

# Aplicación del análisis geostatístico al diseño de programas de muestreo en formaciones salinas

## Use of geostatistical analysis to design sampling programs in salt formations

C. DE LAS CUEVAS

Laboratori d'Investigació en Formacions Salines. Dept. de Geoquímica, Petrologia i Prospecció Geològica.  
Facultat de Geologia. Universitat de Barcelona. 08028 Barcelona

### RESUMEN

El contenido de agua en las formaciones salinas no es una variable aleatoria sino regionalizada, es decir que el valor de la variable en un punto condiciona el valor de ésta en puntos adyacentes. En este trabajo se plantea el estudio de la minimización de las incertidumbres del contenido en agua generadas por la variabilidad espacial y el limitado número de datos, a fin de conocer cuál es el intervalo óptimo de muestreo para el estudio sistemático y representativo de una formación salina. El análisis geostatístico se ha realizado a partir de muestreos tanto a pequeña escala como a gran escala con muestras procedentes de diversas formaciones salinas. El estudio ha mostrado que el análisis geostatístico es justificable cuando las dimensiones del cuerpo geológico a evaluar no excede en más de un orden de magnitud el alcance de la variable regionalizada. En caso contrario, una simple aproximación estadística puede ofrecer un nivel satisfactorio de precisión en la determinación del valor promedio de la variable.

*Palabras clave:* Formación salina. Geoestadística. Semivariograma. Varianza de extensión. Optimización del muestreo.

### ABSTRACT

Water content in salt formations is a regionalised variable rather than a random variable. This means that the yield of the variable in one location varies very slightly from that of close locations. This paper focuses on the minimisation of uncertainties of the water content caused by spatial variability and limited number of samples, in order to assess optimised sampling intervals for representative sampling procedures in salt formations. Geostatistical analyses have been performed using small-scale and large-scale sampling intervals in several salt formations. Results show that geostatistical analyses can be

successfully used, provided that the dimensions of the studied rockbody do not exceed the range of the regionalised variable by more than one order of magnitude. Otherwise, a simple statistical approximation should provide the degree of accuracy needed when estimating sample means.

*Keywords:* Salt Formation. Geostatistics. Semivariogram. Extension variance. Optimised sampling strategies.

### EXTENDED ABSTRACT

A question that frequently arises when numerical data in Geology are used, is to know their degree of uncertainty due to spatial variability and limited amount of data. In this sense Geostatistics offers a robust estimation technique since it combines the advantages of statistical analysis with those of contouring using gridding methods. Furthermore, Geostatistics can be used to improve sampling designs. This paper describes the application of Geostatistics for establishing optimised sampling strategies in rock salt formations. Spatial variability in rock salt has been assessed, using small-scale and large-scale sampling intervals in several salt formations.

Sampling was performed in the Lower Salt Unit of the Cardona and Guendulain Formations, as well as in the salt horizons from the Zaragoza Formation. The variable taken into account was the water content of rock salt.

Sampling took place following whether an horizontal direction (Cardona Fm.) or a vertical direction (Zaragoza and Guendulain Fms.). Sampling interval varied from few centimetres (Cardona and Zaragoza Fms.) to few metres (Guendulain Fm.). Statistical and geostatistical calculations were performed using the GEO-EAS package. The main statistical parameters of the studied formations are displayed in Table 3. Spatial variability has been assessed from detailed study of the semivariograms, which were afterwards fitted to a theoretical model. The bias of the population expressed as the ratio between random variance and spatial component of variance (pitch) was also computed.

Regarding the spatial variability of the Lower Salt Unit of the Cardona Formation, Figure 1 shows its experimental and fitted to a spherical model semivariograms. Best fits were obtained using a nugget of 0.018 %<sup>2</sup>, pitch of 0.023 %<sup>2</sup>, sill value of 0.041%<sup>2</sup> and a range of 1.6 m. Bias of the population was estimated to be of 78%. After this fitting procedure, the extension variance has been calculated for a relationship  $l/a$  which equalled 0.04. By this procedure the standard deviation decreased from 0.20 % (obtained by Statistics) to 0.13 %.

Figure 2 shows the experimental and fitted to a spherical model semivariograms for samples from the salt horizons of the Zaragoza Formation. The fitted parameter were: nugget of 0.006 %<sup>2</sup>, pitch of 0.010 %<sup>2</sup>, sill value of 0.016 %<sup>2</sup> and a range of 1.0 m. Bias of the population was 60 %<sup>2</sup>. The standard deviation decreased from 0.11 % (Statistics) to 0.079 % (after calculation of the extension variance for a relationship  $l/a$  equal to 0.1).

Figure 3 shows the experimental semivariogram of the Guendulain Formation, where random fluctuations can be observed. At this level, the bias of the population tends to infinite, being impossible to determine its extension variance. In this case estimation errors have to be assessed using conventional statistics. The evolution of the standard deviations after using several sampling procedures, obtained for the aforementioned formations, are summarised in Table 4.

In order to design a sampling program for salt formations, a virtual composite formation has been chosen. It has been assumed that the semivariogram is independent of the number and the location of the samples. Base Case parameters were: average of 0.30%, total variance of 0.023%<sup>2</sup>, nugget of 0.010%<sup>2</sup>, pitch of 0.013%<sup>2</sup> and range of 1.0 m. Furthermore, sampling intervals 0.5m, 1m, 2m y 5m have been selected. The

results are shown in Figure 4. From this figure it can be observed that the error obtained with sampling intervals between 0.5 and 2.0 m are very close, due to the large bias of the studied variable.

The estimation error in the grade of the variable has also been studied for larger rockbodies. Two cases have been considered: a rockbody, whose dimensions were 50x50x30m and a second of 10x10x4m size. Best effectiveness would be obtained in the first case using a sampling interval of 8.5 m ( $n=100$ ; error of 9.6%), and in the second case using an interval of 1.5 m ( $n=70$ ; error of 9.7%).

As concluding remarks it should be considered that semivariograms of the water content in rock salt follow the spherical model. Range of the variable is close to 1 m, and bias varies between 60 % and 80%. Furthermore, Geostatistics can be successfully used, provided that the dimensions of the studied rockbody does not exceed in one order of magnitude for the three orthogonal directions in space the range of the variable. If this is not the case, a simple statistical approximation should provide the degree of accuracy needed when estimating sample means.

## INTRODUCCIÓN

Una cuestión fundamental cuando se trabaja con datos numéricos en materiales geológicos es conocer su grado de incertidumbre, causada por la variabilidad espacial y el limitado número de datos. En general, el valor de cualquier variable se determina puntualmente y se suele extrapolar a una estructura geológica de mayores dimensiones. A partir de un número finito de determinaciones puntuales, distribuidas en el espacio, se suele asignar valores de la variable en el conjunto de los puntos intermedios donde se carece de medidas. Tradicionalmente se han utilizado métodos de interpolación (utilizando como factor de ponderación la inversa del cuadrado de la distancia) o mediante poligonalización. El uso de dichas técnicas de interpolación puede dar una idea de la variabilidad a nivel espacial en una estructura geológica, pero no permiten estimar la eficiencia de la estimación. Generalmente para la estimación de cualquier variable se utiliza el análisis estadístico a partir de un muestreo representativo de la población a estudiar. Existen diversas técnicas de muestreo, como tomar muestras al azar o bien realizar un muestreo sistemático a intervalos regulares. Cuando el tamaño de la muestra es grande ( $n>30$ ) las muestras

suelen seguir una distribución normal o su variante lognormal (Rock, 1988). Mediante inferencia estadística pueden llegar a conocerse los parámetros poblacionales (media poblacional, varianza), así como el grado de exactitud de éstos. La eficiencia en la estimación depende en gran manera del tamaño de la muestra quedando, no obstante, una cierta cantidad de error residual, condicionado por el grado de variabilidad intrínseco al medio geológico.

La Geoestadística es el estudio de datos que muestran una determinada correlación en el espacio (Davis, 1986). El análisis geoestadístico tiene la ventaja que permite acoplar las ventajas de la estadística y las de los métodos de interpolación (Rock, 1988), ya que permite establecer el error asociado a la estimación, así como determinar la serie de coeficientes ponderadores para los métodos de extrapolación. Otra cuestión importante, referida al estudio de las variables geológicas, es que no son aleatorias sino regionalizadas. Una variable regionalizada presenta una determinada distribución en el espacio, es decir que el valor de la variable en un punto condiciona el valor de ésta en puntos adyacentes. Estas variables son continuas en el espacio, y presentan un componente aleatorio (causada por las irregularidades locales) y un componente espacial (debido a la tendencia general del fenómeno estudiado). Dado que la Geoestadística ha generado a lo largo de los últimos 20 años una extensa bibliografía, se remite al lector interesado en este tema a los libros de Chica-Olmo (1988), David (1977) o Davis (1986). Asimismo algunos de los conceptos básicos de la Geoestadística están descritos en el Anexo de este trabajo.

La herramienta básica para el estudio de las variables regionalizadas es el semivariograma, que es una función matemática que expresa la evolución de la semivarianza de los incrementos de la variable regionalizada (según una dirección determinada) en función de la distancia entre muestras. A partir del semivariograma se puede calcular el error que se comete cuando se asigna el valor promedio de una serie de valores puntuales a un dominio mayor. Dicho cálculo, que recibe el nombre de varianza de estimación, permite la optimización de las técnicas de muestreo (Chica-Olmo, 1988). A diferencia de la Estadística clásica, en la que la varianza total se utiliza para la estimación del error en la determinación de la media de la población para un nivel de confianza determinado, mediante la Geoestadística dicha varianza se descompone en su componente aleatorio (efecto pepita) y en su componente espacial. Utilizando intervalos de muestreo inferiores o próximos al alcance de la

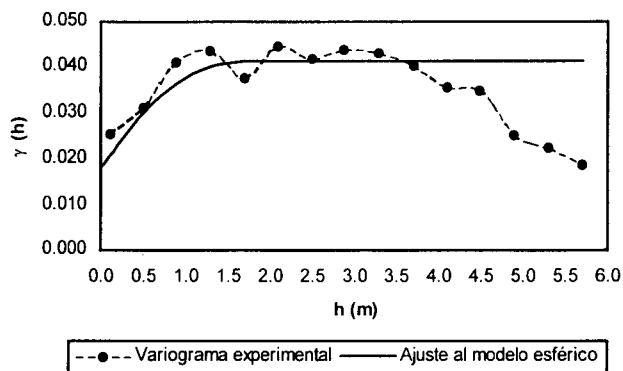


Figura 1. Semivariogramas experimental y ajustado al modelo esférico del contenido en agua de la "Sal de Muro" de la Fm. Salina de Cardona. En el semivariograma experimental, la línea discontinua representa la continuidad de la variable regionalizada determinada a partir de 15 puntos en los que se ha calculado el valor de  $u(h)$ .

Figure 1. Experimental and fitted (spherical model) semivariograms for the water content of the Lower Salt Unit of the Cardona Formation. In the experimental semivariogram, the dotted line shows the regionalized variable which has been extrapolated after computing the value of  $g(h)$  for 15 points.

variable puede reducirse la incertidumbre en el valor real promedio de la variable en un cuerpo geológico.

En este trabajo se plantea el uso del análisis geoestadístico para minimizar las incertidumbres generadas por la variabilidad espacial y el limitado número de datos, a fin de conocer cuál es el intervalo óptimo de muestreo para el estudio sistemático y representativo de una formación salina. El análisis geoestadístico se ha realizado a partir de muestreos tanto a pequeña escala como a gran escala con muestras procedentes de diversas formaciones salinas.

## METODOLOGÍA

La información numérica que se ha utilizado procede de muestras obtenidas en la "sal de muro" de las Fm. Salina de Cardona y de la Fm. Guendulain, así como de los niveles halíticos de la Fm. Zaragoza. La caracterización petrográfica, mineralógica y de contenido en agua, así como sus respectivas metodologías analíticas han sido descritas anteriormente (de las Cuevas y Pueyo, 1991; de las Cuevas, 1992; de las Cuevas y Pueyo, 1995). La variable que se ha decidido estudiar ha sido el contenido de agua en las formaciones salinas. El muestreo ha sido unidimensional, tanto en perfiles horizontales (Fm. Salina de Cardona) como en perfiles verticales (Fm. Zaragoza y Fm. Guendulain). El

intervalo de muestreo ha sido de orden centimétrico (Fm. Salina de Cardona y Fm. Zaragoza) o bien de orden métrico (Fm. Guendulain). Los valores de las medidas, así como sus coordenadas, se encuentran en las tablas 1 y 2.

El tratamiento numérico de los datos analíticos se ha realizado dividiendo las muestras en poblaciones según su pertenencia a las distintas formaciones anteriormente mencionadas. Los cálculos se han realizado mediante la utilización del programa GEO-EAS (EPA, 1988). Dicho programa permite la determinación de los estadísticos muestrales (media, varianza, etc.), así como la realización de cálculos geoestadísticos (semivariogramas, kriging). A partir de la obtención de los estadísticos muestrales se ha calculado el intervalo en el que se encuentra la media de cada población  $m$  (para un nivel de confianza del 95 %) mediante la siguiente fórmula:  $x - 1.96s/\sqrt{n} < \mu < x + 1.96s/\sqrt{n}$ , siendo  $x$  la media muestral,  $s$  la desviación standard muestral y  $n$  el tamaño de la población muestral (Spiegel, 1985).

En lo que respecta a los cálculos geoestadísticos, los semivariogramas experimentales han sido ajustados a un modelo teórico. A partir de la información obtenida en el ajuste, se ha determinado el grado de sesgo de la formación, expresado como la relación entre el componente aleatorio de la varianza y el componente espacial de la varianza (Royle, 1979).

A fin de simplificar la optimización de programas de muestreo se ha considerado que las formaciones salinas estudiadas son isotrópicas en las tres direcciones espaciales, y que el semivariograma en cualquier sector de la formación es similar a los obtenidos experimentalmente.

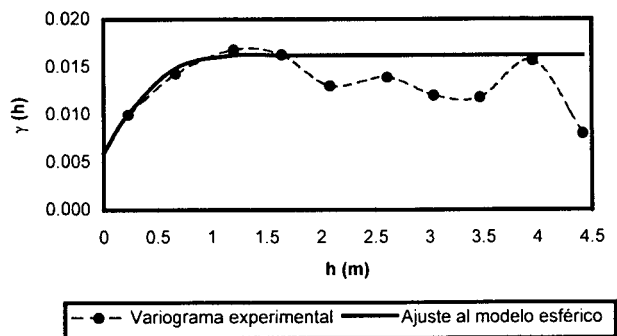


Figura 2. Semivariogramas experimental y ajustado al modelo esférico del contenido en agua en los niveles halíticos de la Fm. Zaragoza

Figure 2. Experimental and fitted (spherical model) semivariograms for the water content of the halitic horizons of the Zaragoza Formation.

## RESULTADOS Y DISCUSIÓN

En lo que respecta al contenido en agua de las rocas salinas, los estadísticos básicos para cada población se muestran en la tabla 3. Para la “Sal de Muro” de la Fm. Salina de Cardona, el rango de contenido en agua varía entre 0,04 y 1,00 % en peso sobre un total de 175 muestras. La distribución es lognormal, estando localizadas las clases modales entre 0,10 y 0,30 %, siendo el contenido medio de 0,32 % y la desviación standard de 0,20 %. Se ha estimado una media poblacional para la Sal de Muro, que oscila entre 0,29 y 0,35 % para un nivel de confianza del 95 %. En la halita de la Fm. Zaragoza, el rango del contenido en agua sobre un total de 34 muestras oscila entre 0,07 y 0,47 %. La distribución es normal, presentando clases modales entre 0,10 y 0,30 %, contenido medio de 0,23 % y una desviación standard de 0,11 %. La media poblacional de la formación oscila entre 0,19 y 0,27 %, para un nivel de confianza del 95 %. Finalmente, en la “Sal de Muro” de la Fm. Guendulain el rango de contenido en agua varía entre 0,01 y 1,38 % para un total de 25 muestras. La distribución es lognormal, presentando clases modales entre 0,00 y 0,20 %, contenido medio de 0,40 % y una desviación standard de 0,32 %. La media poblacional del contenido de agua de la formación oscila entre 0,28 y 0,52 % para un nivel de confianza es del 95 %.

El estudio de la variabilidad espacial en las formaciones estudiadas se ha realizado a partir del estudio de los semivariogramas realizados en la dirección del muestreo. En el caso de la “Sal de muro” de la Fm. Salina de Cardona, se ha realizado un semivariograma compuesto a partir de la información suministrada por 4 sondeos. El número de muestras estudiadas ha sido 175, y el intervalo de muestreo de 7 cm.

La figura 1 muestra los semivariogramas experimental y ajustado obtenidos para la “Sal de Muro” de la Fm. Salina de Cardona. En el semivariograma experimental, la varianza total aumenta hasta alcanzar una meseta en torno a los 2 m. Posteriormente va disminuyendo gradualmente, observándose, no obstante, pequeñas fluctuaciones en el contenido en agua atribuibles debido a efectos de borde. El ajuste del semivariograma a un modelo esférico presenta los siguientes parámetros: efecto pepita de 0,018 %<sup>2</sup>, componente espacial de la varianza de 0,023 %<sup>2</sup>, valor de la meseta de 0,041%<sup>2</sup> y un alcance de 1,6 m. El grado de sesgo de la formación, expresado como la relación entre el componente aleatorio de la varianza y el componente espacial de la varianza, es del 78%.

Tabla 1. Localización espacial y contenido en agua (expresado en %) de las muestras procedentes de la "Sal de Muro" de la Fm. Cardona.

Table 1. Spatial location and water content (expressed in %) of the samples from the Lower Salt Unit of the Cardona Formation.

Sal de Muro de la Fm. Salina de Cardona							
Sondeo 1		Sondeo 2		Sondeo 3		Sondeo 4	
x (m)	H2O (%)	x (m)	H2O (%)	x (m)	H2O (%)	x (m)	H2O (%)
0.30	0.17	0.35	0.14	0.06	0.37	0.39	0.29
0.40	0.31	0.41	0.33	0.15	0.36	0.46	0.42
0.45	0.29	0.62	0.15	0.42	0.11	0.55	0.33
0.55	0.29	0.69	0.15	0.50	0.35	0.62	0.76
0.78	0.24	0.76	0.20	0.70	0.06	0.68	0.33
0.85	0.24	0.83	0.07	0.79	0.18	0.74	0.22
0.95	0.31	1.40	0.13	0.86	0.18	0.78	0.33
1.04	0.32	1.45	0.21	1.36	0.20	0.83	0.53
1.55	0.56	1.67	0.22	1.45	0.24	0.89	0.47
1.60	0.38	1.76	0.34	1.98	0.17	0.96	0.56
1.77	0.48	1.84	0.18	2.05	0.23	1.03	0.49
1.83	0.45	1.90	0.14	2.13	0.10	1.10	0.55
2.02	0.22	2.02	0.27	2.20	0.29	1.18	0.35
2.07	0.22	2.11	0.19	2.43	0.09	1.55	0.46
2.35	0.46	2.55	0.15	2.50	0.15	1.65	0.12
2.40	0.33	2.65	0.31	2.58	0.16	1.74	0.07
2.65	0.28	2.86	0.07	2.65	0.27	1.81	0.44
2.73	0.26	2.95	0.28	2.88	0.31	1.89	0.12
2.82	0.23	3.59	0.05	2.96	0.35	1.96	0.60
2.90	0.28	3.66	0.08	3.14	0.21	2.04	0.46
3.35	0.28	3.79	0.16	3.23	0.09	2.11	0.75
3.40	0.31	3.90	0.09	3.32	0.32	2.25	0.82
3.48	0.32	3.96	0.15	3.39	0.44	2.35	0.38
3.55	0.23	4.05	0.24	3.84	0.39	2.76	1.00
3.64	0.06	4.15	0.14	3.91	0.42	2.83	0.84
3.70	0.14	4.24	0.24	3.99	0.59	2.91	0.65
3.77	0.19	4.31	0.31	4.06	0.17	2.98	0.54
3.83	0.17	4.48	0.11	4.30	0.21	3.06	0.46
4.00	0.85	4.52	0.26	4.39	0.51	3.13	0.64
4.05	0.29	4.58	0.07	4.46	0.19	3.22	0.66
4.14	0.63	4.66	0.14	4.60	0.16	3.30	0.59
4.21	0.99	4.89	0.23	4.68	0.33	3.39	0.44
4.54	0.15	4.94	0.48	4.76	0.25	3.46	0.34
4.63	0.40	5.04	0.14	4.84	0.32	3.74	0.85
4.85	0.33	5.11	0.51	4.91	0.11	3.81	0.73
4.93	0.36	5.19	0.23	4.99	0.14	3.89	0.17
5.03	0.30	5.26	0.47	5.06	0.17	3.96	0.15
5.10	0.27	5.34	0.38	5.14	0.31	4.04	0.66
5.33	0.26	5.41	0.33	5.21	0.21	4.11	0.74
5.40	0.32			5.29	0.28	4.19	0.31
5.50	0.07			5.36	0.15	4.26	0.14
5.60	0.17			5.43	0.19	4.70	0.36
				5.51	0.21	4.78	0.07
						4.88	0.80
						4.97	0.60
						5.07	0.34
						5.15	0.04
						5.24	0.24
						5.33	0.21
						5.42	0.50
						5.51	0.69

Una vez desglosada la varianza en sus dos componentes, se ha pasado a calcular la varianza de extensión ( $\sigma_e^2$ ) para la "Sal de Muro" de la Fm. Salina de Cardona. Esta varianza, que representa el error que se comete al extrapolar una muestra puntual a un dominio mayor, se puede determinar mediante la fórmula :  $\sigma_e^2 = C_0 + (C * \sigma^2)$ , siendo  $C_0$  el valor del efecto de pepita y C el valor del componente espacial de la varianza (David, 1977). El valor de la función auxiliar ( $\sigma^2$ ), que depende de la relación entre distancia entre muestras (l) y alcance de la variable (a), está tabulado en la mayoría de los libros de Geostatística (David, 1977). Concretamente, para este muestreo la relación l/a es igual a 0,04, lo que corresponde a un valor de  $\sigma^2$  de 0,01. La desviación standard obtenida, mediante el cálculo de la varianza de extensión para el muestreo efectuado, es de 0,13 % (valor inferior al de 0,20 %, determinado por los procedimientos clásicos de Estadística).

La figura 2 muestra los semivariogramas experimental y ajustado obtenidos en los niveles halíticos de Fm. Zaragoza. El número de muestras estudiadas ha sido de 34 y el intervalo de muestreo ha sido 10 cm. El semivariograma experimental muestra a grandes rasgos un comportamiento similar al obtenido en "Sal de muro" de la Fm. Salina de Cardona. La varianza total aumenta hasta alcanzar la meseta y disminuye ligeramente. En el extremo final del semivariograma tiende a fluctuar el valor de la varianza, indicando una cierta variabilidad en el contenido en agua a lo largo de la formación. El ajuste del semivariograma a un modelo esférico ha arrojado los siguientes parámetros: efecto pepita de 0,006 %<sup>2</sup>, una componente espacial de la varianza de 0,010 %<sup>2</sup>, una meseta de 0,016 %<sup>2</sup> y un alcance de 1,0 metros. El grado de sesgo de la formación es del 60 %. En lo que respecta a la determinación de la varianza de extensión, la relación l/a es igual a 0,1, lo que corresponde a un valor de  $\sigma^2$  de 0,025. La desviación standard obtenida, mediante el cálculo de la varianza de extensión para el muestreo efectuado, es de 0,079 % (valor inferior al estadístico de 0,11 %).

A diferencia del estudio de la variabilidad espacial, realizados a escala centimétrica en las formaciones anteriormente estudiadas, se ha estudiado la variabilidad del contenido en agua en una formación salina a escala métrica. La figura 3 muestra el semivariograma experimental obtenido en la "Sal de Muro" de la Fm. Guendulain. El número de muestras estudiadas ha sido 25 y el intervalo de muestreo ha sido 5 m. El semivariograma experimental muestra un comportamiento totalmente diferente al observado en las otras formaciones. La

Tabla 2. Localización espacial y contenido en agua (expresado en %) de las muestras procedentes de los niveles halíticos de la Fm. Zaragoza y de la "Sal de Muro" de la Fm. Guendulain.

Table 2. Spatial location and water content (expressed in %) of the samples from the halitic horizons of the Zaragoza Formation and the Lower Salt Unit of the Guendulain Formation.

Niveles halíticos de la Fm. Zaragoza		Sal de Muro de la Fm. Guendulain	
x (m)	H2O (%)	x (m)	H2O (%)
0.05	0.18	0	0.63
0.15	0.13	5	0.53
0.40	0.15	11	1.38
0.65	0.25	16	0.02
0.70	0.23	21	0.54
0.75	0.33	26	0.55
0.80	0.33	30	0.17
0.90	0.08	36	0.46
1.00	0.42	41	0.01
1.20	0.11	46	0.36
1.30	0.10	51	0.37
1.50	0.21	54	0.37
1.75	0.17	60	0.41
2.10	0.26	65	0.20
2.30	0.07	70	0.11
2.60	0.15	75	0.06
2.80	0.13	79	0.37
2.90	0.10	85	1.08
3.10	0.07	90	0.25
3.40	0.27	95	0.85
3.50	0.10	100	0.08
3.60	0.28	105	0.18
3.65	0.43	110	0.51
3.70	0.30	115	0.40
3.80	0.38	120	0.21
3.90	0.47		
4.00	0.39		
4.05	0.22		
4.10	0.35		
4.20	0.23		
4.25	0.25		
4.35	0.37		
4.50	0.20		
4.70	0.09		

varianza total se encuentra ya, desde el primer intervalo de muestreo, en una zona de meseta con pequeñas fluctuaciones, ligadas a la variación del contenido en agua a lo largo de la formación. En este caso podemos hablar de un semivariograma en el que el efecto pepita es igual a la varianza total. A esta escala de muestreo, la variable en estudio sigue un comportamiento errático a lo largo de la formación. Consecuentemente, el grado de sesgo tiende a infinito, siendo imposible la determinación de la varianza de extensión. Por dicho motivo, para la determinación de los errores de estimación en el muestreo hay que utilizar los estadísticos convencionales.

Caso de existir algún tipo de estructura a gran escala a lo largo de la formación, esta debería tener un alcance superior a los 120 m. Asimismo, asumiendo que el alcance de la variable a pequeña escala fuera de 1 m (similar a los obtenidos para la "Sal de Muro" de la Fm. Salina de Cardona y en los niveles halíticos de la Fm. Zaragoza), la relación  $l/a$  sería de 5, lo que corresponde a un valor de  $\sigma^2$  de 1. En dicho caso la varianza de extensión correspondería a la varianza total, tal como se verifica en "Sal de Muro" de la Fm. Guendulain para este intervalo de muestreo de 5 m. Por dicho motivo es verosímil pensar que la "Sal de Muro" de la Fm. Guendulain debe presentar un comportamiento de la variabilidad espacial similar a los observados en las dos formaciones estudiadas a nivel centimétrico.

La tabla 4 presenta un sumario donde se comparan las desviaciones standard de las 3 muestras poblacionales, utilizando el análisis estadístico convencional, la obtenida mediante análisis geoestadístico utilizando el intervalo de muestreo realizado y las que se obtendrían utilizando otros intervalos de muestreo correspondientes a una relación  $l/a$  de 0,5 y de 1.

Finalmente, a partir del estudio de la varianza de extensión, puede plantearse la optimización del muestreo. Una densidad óptima de muestreo es aquella en la cual, a

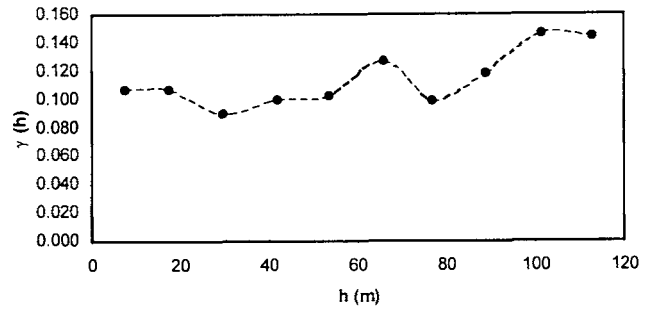


Figura 3. Semivariograma experimental del contenido en agua de la "Sal de Muro" de la Fm. Guendulain.

Figure 3. Experimental semivariogram for the water content of the Lower Salt Unit of the Guendulain Formation.

partir de un número mínimo de muestras, se obtiene un grado de fiabilidad determinado. En este caso se ha utilizado una formación salina virtual, a la que se le han asignado una serie de parámetros similares a los obtenidos en el estudio geoestadístico, realizado en las tres formaciones anteriores. Se ha asumido que el semivariograma virtual es independiente del número y de la localización de las muestras. Los parámetros de base han sido: media de 0,30%, varianza total de 0,023%<sup>2</sup>, efecto pepita de 0,010%<sup>2</sup>, un componente espacial de la varianza de 0,013%<sup>2</sup> y alcance de la variable de 1 m. Asimismo se han considerado intervalos de muestreo de 0,5m, 1m, 2m y 5m. A partir de valores de intervalo de muestreo iguales o superiores a 5 metros, hay que considerar que el valor de la varianza de extensión es la misma que la varianza obtenida por los métodos estadísticos clásicos.

La estimación del valor medio de la variable para todo el cuerpo geológico se ha realizado mediante la extrapolación del valor promedio de n muestras. En este caso, la varianza de estimación se obtiene dividiendo el valor de la varianza de extensión por el número de bloques que conforman el cuerpo geológico en estudio,

Tabla 3. Estadísticos del contenido en agua (expresado en %) de las formaciones salinas estudiadas.

Table 3. Water content statistics (expressed in %) of the studied salt formations.

Formación salina	Rango	Media	Mediana	Desv. Standard	N
S.M. Fm. Cardona	0.04-1.00 %	0.32 %	0.28 %	0.20 %	175
Fm. Zaragoza	0.07-0.47 %	0.23 %	0.22 %	0.11 %	34
S.M. Fm. Guendulain	0.01-1.38 %	0.40 %	0.37 %	0.32 %	25

Tabla 4. Evolución de la variabilidad en las 3 formaciones estudiadas (expresada como desviación standard) en función del análisis (estadístico o geoestadístico) utilizado. En los análisis geoestadísticos la opción A representa la relación l/a utilizada a partir del muestreo (0.04 para la "Sal de Muro" de la Fm. Cardona y 0.1 para los niveles halíticos de la Fm. Zaragoza). Las opciones B y C representan la variabilidad que se obtendría para unas relaciones l/a de 0.5 y 1 respectivamente.

Table 4. Variability in the 3 studied formations (expressed as standard deviation) versus type of analysis (statistical or geostatistical) performed. For the geostatistical analyses option A stands for the actual l/a ratio (0.04 for the Lower Salt Unit of the Cardona Formation and 0.1 for the halitic horizons of the Zaragoza Formation). Options B and C represent variability which would be obtained for l/a ratios of 0.5 and 1 respectively.

Formación salina	Estadística	Geoestadística		
		Opción A	Opción B	Opción C
S.M. Fm. Cardona	0.201	0.130	0.142	0.153
Fm. Zaragoza	0.114	0.079	0.085	0.095
S.M. Fm. Guendulain	0.324	-	-	-

asumiendo que cada bloque contiene una muestra tomada al azar (Wellmer, 1998). La figura 4 muestra la curva de la optimización de muestreo, para un nivel de confianza del 95 %, a partir de la cual se puede determinar el mínimo número de muestras que ofrezca un grado satisfactorio de fiabilidad. En dicha figura se observa, para, todos los intervalos de muestreo estudiados, cómo a partir de un número determinado de muestras (entre 30 y 40), la curva se vuelve asintótica. El ampliar la campaña de muestreo tiene, por consiguiente, un efecto limitado en reducir la incertidumbre. Otro aspecto significativo es que el error cometido por muestreos con intervalos de 0,5 a 2,0 m es muy próximo. Este aspecto hay que atribuirlo al alto grado de sesgo que presenta la variable en estudio.

Los resultados obtenidos se han querido extrapolar para ubicar el valor promedio de la variable en un volumen determinado. Se han considerado 2 casos: Un cuerpo cuyas dimensiones son 50x50x15 m y otro de dimensiones 10x10x4 m. En ambos casos se ha intentado definir la relación óptima entre el número de muestras y la eficiencia en la estimación, para un error relativo del 10%. En el primer caso, un muestreo sistemático cada 5 o 10 m requeriría 300 o 75 muestras respectivamente. El error relativo sería de 5,58% y 11,1% respectivamente. Un muestreo sistemático cada 8,5 m ( $n \approx 100$ ) permitiría optimizar el error por debajo del 10% (9,6%). Para el segundo caso, un muestreo sistemático cada 2 o 5 m requeriría 50 o 4 muestras respectivamente. El error relativo sería de 11,7 % y 48,3 % respectivamente. Un muestreo sistemático cada 1,5 m ( $n \approx 70$ ) permitiría optimizar el error por debajo del 10% (9,7%). Se observa cómo para el primer caso la aproximación estadística

clásica es suficiente para estimar el valor promedio de la variable con una precisión aceptable, mientras que para el segundo caso la minimización del error conlleva la utilización del análisis geoestadístico.

Lo anteriormente expuesto permite extraer una conclusión interesante: el uso de la Geoestadística, para la minimización de los errores de estimación, es justificable cuando las dimensiones del cuerpo geológico a evaluar no excede en un orden de magnitud el alcance de la variable en las tres direcciones ortogonales del espacio. En caso contrario, las medidas pueden considerarse independientes, por lo que la Estadística convencional puede ofrecer un nivel satisfactorio de precisión en la determinación del valor promedio de la variable.

## CONCLUSIONES

El contenido de agua en las formaciones salinas no es una variable aleatoria sino regionalizada, es decir que el valor de la variable en un punto condiciona el valor de ésta en puntos adyacentes. Por dicho motivo, mediante el uso del análisis geoestadístico, la varianza en una población estudiada puede descomponerse en su componente aleatorio (efecto pepita) y en su componente espacial. En este trabajo se ha realizado un análisis geoestadístico asumiendo que las formaciones salinas estudiadas son isótropas en las tres direcciones espaciales y que el semivariograma en cualquier sector de la formación es similar a los obtenidos experimentalmente.

Los semivariogramas realizados presentan diferencias según si el intervalo de muestreo sea a escala centimétrica



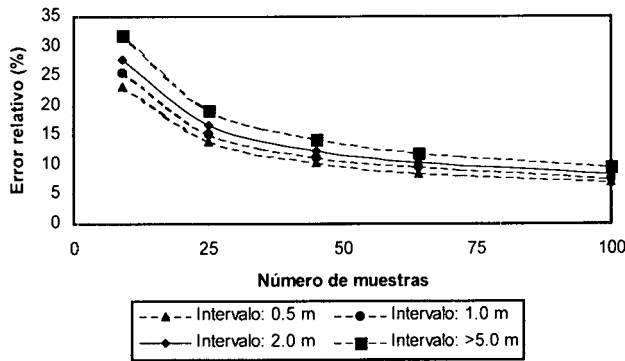


Figura 4. Ejemplo para la optimización del muestreo en una formación salina virtual que presenta los siguientes parámetros de variabilidad: media de 0.30%, varianza total de 0.023%, efecto pepita de 0.010%, un componente espacial de la varianza de 0.013% y alcance de la variable de 1 m. En este ejemplo se han considerado intervalos de muestreo de 0.5m, 1m, 2m y 5m

Figure 4. Optimised sampling procedure in a virtual salt formation which exhibits following parameter: average of 0.30%, total variance of 0.023%, nugget of 0.010%, pitch of 0.013% and range of the variable of 1 m. Sampling intervals of 0.5m, 1m, 2m and 5m have been considered.

o a escala métrica. Para intervalos de muestreo centimétrico, los semivariogramas pertenecen al tipo esférico y el alcance de la variable regionalizada es del orden de un metro. El grado de sesgo, expresado como la relación entre el componente aleatorio de la varianza y el componente espacial de la varianza, oscila entre el 60 y el 80%. Para intervalos de muestreo métrico, los semivariogramas no presentan ninguna estructura, siendo el comportamiento de la variable a esta escala totalmente aleatorio. Caso de existir algún tipo de estructura a gran escala a lo largo de la formación, ésta debería tener un alcance superior a los 120 m.

A partir del estudio de los semivariogramas realizados con intervalo de muestreo a escala centimétrica, el estudio de la varianza de extensión muestra que la extrapolación del valor de un punto es solo válida para distancias inferiores a 5m. Dicho fenómeno ha podido ser verificado en el semivariograma que presentaba un intervalo de muestreo a escala métrica.

El estudio de la optimización del muestreo en una formación salina virtual, a la que se le han asignado una serie de parámetros similares a los obtenidos en el estudio geoestadístico, ha sido realizado para intervalos de muestreo de 0,5m, 1m, 2m y 5m. El estudio ha mostrado que, para todos los intervalos de muestreo mencionados, a partir de un número determinado de

muestras (entre 30 y 40), la eficiencia en la disminución del error de estimación decrece significativamente. Asimismo, debido al alto grado de sesgo de la variable, el error cometido por muestreos con intervalos de 0,5 a 2,0 m es muy próximo.

Del estudio de la optimización del muestreo, se ha podido extraer una conclusión interesante. El uso de la Geoestadística, para la minimización de los errores de estimación, es justificable cuando las dimensiones del cuerpo geológico a evaluar no excede en un orden de magnitud el alcance de la variable en las tres direcciones ortogonales del espacio. En caso contrario, la Estadística convencional puede ofrecer un nivel satisfactorio de precisión en la determinación del valor promedio de la variable.

#### AGRADECIMIENTOS

El autor quiere agradecer a los Drs. K. Bitzer, J. Carrera y M. Viladevall, la lectura crítica del manuscrito así como sus sugerencias, las cuales han permitido mejorar la calidad del mismo.

#### BIBLIOGRAFÍA

- Chica-Olmo, M., 1988. Análisis geoestadístico en el estudio de la explotación de los recursos minerales. Granada, Ed. Univ. Granada, 387p.
- David, M., 1977. Geostatistical ore reserve estimation. Developments in Geomathematics, 2, Amsterdam, Elsevier, 364p.
- Davis, J. C., 1986. Statistics and data analysis in Geology. New York, John Wiley and Sons, 646p.
- De las Cuevas, C., 1992. Análisis del contenido en agua en formaciones salinas. Su aplicación al almacenamiento de residuos radiactivos. Tesis doctoral. Univ. Barcelona, 175 p.
- De las Cuevas, C., Pueyo, J. J., 1991. Caracterización y cuantificación termogravimétrica de los diferentes tipos de agua presentes en formaciones salinas. Bol. Soc. Esp. Min. 14, 15-21.
- De las Cuevas, C., Pueyo, J. J., 1995. The influence of mineralogy and texture in the water content of rock salt formations. Its implications in radioactive disposal. Appl. Geochemistry, 10, 317-327.
- EPA, 1988. GEO-EAS (Geostatistical Environmental Assessment Software) User's Guide. Las Vegas, EPA, 183p.
- Rock, N. M. S., 1988. Numerical Geology. Lecture Notes in Earth Sciences, 18. Berlin, Springer Verlag, 427p.
- Royle, A. G., 1979. Why Geostatistics? Eng. Min. Journal, 114, (5), 92-101.
- Spiegel, M. R., 1985. Estadística. Madrid, Mc Graw-Hill, 357p.
- Wellmer, F. W., 1998. Statistical evaluation in exploration for mineral deposits. Berlin, Springer Verlag, 379p.

## ANEXO

Las variables regionalizadas son continuas entre punto y punto, pero los cambios en la variable son tan complejos que no pueden ser descritos por ninguna función determinística. Aunque una variable regionalizada es continua en el espacio, ésta se suele determinar a través de un muestreo en puntos geográficamente localizados. La función matemática que mide el grado de dependencia espacial entre muestras según una dirección determinada es denominada semivariograma. Dicha función expresa la semivarianza de los incrementos de la variable regionalizada en relación con su distancia. Si se admite que en el punto  $x$  la variable tiene un valor  $Z(x)$  y en el punto  $x+h$  la variable tiene un valor igual a  $Z(x+h)$ , la semivarianza de la variable regionalizada, para una distancia  $h$ , es igual a:

$$\gamma(h) = \frac{1}{2N(h)} \sum [Z(x) - Z(x+h)]^2$$

donde

- $h$  es un vector que representa el intervalo de separación entre dos parejas de puntos de muestreo;
- $N(h)$  es el número de parejas de puntos de muestreo para un intervalo de separación  $h$ .

Si se calculan las semivarianzas para diferentes valores de  $h$ , se puede representar los resultados en forma de un gráfico denominado semivariograma tal como muestran las figuras 1 y 2. En él se observa cómo, a medida que aumenta el valor de  $h$ , los puntos comparados

presentan cada vez menor correlación entre ellos y la diferencia entre los valores de la variable son mayores. A una cierta distancia los puntos comparados no suelen presentar correlación, por lo que la semivarianza ya no aumenta y el semivariograma se estabiliza, dando una región plana, denominada meseta. Dicha distancia es conocida con el nombre de alcance y correspondería a la zona de influencia de una muestra. En algunos casos, para valores de  $h$  grandes (grandes distancias), el semivariograma no se estabiliza, creciendo  $\gamma(h)$  más que el valor  $h$ . En esos casos se habla de que la variable regionalizada presenta una deriva.

Teóricamente, la varianza de los incrementos en el origen es igual a 0 (y, por tanto, la correlación es perfecta), aunque a veces el semivariograma puede presentar una discontinuidad en el origen. Dicha discontinuidad, a pequeña escala, es denominada "efecto pepita". En el caso de que el semivariograma proporcione una línea horizontal existe aleatoriedad pura, lo que implica que no existirá correlación entre los datos por muy próximos que estén. En dicho caso las diferencias entre los puntos medidos son debidas únicamente al azar y no existe una distribución geográfica estructurada en el fenómeno natural.

Una vez obtenidos los semivariogramas experimentales, éstos deben de ser posteriormente ajustados a un modelo teórico. Entre los más frecuentes se encuentran el modelo esférico, el exponencial, el gaussiano y el lineal, pudiendo en todos los casos presentar o no efecto pepita.