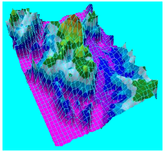
GEOESTADÍSTICA APLICADA

Tema: Análisis Variográfico

Dr. Martín A. Díaz Viera,
Dr. Ricardo Casar González

UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO
mdiazv@imp.mx



Contenido

Introducción

- El análisis variográfico o estructural es uno de los tópicos más importantes de la geoestadística puesto que se encarga de la caracterización de la estructura espacial de una propiedad o fenómeno.
- Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia.

Introducción

- El análisis estructural consiste en estimar y modelar una función que refleje la correlación espacial de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad.
- En dependencia de las características de estacionaridad del fenómeno se modelará la función de covarianzas o la de semivarianzas.
- Por su importancia y generalidad estudiaremos el proceso de estimación y modelación de la función de semivarianzas o semivariograma.

El semivariograma. Definición

- El **semivariograma**, conocido también como **variograma**, es la herramienta central de la geoestadística. Dada una FA $\mathbf{Z}(\underline{x})$ que cumpla la Hipótesis Intrínseca entonces existe la función semivarianza y se define como sigue:

$$\gamma(\underline{h}) = \frac{1}{2} Var\left[Z(\underline{x}) - Z(\underline{x}+\underline{h})\right] = \frac{1}{2} E\left[\left\{Z(\underline{x}) - Z(\underline{x}+\underline{h})\right\}^2\right] \tag{1}$$

Estimación del semivariograma

- Existen varios estimadores pero la forma de estimación más común está dada por

$$\gamma(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} \left[Z(\underline{x}_i + \underline{h}) - Z(\underline{x}_i)\right]^2 \tag{2}$$

- $N(\underline{h})$ es el número de pares $Z(\underline{x}_i)$ y $Z(\underline{x}_i + \underline{h})$ separados a una distancia $h = |\underline{h}|$

Estimación del semivariograma

- Debido a que el estimador es esencialmente una media muestral, tiene todas las desventajas comúnmente asociadas a este tipo de estimadores como es la no robustez.
- A menudo el empleo de este estimador produce variogramas experimentales erráticos, lo cuál se debe a desviaciones del caso ideal para la aplicación del mismo.

Estimación del semivariograma

- Estas desviaciones pueden ser enumeradas por su importancia en el orden siguiente:

1. Desviaciones en la distribución (asimetría)
2. No homogeneidad de la varianza
3. Desviaciones en el muestreo
4. Existencia de valores atípicos (outliers)

Forma general del variograma

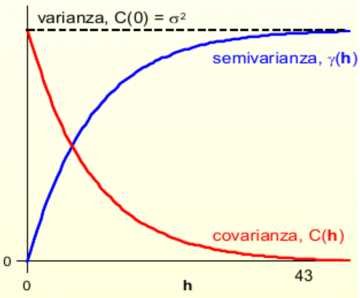


Figura 1: Relación de las funciones de covarianzas y semivarianzas.

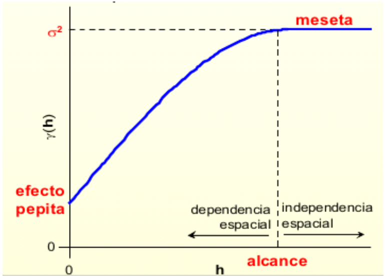


Figura 2: Parámetros del variograma: meseta (sill), alcance (range) y efecto pepita (nugget effect).

Consideraciones para el cómputo del variograma

- Los pares de las observaciones se agrupan según la distancia dentro de un intervalo con una tolerancia y dentro de una dirección con una tolerancia angular.
- Se estima para valores menores que la mitad de la distancia máxima.
- 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 variograma.
- Se considera que debe haber entre 30 y 50 pares de puntos como mínimo por intervalo.
- Los valores estimados para cada intervalo se deben graficar contra la distancia promedio de todos los pares que se encuentran dentro de dicho intervalo.

Cómputo del variograma

Figura 3: Ejemplo del cómputo del variograma en 1D.

$$\mathbf{h} = 1: (10-7)^2+(11-10)^2+(13-11)^2 +(12-13)^2+(14-12)^2+(12-14)^2 +(13-12)^2+(10-13)^2+(11-10)^2 +(9-11)^2+(8-9)^2 = 39$$

$$N(\mathbf{1}) = 11$$

$$\hat{\gamma}(\mathbf{1}) = 39/22 = 1.78$$

Consideraciones para el cómputo del variograma

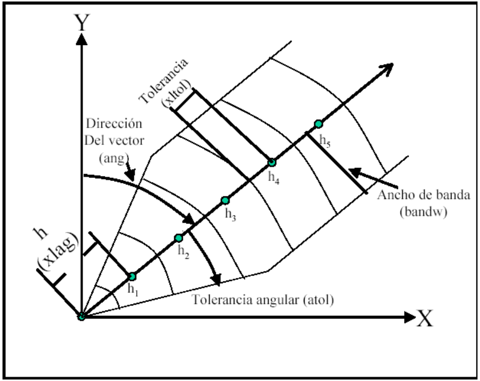


Figura 4: Diagrama del cómputo del variograma.

Distribución Espacial

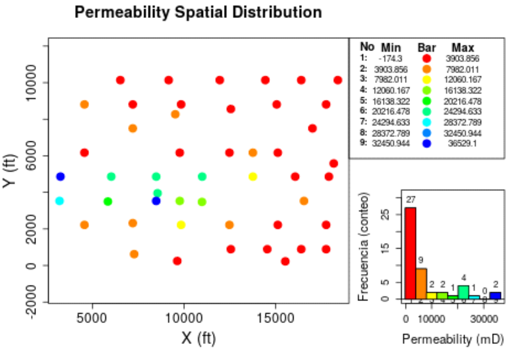


Figura 5: Distribución espacial.

Ejemplo de variograma estimado

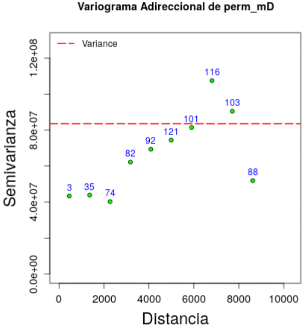


Figura 6: Ejemplo de variograma estimado.

Estadígrafo	Valor
Muestras	48
Mínimo	7.4
1° cuartil	1002.45
Mediana	3482.205
Media	6818.24521
3° cuartil	7743.625
Máximo	36347.4
Rango	36340
Rango intercuartil	6741.175
Varianza	83532706.36
Desviación estándar	9139.62288
Simetría	1.83579
Curtosis	5.62603

Tabla 1: Estadística básica.

Ejemplo de variograma estimado

Dirección = 0 grados, Tolerancia = 90.0 grados, Intervalo = 900 ft.

No. lag	**No. pares**	**Intervalos (lags)**	**Semivarianzas**
1	3	453.71	43,357,868.50
2	35	1361.14	43,869,273.09
3	74	2268.56	40,249,566.62
4	82	3175.98	62,199,450.34
5	92	4083.41	69,343,355.94
6	121	4990.83	74,381,397.02
7	101	5898.25	81,429,116.65
8	116	6805.68	107,465,654.01
9	103	7713.10	90,459,560.73
10	88	8620.52	51,904,218.42

Tabla 2: Ejemplo de variograma estimado.

Modelos autorizados

- Sólo ciertas funciones pueden ser consideradas modelos válidos de variogramas y se dicen que son **modelos autorizados**.
- El hecho de probar si una función dada es aceptable o no, está relacionado con el examen de su Transformada de Fourier. [Christakos,1984] obtuvo las condiciones que el espectro de la función debe reunir para ser un modelo autorizado.
- Como una propiedad importante se debe destacar que cualquier combinación lineal de modelos autorizados es un modelo autorizado.

Modelos transitivos o acotados

- Este grupo de modelos se deriva a partir de la noción de autocorrelación entre los valores promedios dentro de los bloques.
- La idea es que la función aleatoria, de la cual la propiedad medida es una realización, depende del grado de traslape de los dos bloques.

Modelo Esférico

$$\gamma(h) = \begin{cases} \frac{S}{2}\left\{3\left(\frac{h}{a}\right) - \left(\frac{h}{a}\right)^3\right\} & \text{para } 0 \leq h \leq a \\ S & \text{para } h > a \end{cases} \tag{4}$$

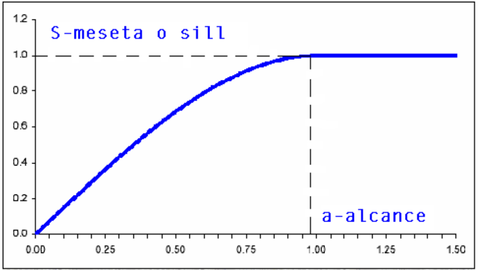


Figura 7: Modelo Esférico.

Modelo Exponencial

$$\gamma(h) = S\left[1 - exp\left(-\frac{3h}{a}\right)\right] \text{ para } h \geq 0 \qquad (5)$$

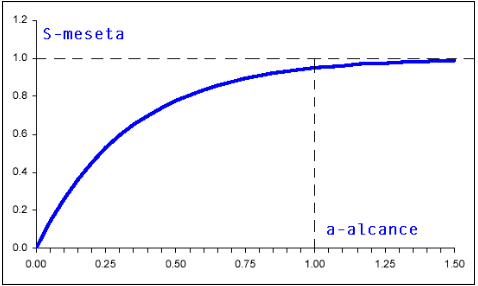


Figura 8: Modelo Exponencial.

Modelo Gaussiano

$$\gamma(h) = S\left[1 - exp\left(-\left(\frac{3h}{a}\right)^2\right)\right] \text{ para } h \geq 0 \tag{6}$$

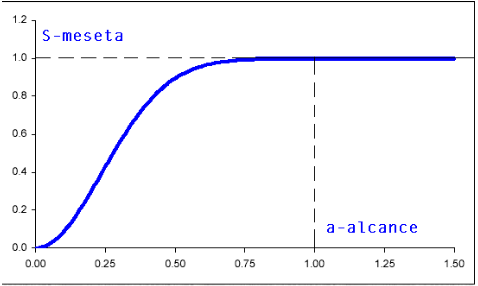


Figura 9: Modelo Gaussiano.

Modelo Efecto Agujero (Hole)

$$\gamma(h) = S\left(1 - cos\left(\frac{\pi h}{a}\right)\right) \text{ para } h \geq 0 \tag{7}$$

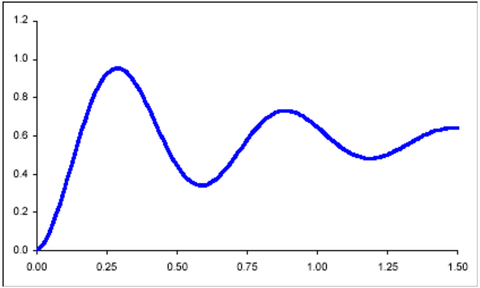


Figura 10: Modelo Efecto Agujero.

Modelo Efecto Pepita (Nugget)

$$\gamma(h) = S\,(1 - \delta(h)) \tag{8}$$

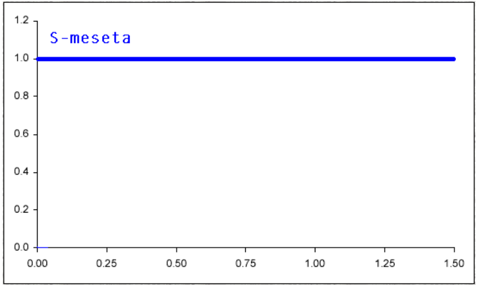


Figura 11: Modelo Efecto Pepita (Nugget).

Influencia del Alcance

Dominio de 100x100 metros. Modelo de variograma con diferentes alcances

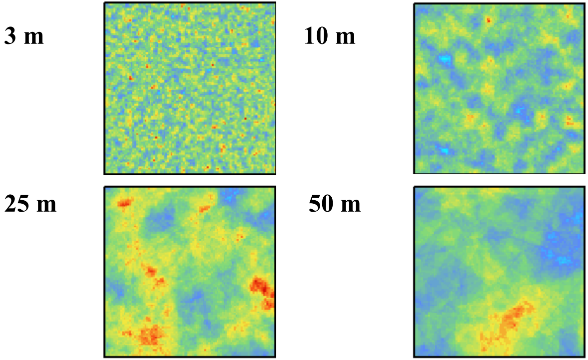


Figura 12: Influencia del Alcance.

Influencia del Efecto Pepita

Sin Efecto Nugget

33% Nugget

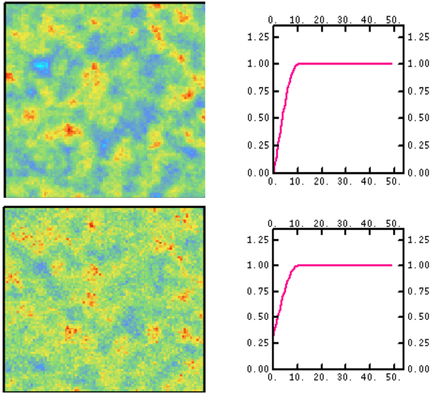


Figura 13: Influencia del Efecto Pepita.

Influencia del Efecto Pepita

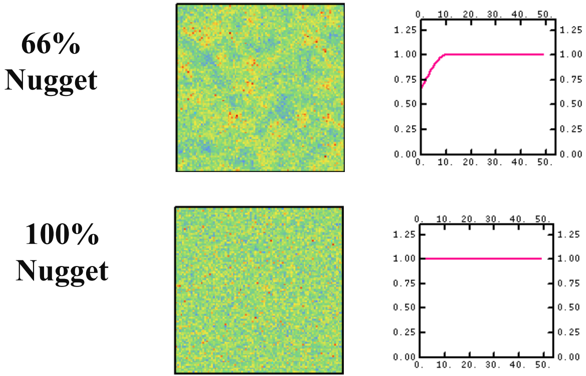


Figura 14: Influencia del Efecto Pepita.

Modelos no acotados

- Existen casos en que la varianza aparenta incrementarse indefinidamente.
- Si se toma cada vez un menor intervalo de muestreo, siempre existe alguna variación que queda sin resolver.
- [Mandelbrot,1982] llamó al resultado de tales procesos fractales.
- Ejemplos: El ruido Gaussiano fraccional (fGn) y el movimiento Browniano fraccional (fBm).
- Para conocer más detalles ver la presentación CG8a sobre Geoestadística Fractal.

Modelos no acotados

- Modelo Potencia

$$\gamma(h) = \frac{1}{2} h^{\theta} \text{para } 0 < \theta < 2 \tag{9}$$

- Relación entre la dimensión fractal **D** (Hausdorff-Besicovitch) y el parámetro θ

$$D = 2 - \left(\frac{\theta}{2}\right) \tag{10}$$

Casos Extremos:

1. $\theta = 2$ y **D**$=1$, es una parábola, no representa un proceso aleatorio.
2. $\theta = 0$ y **D**$=2$, ruido puro (efecto nugget).

Combinación de Modelos

- Una combinación lineal de modelos de variogramas autorizados con coeficientes positivos representa un modelo de variograma válido.
- Usualmente los modelos anteriores los encontramos como combinaciones del tipo:

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) \tag{11}$$

Donde $\gamma_0(h)$ es el efecto nugget y $\gamma_1(h)$ es otro modelo.

Combinación de Modelos

- La combinación formada por variogramas con diferentes rangos se conoce como **estructura anidada**.
- Describen variaciones espaciales a diferentes escalas y que se deben por lo tanto, a factores de naturaleza diferente.
- Por ejemplo:

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) + \gamma_2(h) \tag{12}$$

Donde $\gamma_0(h)$ es el efecto nugget, $\gamma_1(h)$ es el modelo esférico 1 y $\gamma_2(h)$ es el modelo esférico 2.

Combinación de Modelos

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) + \gamma_2(h)$$

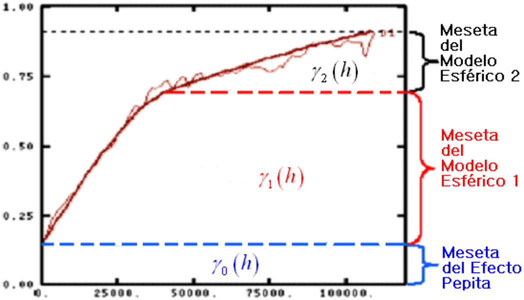


Figura 15: Combinación de Modelos.

Anisotropía Geométrica

- Existen numerosas situaciones en que la variación es anisotrópica, es decir depende de la dirección.
- Cuando la anisotropía se puede tener en cuenta mediante una transformación lineal simple de las coordenadas, entonces se dice que la **anisotropía es geométrica** o afín.
- La anisotropía se refleja en diferentes alcances según la dirección. El gráfico direccional de los rangos forma una elipse.

Variogramas Anisotrópicos

- Fórmula de la transformación:

$$\Omega(\theta) = \left(\mathbf{A}^2 cos^2(\theta - \phi) + \mathbf{B}^2 sen^2(\theta - \phi)\right)^{\frac{1}{2}} \tag{13}$$

A es el eje mayor (variabilidad es más lenta).
B es el eje menor (variabilidad es más rápida)
ϕ es la dirección (ángulo) del eje mayor.

- Se aplica como un factor al argumento h del variograma en los modelos acotados o al gradiente en los modelos no acotados.
- $\lambda = \mathbf{A}/\mathbf{B}$ es una medida de la anisotropía.

Anisotropía Geométrica

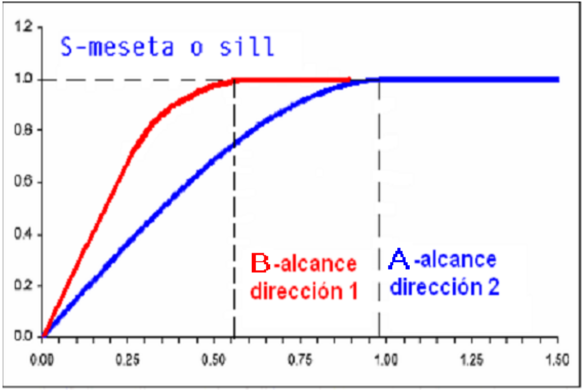


Figura 16: Anisotropía Geométrica.

Anisotropía Geométrica

- En la práctica se estiman los variogramas en 4 direcciones principales (0, 45, 90 y 135 grados).
- Se determinan los rangos para cada dirección.
- Luego se construye el gráfico direccional de los rangos para decidir si hay anisotropía geométrica presente o no.
- Finalmente, se determinan A, B y la dirección (ángulo) de mayor alcance.

Anisotropía Geométrica

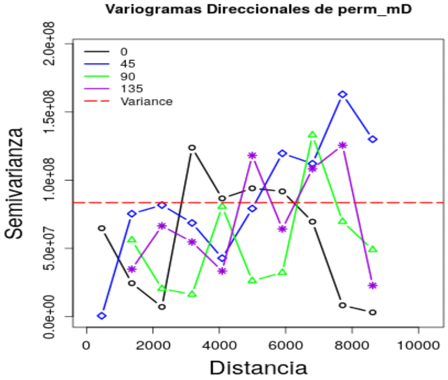


Figura 17: Variogramas direccionales en cuatro direcciones: 0, 45, 90 y 135 grados.

Anisotropía Zonal

- Cuando la anisotropía se refleja en la meseta, es decir, en dependencia de la dirección el variograma presenta diferentes mesetas.
- Un ejemplo típico es la situación cuando tenemos medida cierta propiedad en diferentes pozos.
- Por lo general el variograma en la vertical (en profundidad) presenta una mayor variabilidad (meseta) que en la dirección horizontal.

Anisotropía Zonal

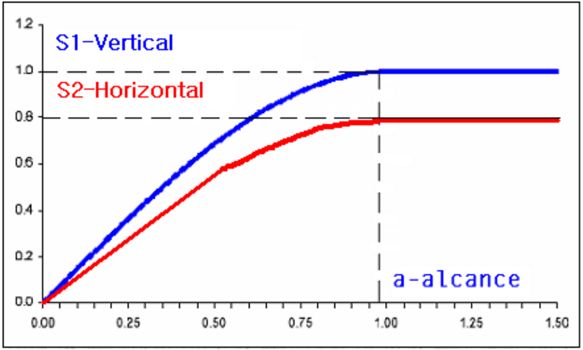


Figura 18: Anisotropía Zonal.

Modelación del variograma

- Algunos geoestadísticos ajustan los modelos de forma visual.
- Es recomendable auxiliarse con algún procedimiento estadístico.
- Por ejemplo, un ajuste con un Método de Mínimos Cuadrados.
- Se requiere de un criterio que considere la bondad del ajuste y la complejidad del modelo.
- El criterio de información de Akaike (AIC), que se define como

$$\mathrm{AIC} = -2\ \ln(\textit{máxima verosimilitud}) + 2\,(\textit{número de parámetros}) \tag{14}$$

Ajuste con mínimos cuadrados

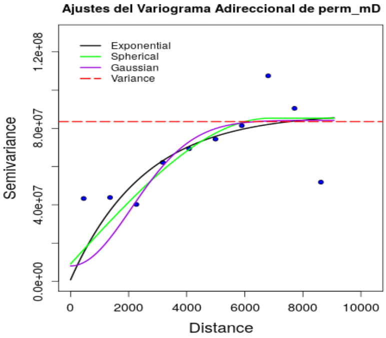


Figura 19: Ajuste con mínimos cuadrados.

Modelo	Nugget	Meseta+Nugget	Alcance	SCE
Exponential	814,495.7	87,992,879	2,533.95	2.79 e+15
Spherical	9,113,830.9	85,317,292	6,882.69	2.56 e+15
Gaussian	8,103,578.2	84,238,551	2,952.40	3.26 e+15

Tabla 3: Ajuste con mínimos cuadrados.

Mejor ajuste con mínimos cuadrados

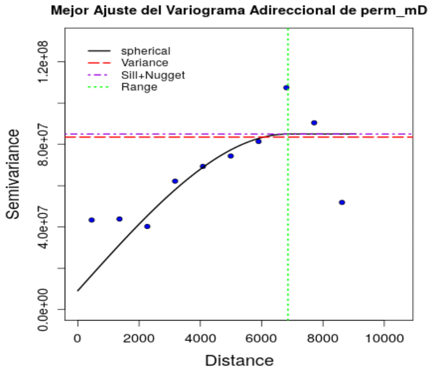


Figura 20: Mejor ajuste con mínimos cuadrados.

Ajuste con visual

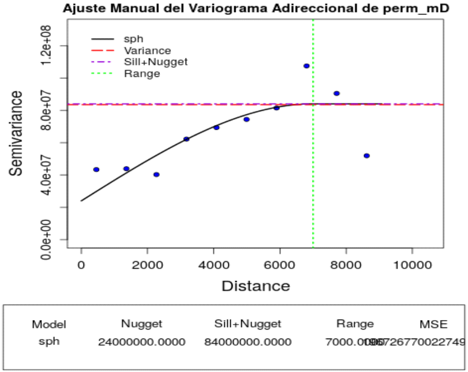


Figura 21: Mejor ajuste visual.

Validación cruzada del modelo

- Un método sencillo y eficiente es el método de **validación cruzada** conocido como **leave one out**.
- Consiste en sacar un punto de la muestra y estimar con **Kriging** el valor en ese punto usando el modelo de variograma obtenido.
- De forma análoga se actúa para el resto de los elementos de la muestra.
- Como resultado se obtiene un mapa y la estadística de las diferencias $(Z(\underline{x}_i) - Z^*(\underline{x}_i))$ entre el valor observado $Z(\underline{x}_i)$ y el valor estimado $Z^*(\underline{x}_i)$ para cada ubicación $\underline{x}_i$.

Validación cruzada del modelo

- Si el modelo del variograma refleja adecuadamente la estructura espacial del conjunto de datos, entonces los valores estimados $Z^*(\underline{x}_i)$ deben ser cercanos a los valores observados $Z(\underline{x}_i)$.
- Esta **cercanía** puede ser caracterizada según los siguientes criterios:

$$\frac{1}{n}\sum_{i=1}^{n}\{Z(\underline{x}_i) - Z^*(\underline{x}_i)\} \qquad \text{cercano a 0} \tag{15}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{Z(\underline{x}_i) - Z^*(\underline{x}_i)\}^2 \qquad \text{pequeño} \tag{16}$$

- El histograma de las diferencias $Z(\underline{x}_i) - Z^*(\underline{x}_i)$ permite identificar **valores atípicos espaciales**, datos sospechosos o anomalías de otra naturaleza.

Validación cruzada del modelo

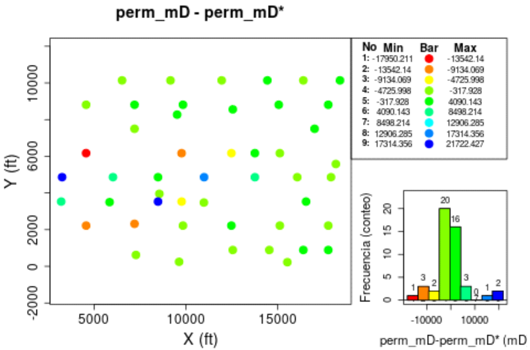


Figura 22: Mapa de la validación cruzada del modelo.

Validación cruzada del modelo

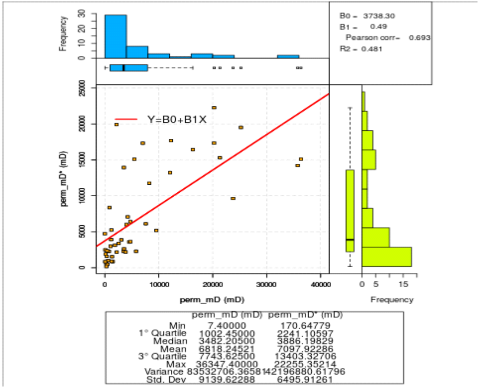


Figura 23: Valores estimados vs. valores reales.

Validación cruzada del modelo

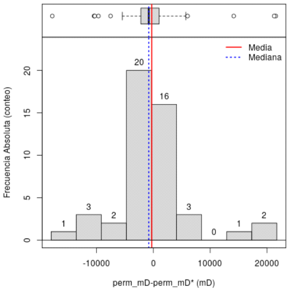


Figura 24: Histograma de las diferencias.

Estadígrafo	**Valor**
Muestras	48
Mínimo	-17753.81
1° cuartil	-2101.11
Mediana	-799.22
Media	-279.68
3° cuartil	985.77
Máximo	21526.03
Rango	39279.84
Rango intercuartil	3086.88
Varianza	43409713.43
Desviación estándar	6588.60
Simetría	1.13
Curtosis	7.07

Tabla 4: Estadística de las diferencias.

Referencias

Christakos(1984)
On the Problem of Permissible Covariance and Variogram Models
Water Resour. Res. 20(2), 251– 265.

Mandelbrot(1982)
The fractal geometry of nature
Earth Surf. Process. Landforms 8: 406-406.

Agradecimiento especial

Al estudiante de doctorado M. en C. Daniel Vázquez Ramírez, por su desinteresado apoyo en la conversión de esta presentación del curso de Powerpoint a Latex con Beamer.

Siguiente tema: Estimación Univariada