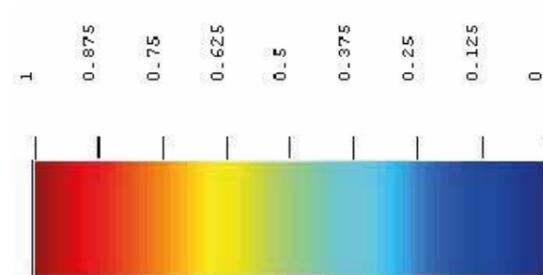
ESPESOR > 10 m.



ESPESOR > 20 m.

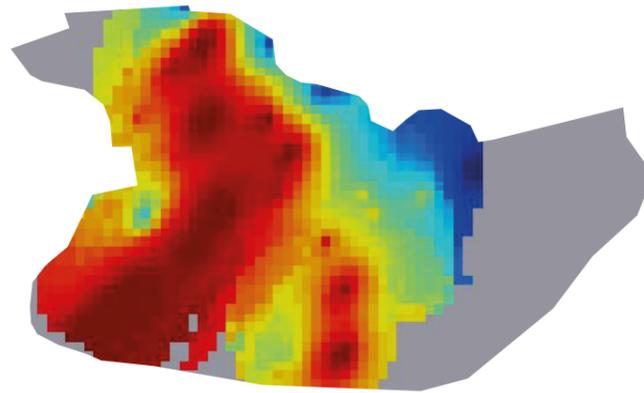


ESPESOR > 40 m.

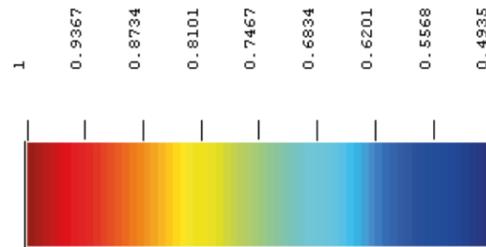


PROBABILIDAD (de 0 a 1)

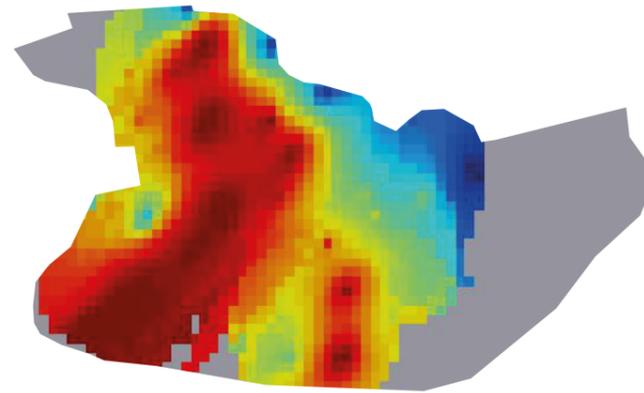
FECHA 28-8-17	AUTOR: CRISTINA AZCARAY FERNÁNDEZ	FIRMA: <i>C. Azcaray</i>	ESCUOLA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
TÍTULO Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia			 UNIVERSITAT POLITÈCNICA DE VALÈNCIA UNIVERSIDAD POLITÉCNICA DE VALENCIA
NOMBRE DEL PLANO MAPAS DE RIESGO SIMULADOS (SIMULACIÓN GAUSIANA SECUENCIAL)			ESCALA: S/E
			Nº PLANO: 3



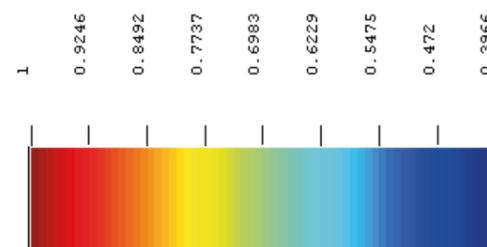
ESPESOR > 1 m.



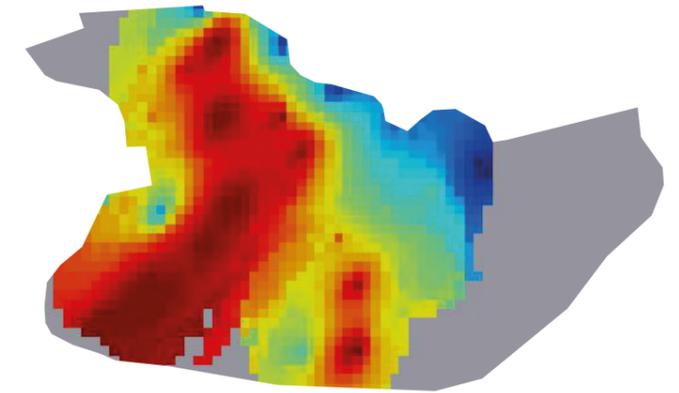
PROBABILIDAD (de 0 a 1)



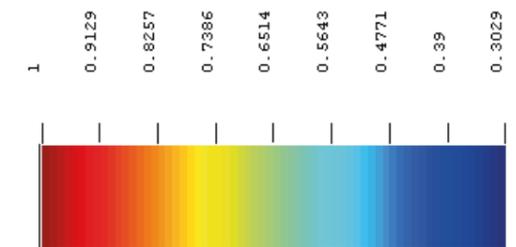
ESPESOR > 4 m.



PROBABILIDAD (de 0 a 1)

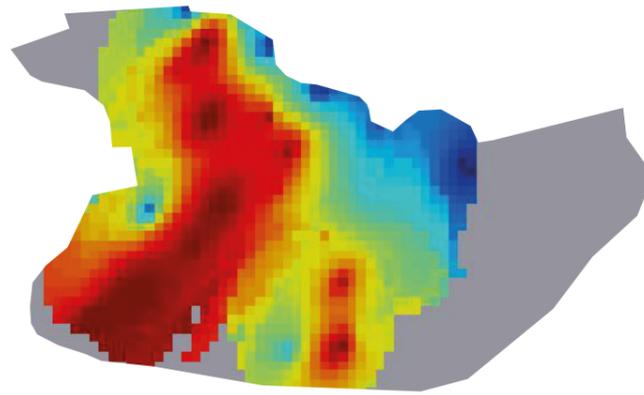


ESPESOR > 7 m.



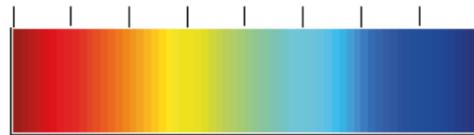
PROBABILIDAD (de 0 a 1)

FECHA 28-8-17	AUTOR: CRISTINA AZCARAY FERNÁNDEZ	FIRMA: <i>C. Azcaray</i>	ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
TÍTULO Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia			 UNIVERSITAT POLITÈCNICA DE VALÈNCIA UNIVERSIDAD POLITÉCNICA DE VALENCIA
NOMBRE DEL PLANO MAPAS DE RIESGO ESTIMADOS (KRIGEADO ORDINARIO)			ESCALA: S/E
			Nº PLANO: 4

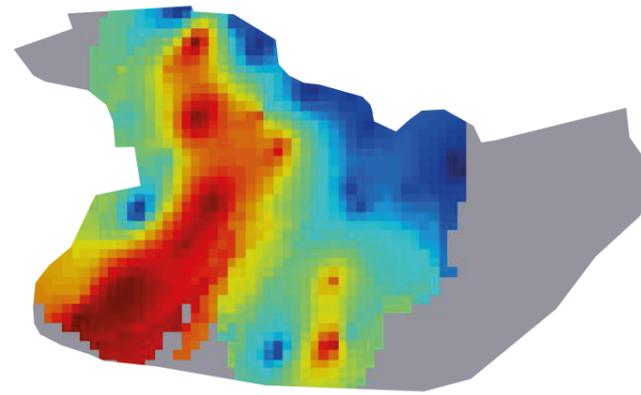


ESPESOR > 10 m.

1
0.9026
0.8051
0.7077
0.6103
0.5128
0.4154
0.318
0.2205

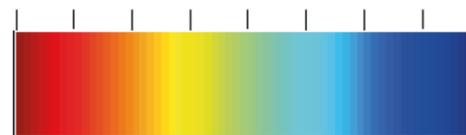


PROBABILIDAD (de 0 a 1)

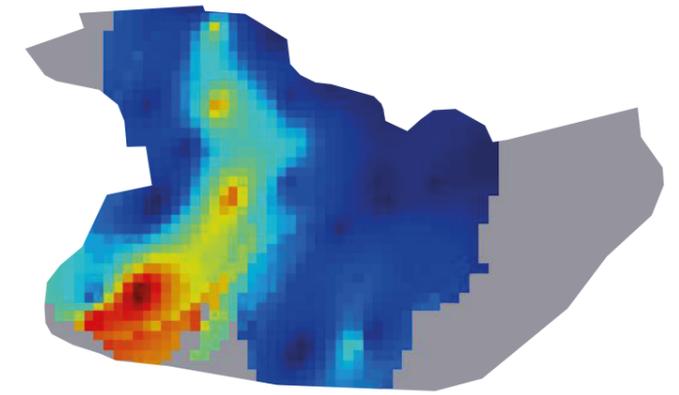


ESPESOR > 20 m.

0.9997
0.8814
0.763
0.6447
0.5263
0.4079
0.2896
0.1712
0.05287

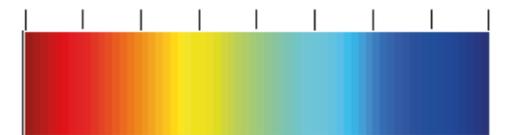


PROBABILIDAD (de 0 a 1)



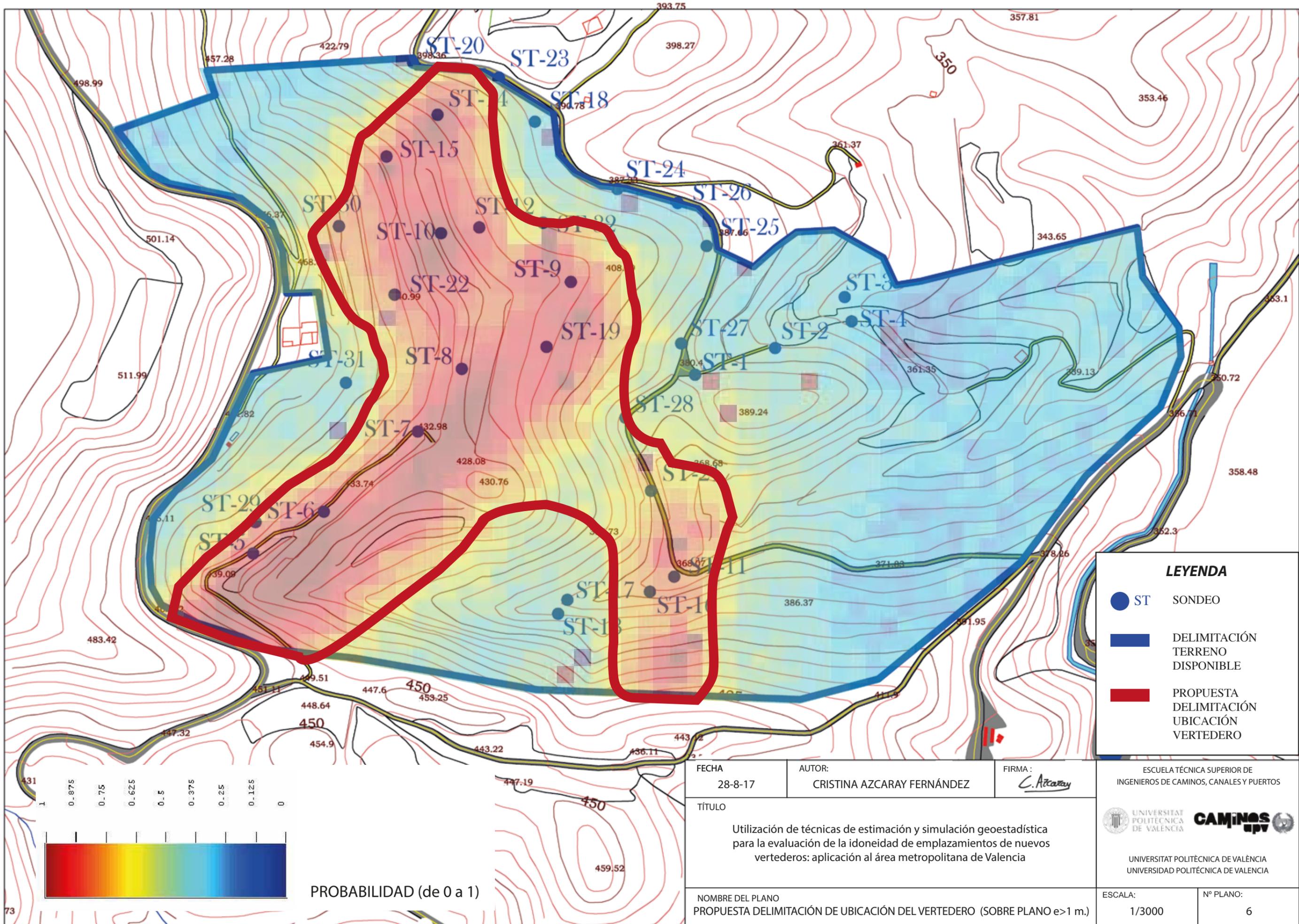
ESPESOR > 40 m.

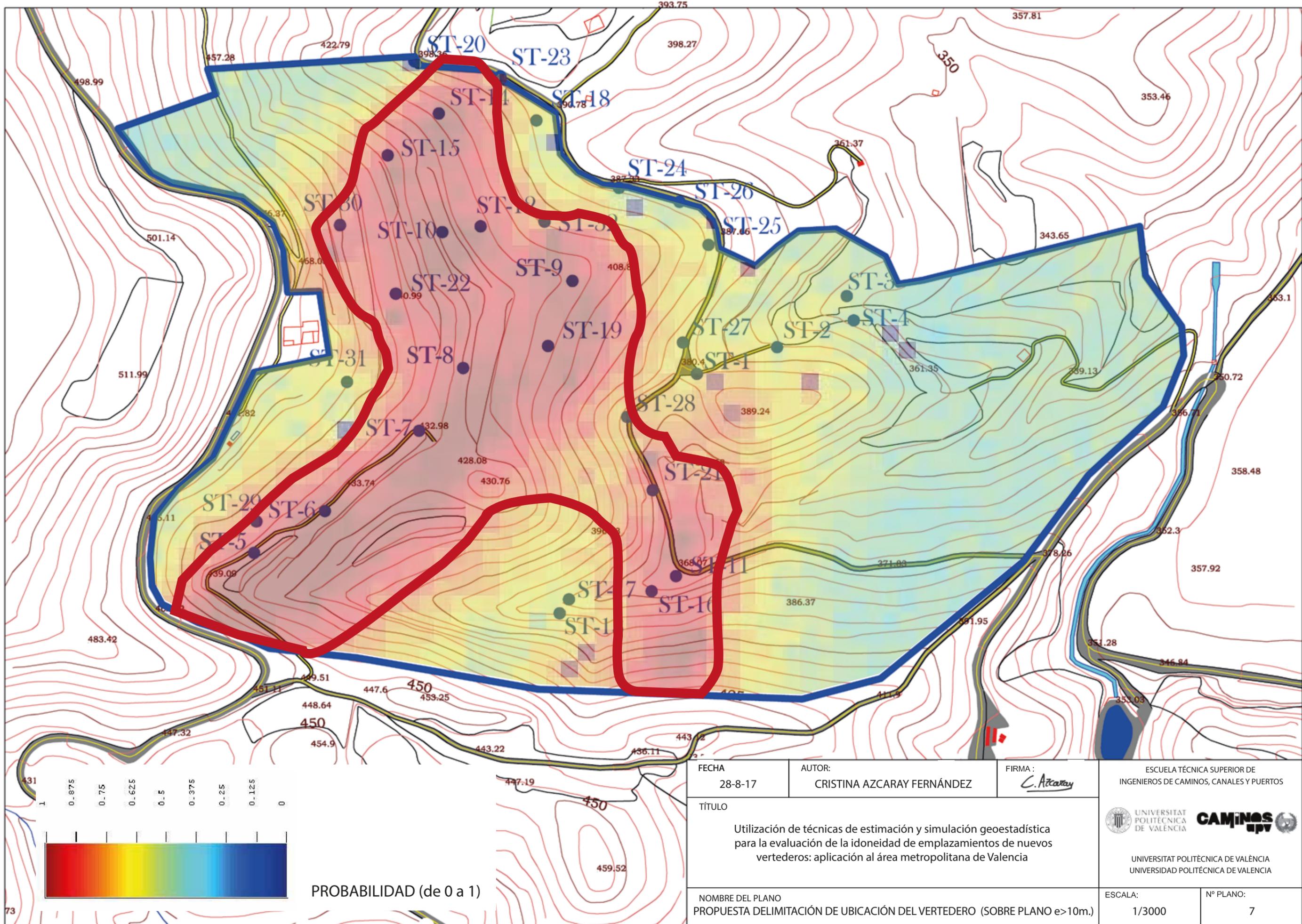
0.9743
0.8526
0.7309
0.6091
0.4874
0.3657
0.2439
0.1222
0.000465



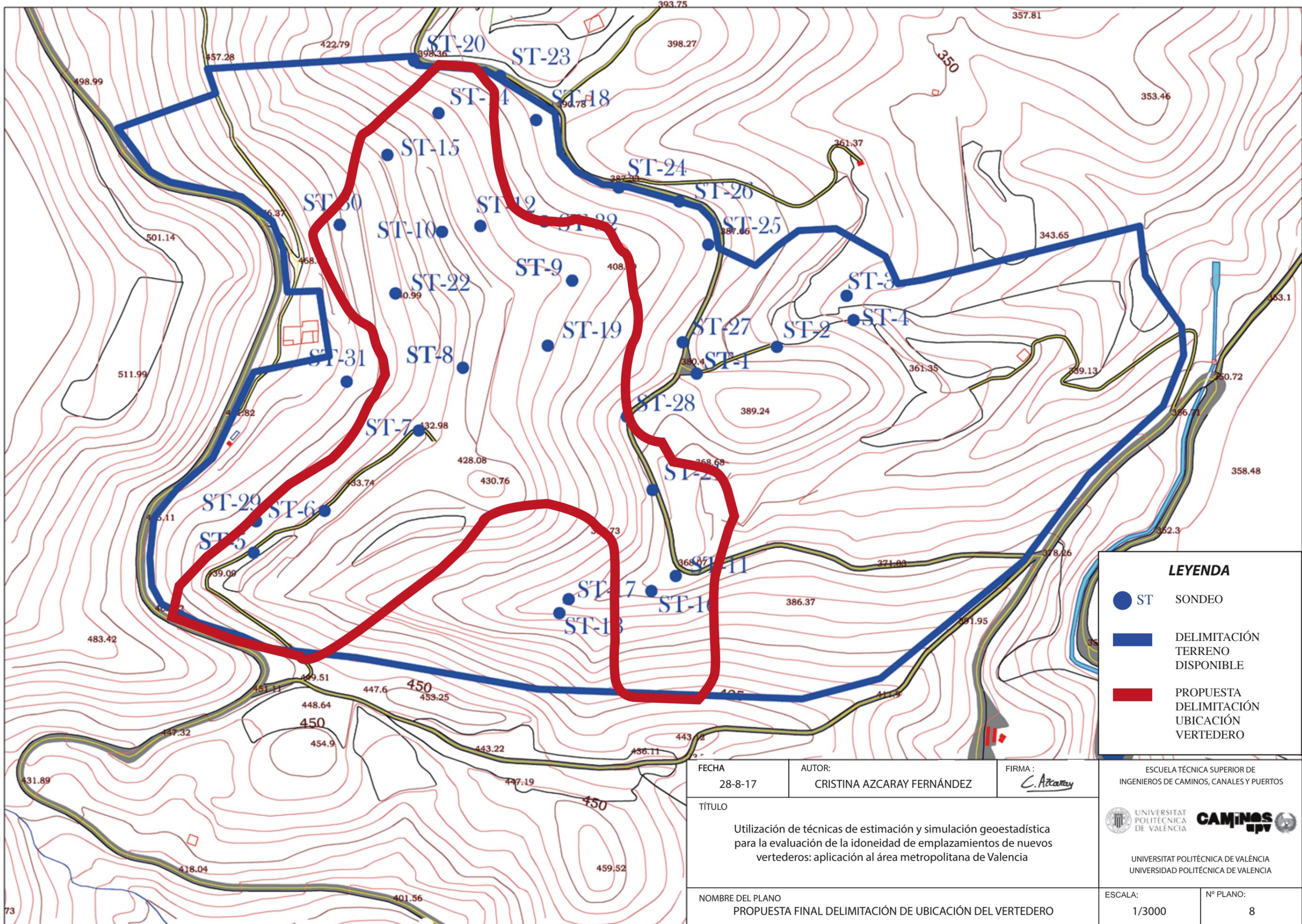
PROBABILIDAD (de 0 a 1)

FECHA 28-8-17	AUTOR: CRISTINA AZCARAY FERNÁNDEZ	FIRMA: <i>C. Azcaray</i>	ESCUOLA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
TÍTULO Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia			 UNIVERSITAT POLITÈCNICA DE VALÈNCIA UNIVERSIDAD POLITÉCNICA DE VALENCIA
NOMBRE DEL PLANO MAPAS DE RIESGO ESTIMADOS (KRIGEADO ORDINARIO)			ESCALA: S/E
			Nº PLANO: 5





FECHA 28-8-17	AUTOR: CRISTINA AZCARAY FERNÁNDEZ	FIRMA: <i>C. Azcaray</i>	ESCUOLA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
TÍTULO Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia			
NOMBRE DEL PLANO PROPUESTA DELIMITACIÓN DE UBICACIÓN DEL VERTEDERO (SOBRE PLANO e>10m.)			ESCALA: 1/3000 Nº PLANO: 7



LEYENDA	
● ST	SONDEO
—	DELIMITACIÓN TERRENO DISPONIBLE
—	PROPUESTA DELIMITACIÓN UBICACIÓN VERTEDERO

FECHA	AUTOR:	FIRMA:
28-8-17	CRISTINA AZCARAY FERNÁNDEZ	<i>C. Azcaray</i>
TÍTULO		
Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia		
NOMBRE DEL PLANO		
PROPUESTA FINAL DELIMITACIÓN DE UBICACIÓN DEL VERTEDERO		

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS	
UNIVERSITAT POLITÈCNICA DE VALÈNCIA UNIVERSIDAD POLITÉCNICA DE VALENCIA	
ESCALA:	Nº PLANO:
1/3000	8



Trabajo final de master

Utilización de técnicas de estimación y simulación geoestadística para la evaluación de la idoneidad de emplazamientos de nuevos vertederos: aplicación al área metropolitana de Valencia

Firmado digital por Cristina Azcaray Fernández

D.N.I. 78937486Z

Valencia, septiembre de 2017

TITULACIÓN: Master en Ingeniería de Caminos, Canales y Puertos

Curso: 2016/2017

AUTORA: Azcaray Fernández, Cristina

TUTOR: Rodrigo-Ilarri, Javier

COTUTORA: Rodrigo-Clavero, M^a Elena