

# Un caso de estudio: Métodos de interpolación espacial en un área forestal del norte de Portugal

Workshop on Computational Intelligence Trends and Methodologies

Carmen Hernández

Grupo de Inteligencia Computacional (UPV/EHU) y Departamento Florestal (UTAD)

11 de diciembre de 2009

- Proyecto: PTDC/AGR-CFL/68186/2006

*Florestas mistas. Modelação, dinâmica e distribuição geográfica da produtividade e da fixação do carbono nos ecossistemas florestais mistos em Portugal.*

**Domingos Manuel Mendes Lopes y Helder Viana**

Departamento Florestal

Universidade de Trás-os-Montes e Alto Douro UTAD

**Manuel Graña, Miguel Angel Veganzones y Carmen Hernández**

Grupo de Inteligencia Computacional

Universidad del País Vasco UPV/EHU

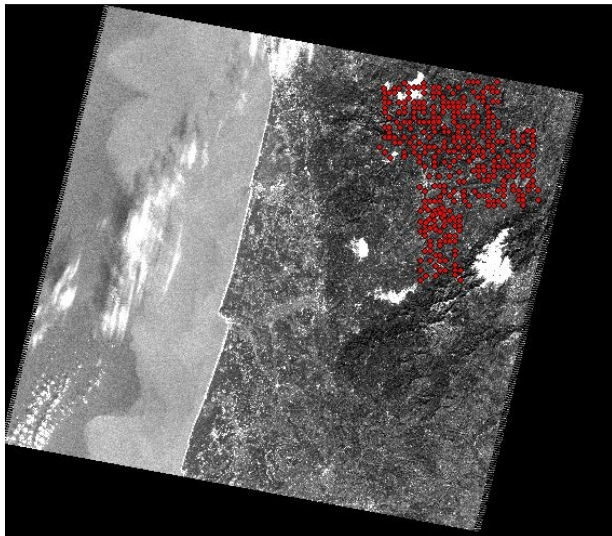
- 1 Objetivo
- 2 Descripción de los datos utilizados
- 3 Métodos de interpolación espacial
- 4 Resultados obtenidos
- 5 Validación de los métodos de interpolación
- 6 Conclusiones y Líneas Futuras
- 7 Referencias bibliográficas y Anexos

- El objetivo de esta presentación es la descripción y análisis de los primeros resultados obtenidos de la aplicación de técnicas de interpolación espacial en la estimación del volumen de biomasa de un área forestal del norte de Portugal a partir de los datos recogidos en diversas parcelas durante el año 2006.
  - La vegetación de este área es fundamentalmente matorral y árboles (*Pinus Pinaster*, Roble, Eucalipto y Mixto).
- Utilizaremos diversos métodos de interpolación espacial así como un índice de vegetación inferido a partir de una imagen de detección remota.

- Inicialmente, disponemos de información del volumen de biomasa recogida durante el año 2006 en 351 parcelas de la zona de estudio.
- No obstante, después de realizar una selección de los datos, obtenemos 280 localizaciones para las cuales conocemos:
  - Posición ( $x, y$ )
    - Projected coordinate system: Lisboa\_Hayford\_Gauss\_IGeoE
    - Geographic coordinate system: GCS\_Datum\_Lisboa\_Hayford
    - Map Projection: Transverse Mercator
  - Tipo de vegetación (matorrales y árboles)
  - Peso de matorral verde y seco:
    - Peso (**v**) de matorral verde recolectado por hectárea ( $ton/ha$ ).
    - Peso (**s**) de matorral seco recolectado por hectárea ( $ton/ha$ ).
  - Metros cúbicos (**m3**) de biomasa recolectada por hectárea ( $m^3/ha$ ).

- Las variables de peso (**v** y **s**) fueron cuantificadas en el campo y mediante la transformación de masa volumétrica, la cual mide el grado de concentración de masa en un determinado volumen, se obtuvo la variable metros cúbicos de biomasa (**m<sup>3</sup>**).

# Datos del terreno



- Estudio estadístico de los datos observados:
  - Estadística descriptiva de los datos:

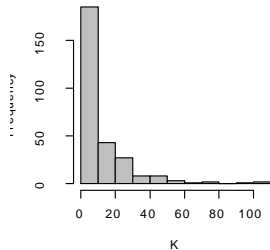
	Min	1Qu.	Median	Mean	3Qu.	Max	SD
v	0.08	2.277	5.805	11.89	15.79	108	16.18
s	0.03	0.91	2.32	4.76	6.31	43.2	6.47
m3	2.21	194.8	685.9	2375	2638	35100	4602.48

- Histogramas

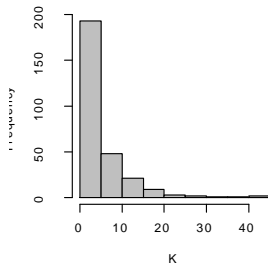


# Datos del terreno: Histogramas

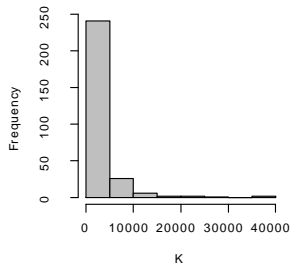
### Histogram (v)



### Histogram (s)



### Histogram (m3)



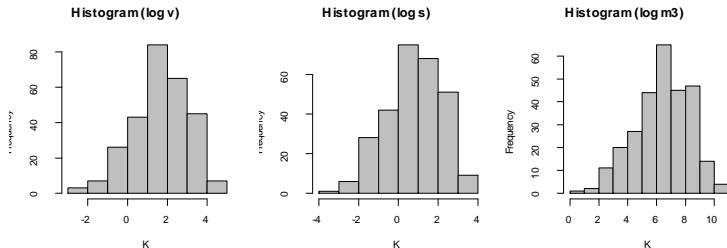
# Datos del terreno: Utilización de logaritmo

- Las distribuciones de los datos están claramente sesgadas hacia valores bajos de las variables y los valores más elevados presentan un importante rango de variación.
- Estudio estadístico de los datos observados (logaritmo de los datos):
  - Estadística descriptiva de los datos:

	Min	1Qu.	Median	Mean	3Qu.	Max	SD
logv	-1.1	0.36	0.76	0.73	1.20	2.03	0.61
logs	-1.52	-0.04	0.37	0.33	0.80	1.63	0.61
logm3	0.34	2.29	2.84	2.79	3.42	4.54	0.82

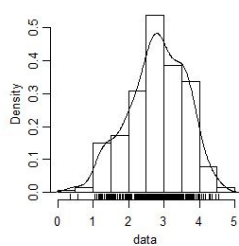
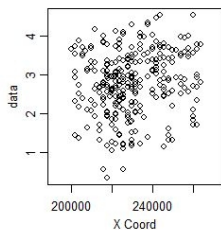
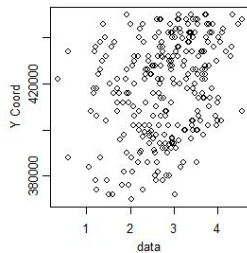
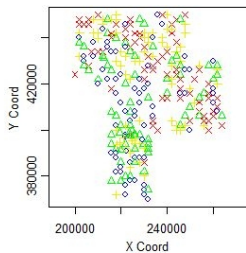
- Histogramas

# Datos del terreno: Histogramas

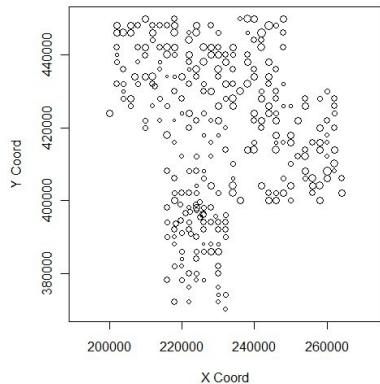


- La distribución logarítmica de los datos observados permite comprobar que estos datos presentan una distribución lognormal (normalización de los datos).

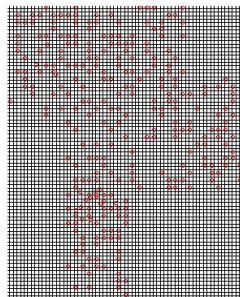
# Datos del terreno: Logaritmo



# Datos del terreno: Grid de datos



Interpolation Grid and Sample Points



- Imagen LandSat5 TM

- Nombre: L5204032\_03220061210\_B123457 (*U.S. Geological Survey*)
- Formato: GEOTIFF (7 imágenes, una por cada banda)
- Número de bandas: 7 bandas
- Píxel: 30 metros
- Data: 2006-12-10
- Map projection: UTM, Zone 29 North
- Datum: WGS-84
- Dimensiones: 8241 x 7221 x 6 bandas espectrales y una banda espectral (la sexta) de 4121 x 3611 píxeles, todas ellas en formato BSQ



# Transformación de la imagen

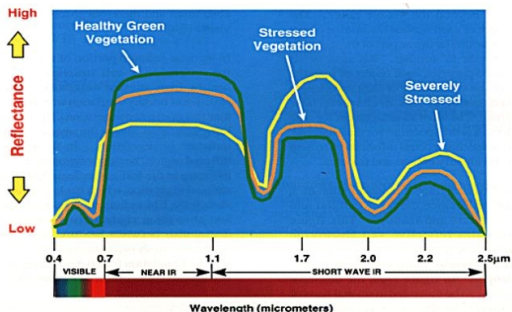
- Nombre: L5204032\_03220061210\_B123457Geo
- Formato: GEOTIFF
- Píxel: 30 metros
- Número de bandas: 6 bandas espectrales (eliminada la banda 6)
  - Map projection: UTM, Zone 29 North
  - Datum: WGS-84
- Dimensiones: 8241 x 7221 x 6 bandas espectrales en formato BSQ



- La teledetección (*Remote Sensing*, RS) se ha utilizado en muchas ocasiones en el seguimiento de la dinámica vegetal, ya que permite analizar amplias áreas con una gran frecuencia temporal.
- Observamos respuesta de la vegetación a la radiación en las regiones del espectro electromagnético del visible y del infrarrojo cercano.
  - La radiación visible es absorbida por la vegetación para llevar a cabo sus procesos fotosintéticos, mientras que en la región del infrarrojo cercano la radiación es reflejada debido a la estructura interna de las hojas [13].

# Índices de vegetación

- El comportamiento espectral de la vegetación (tanto natural como cultivada) es especialmente variable, dependiendo de factores como estado fenológico, forma y contenido en humedad.
  - Presenta una reducida reflectividad en las bandas visibles, con un máximo relativo en la región del verde ( $0.55\mu m$ ), debido al efecto absorbente de los pigmentos fotosintéticos de las hojas.



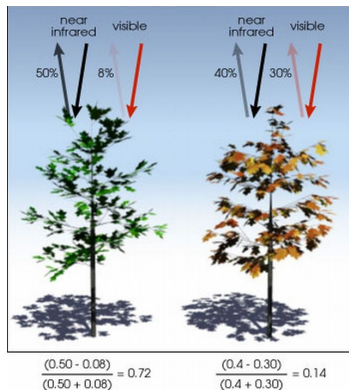
## Definition

El Índice Diferencial de Vegetación Normalizado (*Normalized Difference Vegetation Index*, NDVI) es el cociente normalizado entre los valores de reflectancia de la banda del rojo (baja reflexión) y el infrarrojo cercano (alta reflexión) y proporciona un realce tanto visual como numérico del estado de verdor de la vegetación [20].

$$NDVI = \frac{NIR - R}{NIR + R}$$

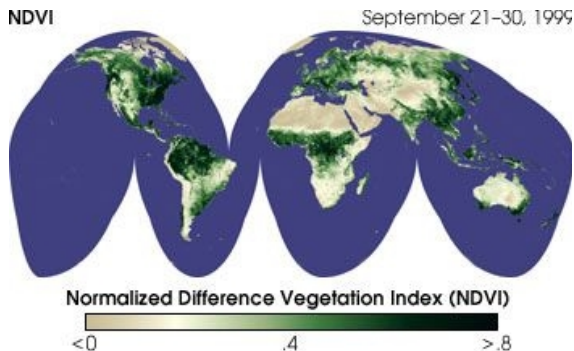
# Índices de vegetación

- Las hojas contienen clorofila que absorben las longitudes de onda del rojo y azul. Por eso, en vegetación no vigorosa, la clorofila no absorbe tanto y la hoja se torna un color amarillento (verde + rojo) [3].

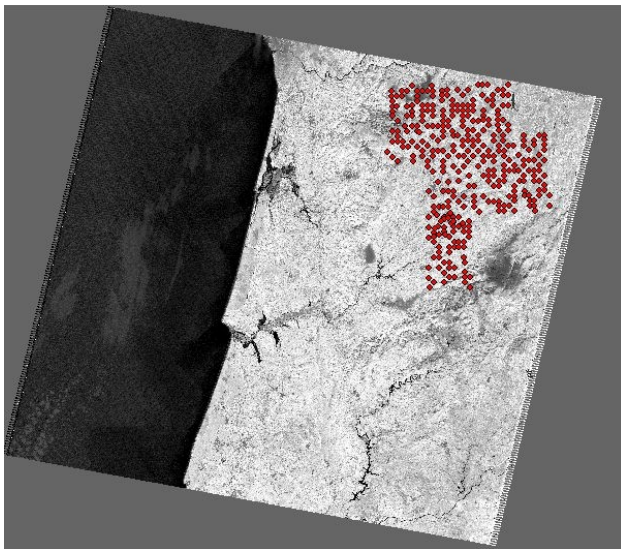


# Índice de vegetación NDVI

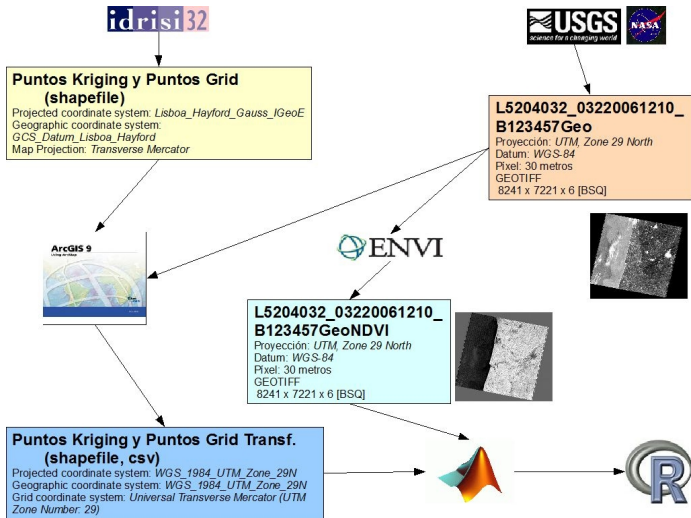
- Este índice se ha utilizado frecuentemente en la metodología de estimación de biomasa ya que existe una estrecha relación entre el índice NDVI y la biomasa vegetal.



# Índice de vegetación NDVI



# Pipeline: Adquisición y transformación de datos



- Existe en la literatura un gran número de métodos de interpolación espacial recogidos en tres categorías principales:
  - 1 Métodos no geoestadísticos (también denominados métodos determinísticos): generan superficies a partir de las observaciones puntuales en función de la similitud (*Inverse Distance Weighted*, IDW), del grado de suavizado (*Radial Basic Functions*, RBF), etc.
  - 2 Métodos geoestadísticos: utilizan propiedades estadísticas de las observaciones puntuales (p.e. la autocorrelación espacial).
  - 3 Métodos mixtos.

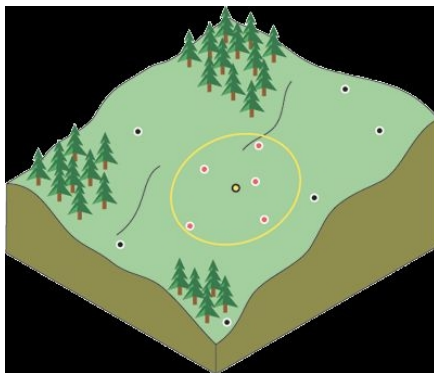


# Métodos de interpolación espacial

Non-geostatistical	Geostatistical		Combined method
	Univariate	Multivariate	
Nearest neighbours	Simple kriging	Universal kriging	Classification combined other interpolation methods
Triangular irregular network related interpolations	Ordinary kriging	SK with varying local means	Trend surface analysis combined with kriging
Natural neighbours	Block kriging	Kriging with an external drift	Lapse rate combined with kriging
Inverse distance weighting	Factorial kriging	Simple cokriging	Linear mixed model
Regression models	Dual kriging	Ordinary cokriging	Regression trees combined with kriging
Trend surface analysis	Indicator kriging	Standardised OCK	Residual maximum likelihood-empirical best linear unbiased predictor
Splines and local trend surfaces	Disjunctive kriging	Principal component kriging	Regression kriging
Thin plate splines	Model-based kriging	Colocated cokriging	Gradient plus inverse distance squared
Classification	Simulation	Kriging within strata	
Regression tree		Multivariate factorial kriging	
Fourier series		Indicator kriging	
Lapse rate		Indicator cokriging	
		Probability kriging	
		Simulation	

# Inverse Distance Weighted (IDW): Idea intuitiva

- Las observaciones más cercanas al punto de interpolación (punto de interés) deben tener una mayor influencia sobre los valores estimados en ese punto que las más distantes.



# Inverse Distance Weighted (IDW)

- Este método de la distancia inversa ponderada (*Inverse Distance Weigthed*, IDW) estima los valores de un atributo en una localización no muestreada a partir de una combinación lineal de los valores en los puntos muestreados donde los pesos o coeficientes vienen dados por una función inversa de la distancia del punto de interés a los puntos observados.

# Inverse Distance Weighted (IDW)

- El estimador IDW es [6]:

$$\hat{Z}(s_0) = \sum_{i=1}^n \frac{w_i(s_0)}{\sum_{k=1}^n w_k(s_0)} \cdot Z(s_i)$$

con

$$w_i(s_j) = \frac{1}{d(s_i, s_j)^p}$$

donde  $p$  es el parámetro *potencia* y  $d(s_i, s_j)$  es la distancia entre las localizaciones  $s_i$  y  $s_j$ .

- La teoría de las variables regionalizadas [16] y los métodos geoestadísticos basados en ella son una herramienta efectiva para el estudio de la distribución espacial de variables regionalizables y la interpolación en una malla regular.
- Matheron presentó esta técnica, alrededor de 1960, inspirado en los trabajos de D.G. Krige el cual fue, probablemente, el primero que utilizó la correlación espacial.

- **Kriging** o **krigeado** es un método estadístico de interpolación, también conocido como el *mejor estimador lineal insesgado*.
  - Se utiliza para la interpolación de datos de un campo aleatorio en una posición para la cual no se tiene información de las observaciones en las posiciones vecinas.
  - Aporta un entorno matemático muy sencillo y eficiente para la predicción de los valores del campo en posiciones no observadas a partir de un modelo de variograma.

- Los estudios geoestadísticos se utilizan en muchas aplicaciones tales como:
  - 1 Aplicaciones con datos hidrológicos [18], [12]
  - 2 Aplicaciones en minería [11], [19]
  - 3 Estudios de la calidad del aire [8], [14]
  - 4 Estudios y aplicaciones en la ciencia del suelo [22]
  - 5 Aplicaciones en biología [5], [17]
  - 6 Aplicaciones en economía [1]

- El objetivo del kriging es estimar el valor de una función aleatoria  $Z(s)$  en uno o más puntos no muestreados de una región  $D$  a partir de un conjunto de datos observados  $\{Z(s_1), \dots, Z(s_n)\}$  en los puntos  $s_1, \dots, s_n$ .
  - Asumimos que tenemos un proceso estacionario intrínseco y que queremos predecir el valor de la función en la localización  $s_0$ .
- El predictor de kriging,  $\hat{Z}(s_0)$ , es una combinación lineal de los puntos observados:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i \cdot Z(s_i)$$



## 1 Modelado de la variación espacial:

- 1 Generar un modelo de la estructura de la correlación (variograma) que exprese la variabilidad entre pares de observaciones como una función de su vector de separación.

## 2 Predicción espacial:

- 1 A partir del modelo obtenido, predecir los valores en las localizaciones no muestreadas.
  - La técnica estándar del kriging incorpora el variograma en el sistema de ecuaciones lineales a resolver y produce el mejor predictor lineal (*Best Linear Unbiased Predictor*, BLUP).

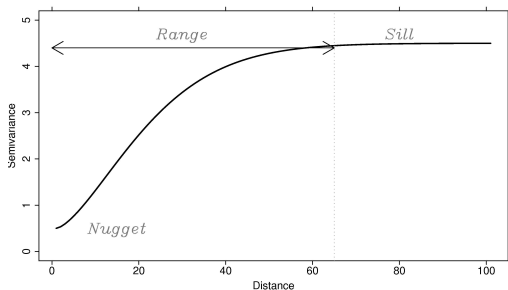
## Definition

El semivariograma  $\gamma(h)$  es un momento de segundo orden de una función aleatoria  $Z(s)$  [10], [21]:

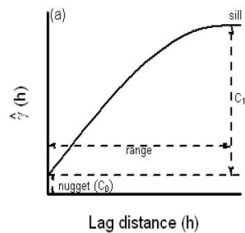
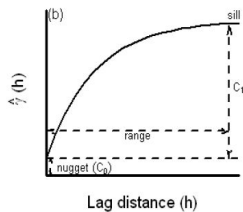
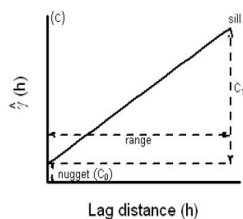
$$2\gamma(s_i, s_j) = \text{var} [Z(s_i) - Z(s_j)] = E \left[ (Z(s_i) - Z(s_j))^2 \right]$$

- A partir de los datos observados, generamos el **variograma empírico** o **experimental**.
- Posteriormente, debemos seleccionar un **variograma teórico** que se adecúe al variograma experimental ya que este último no es una función, en general, a partir de la cual se puedan realizar interpolaciones.
  - Este modelo describe el grado de dependencia espacial de un campo aleatorio espacial o de un proceso estocástico  $Z(s)$ .
  - En este paso debemos tener en cuenta los tres elementos más característicos del variograma teórico: el rango, la asíntota (*sill*) y el efecto pepita (*nugget*).

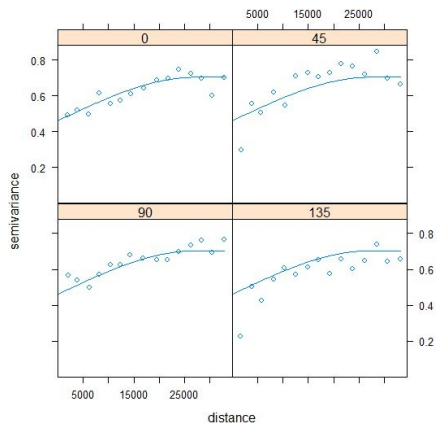
# Kriging: Variograma



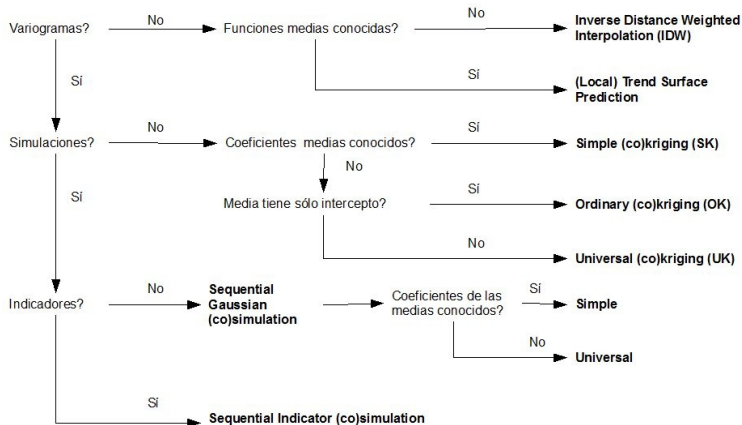
# Kriging: Tipos de variograma



# Kriging: Variograma empírico y variograma teórico



# Kriging: Tipos



- Existen varios métodos de kriging que trabajan bajo ciertas asunciones sobre la estructura de la media del modelo:

$$E [Z (s)] = \mu (s)$$

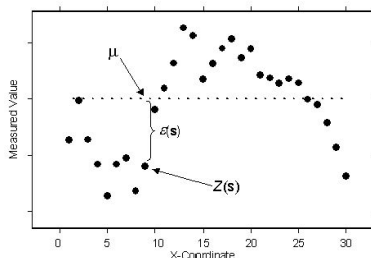
- Kriging Simple (*Simple Kriging*, SK): La media es una constante conocida.
- Kriging Ordinario (*Ordinary Kriging*, OK): La media es constante pero desconocida y debe ser estimada.
- Kriging Universal (*Universal Kriging*, UK): La media es desconocida pero es una combinación lineal de funciones conocidas de las localizaciones.



- Inicialmente, utilizaremos el **kriging ordinario** y el **universal** como métodos de interpolación espacial.
- Posteriormente, utilizaremos otro método de kriging denominado **kriging con factor externo** (*External Drift Kriging*, EDK) que utiliza el índice de vegetación NDVI.

# Ordinary Kriging

- Este método se conoce como el *método de kriging lineal con media estacionaria pero desconocida*<sup>1</sup> y depende solamente de las propiedades de segundo orden del campo aleatorio.



<sup>1</sup>Generalmente, el valor de la media es desconocido y, por lo tanto, no se puede utilizar el kriging simple. El kriging ordinario establece una condición adicional al sistema de ecuaciones del kriging simple a fin de controlar la media desconocida.

- Puede utilizarse cuando existe:
  - estacionariedad de segundo orden o
  - estacionariedad intrínseca

- El estimador de kriging ordinario (OK) es [10]:

$$Z_{OK}^*(s_0) = \sum_{i=1}^n \lambda_i^{OK}(s_0) \cdot Z(s_i) \quad \text{con} \quad \sum_{i=1}^n \lambda_i^{OK}(s_0) = 1$$

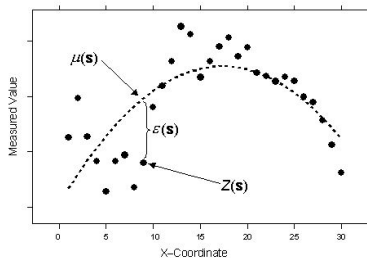
siendo  $s_0$  la localización de interés.

- La varianza del kriging ordinario es, por lo general, mayor que la varianza del kriging simple y puede ser expresada como:

$$\sigma_{OK}^2(s_0) = C(0) - \sum_{i=1}^n \lambda_i \cdot C(s_i - s_0) - \mu_{OK}(s_0)$$

con  $C(0) = \text{var}[Z(s)]$ .

- Este método de interpolación es un kriging con tendencia y es muy similar al kriging ordinario salvo que, en vez de tomar el vecindario del punto de interés, se realiza mediante un ajuste lineal o de orden mayor en las coordenadas de los datos.

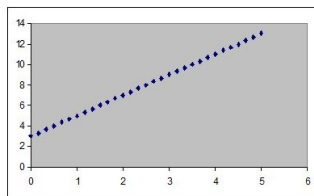


- Los datos espaciales  $Z(s)$  pueden ser vistos como un campo aleatorio con una componente de tendencia,  $m(s)$ , y una componente residual:

$$R(s) = Z(s) - m(s)$$

- El método de interpolación de kriging estima el residual en la posición de interés  $s_0$  como la suma pesada de las componentes residuales en los puntos vecinos.
- Los pesos kriging,  $\lambda_i(s)$ , se obtienen a partir de la función de covarianza o semivariograma la cual caracteriza a la componente residual.

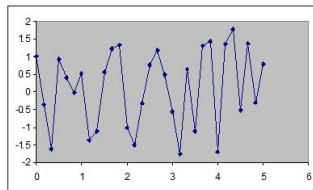
# Universal Kriging



$m$

+

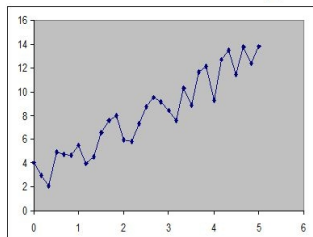
=



$R$

$Z(x)$

$Z$





- El estimador de kriging universal (UK) es:

$$Z_{UK}^*(s_0) = \sum_{i=1}^n \lambda_i^{UK}(s_0) \cdot Z(s_i)$$

donde

$$\mu_{UK}(s_0) = \sum_{i=1}^n \lambda_i^{UK}(s_0) \cdot m(s_i) = \sum_{k=1}^p a_k \cdot \sum_{i=1}^n \lambda_i^{UK}(s_0) \cdot f_k(s_i)$$

con la condición

$$\sum_{i=1}^n \lambda_i^{UK}(s_0) \cdot f_k(s_i) = f_k(s_0) \quad k = 0, \dots, p$$

donde  $f_k$  son funciones conocidas de la posición y  $a_k$  son parámetros desconocidos a estimar.

- La varianza del kriging universal es:

$$\sigma_{UK}^2(s_0) = C_R(0) + \sum_{i,j=1}^n \lambda_i \cdot \lambda_j \cdot C_R(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i \cdot C_R(s_i - s_0)$$

siendo  $C_R(\cdot)$  la covarianza del residual.

- En el espacio bidimensional se suelen utilizar polinomios de bajo orden.
- Por ejemplo, dada la posición de interés  $s_0 = (x, y)$ , la media o tendencia viene dada por el polinomio:

$$\mu_{UK}(s_0) = m(x, y) = a_0 + a_1x + a_2y$$

- El método del kriging con factor externo (*External Drift Kriging*, EDK) permite utilizar una variable secundaria espacial (exhaustiva) que muestra la tendencia espacial de la primera variable [7].
  - Cuando una de las variables es precisa pero poco muestreada (primera variable,  $Z$ ) y la otra es más imprecisa pero densamente muestreada (segunda variable,  $q$ ), podemos utilizar ambas fuentes de información para el estudio del fenómeno en cuestión.

$$\mu_{EDK}(s_0) = a + b \cdot q(s)$$

- El estimador del kriging con factor externo es:

$$Z_{EDK}^*(s_0) = \sum_{i=1}^n \lambda_i^{EDK}(s_0) \cdot Z(s_i)$$

con

$$\sum_{i=1}^n \lambda_i^{EDK}(s_0) \cdot q_k(s_i) = q_k(s_0) \quad k = 1, \dots, p$$

donde  $Z$  son los datos espaciales (valores de la primera variable),  $q_k$  son las  $p$  variables predictoras y

$$\mu_{EDK}(s_0) = \sum_{k=0}^p \mu_k(s_0) \cdot q_k(s_0)$$

siendo  $\mu_k(s_0)$  los parámetros de Lagrange.

- La varianza del estimador de kriging con factor externo es:

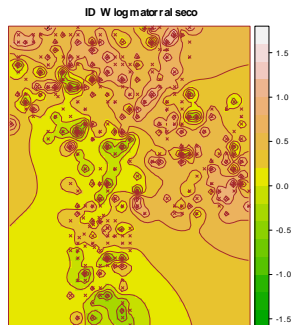
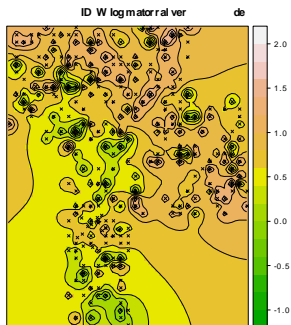
$$\sigma_{EDK}^2(s_0) = C(0) - \sum_{i=1}^n \lambda_i^{EDK}(s_0) \cdot C(s_i - s_0) + \sum_{k=0}^p \mu_k(s_0) \cdot q_k(s_0)$$

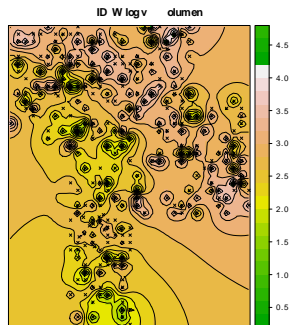
donde  $C(0)$  es la covarianza estacionaria residual,  $\mu_k(s_0)$  son los parámetros de Lagrange y  $q_k(s_0)$  son las funciones de tendencia.

- Si dos o más variables se encuentran correlacionadas, es posible aprovechar la información espacial no solamente de una de ellas sino de ambas.
- Para ello, se obtiene un semivariograma cruzado donde la varianza representada se obtiene a partir de puntos de una variable con respecto a la otra.
  - Si el variograma cruzado da una estructura interpretable, esta información puede usarse para la predicción de la primera variable en una técnica que se conoce como **cokriging**.

- La información que tiene una variable sobre otra es siempre menor que la que tiene una variable sobre sí misma.
  - Por ello, el cokriging rara vez mejora la predicción del kriging salvo que una de las dos variables haya sido muestreada con menor intensidad que la otra.
  - En este caso el cokring puede ser muy útil.



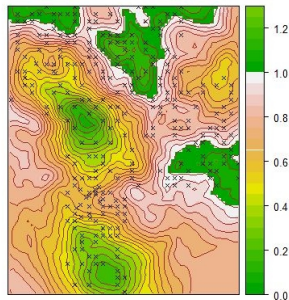




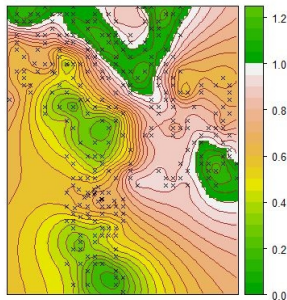
# Ordinary Kriging: logv

Predictions: Ordinary Kriging log matorral verde

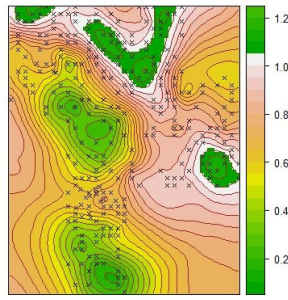
Linear



Exponential

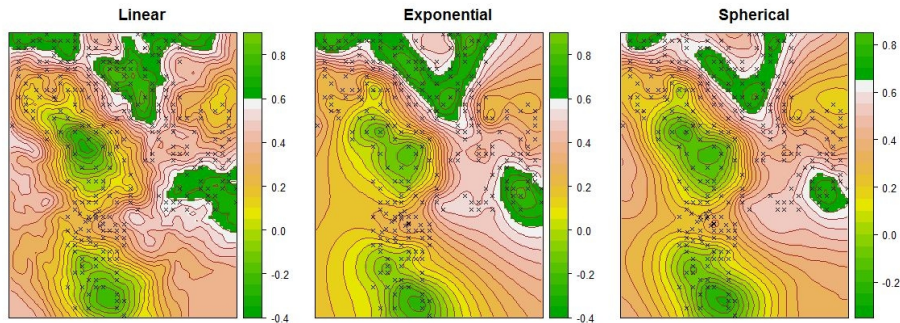


Spherical



# Ordinary Kriging: logs

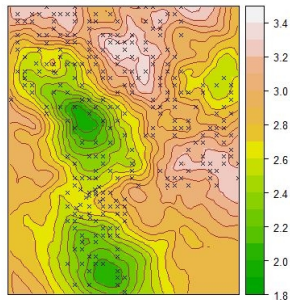
Predictions: Ordinary Kriging log matorral seco



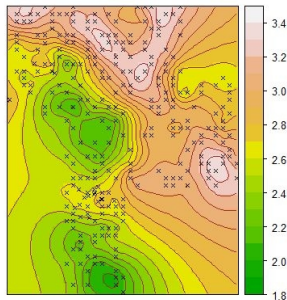
# Ordinary Kriging: logm3

Predictions: Ordinary Kriging log volumen

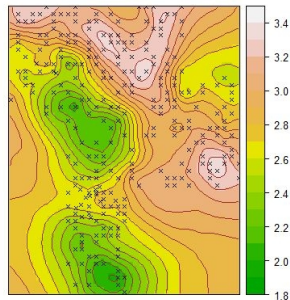
Linear



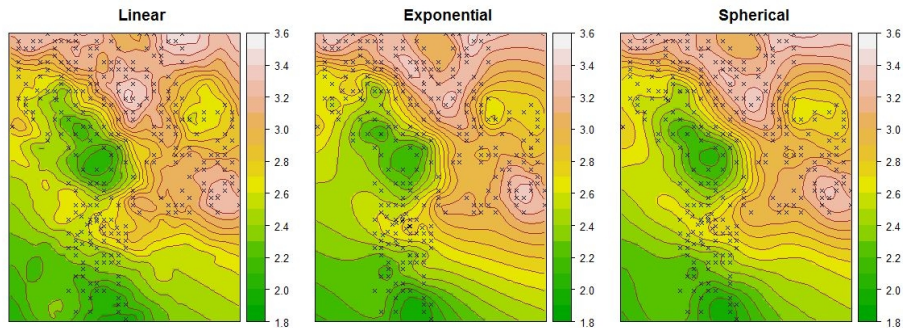
Exponential



Spherical

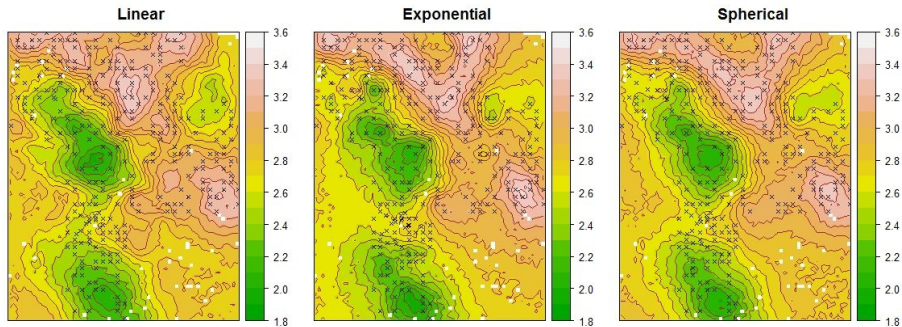


## Predictions: Universal Kriging log volumen



# External Drift Kriging (NDVI): logm3

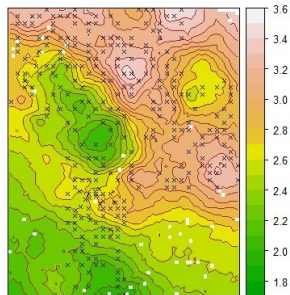
Predictions: External Drift Kriging (NDVI) log volumen



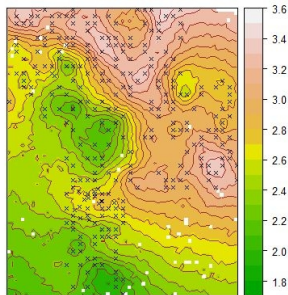
# External Drift Kriging (NDVI+x+y): logm3

Predictions: External Drift Kriging (NDVI+x+y) log volumen

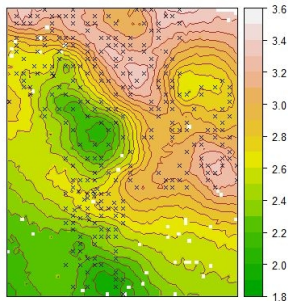
Linear



Exponential

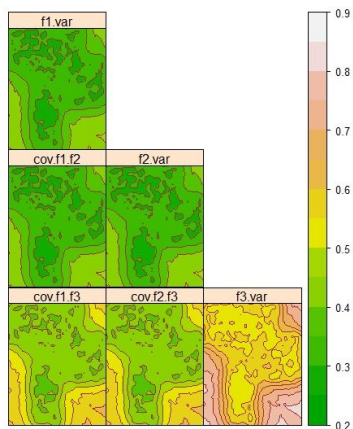


Spherical





# Cokriging: logv, logs, logm3 (exp)



# Validación del método de interpolación espacial

- Existen varios criterios a la hora de evaluar la "bondad" de un método de estimación.
- La selección del "mejor" método depende de las siguientes medidas:
  - Minimizar el error medio absoluto global o la desviación estándar del error global.
  - Minimizar el error medio absoluto local o la desviación estándar local de una subárea específica de interés.
  - Evitar la sobreestimación o subestimación en subáreas específicas.

- Para la validación de los métodos de interpolación espacial precisamos de un conjunto de datos de control (**conjunto de validación**).
- No obstante, en muchos casos, no disponemos de este conjunto de validación y, por ello, podemos:
  - calcular ciertas medidas del error de la predicción para cada método
  - establecer diferentes medidas de calidad entre los diferentes métodos de interpolación
  - utilizar la validación cruzada (*Cross-validation Kriging*, CVK) como medida de la evaluación de la calidad del método de interpolación.

- La validación cruzada compara la efectividad y la precisión de los métodos de estimación y evalúa el comportamiento del modelo.
- Utilizaremos el método de *Cross-validation Kriging* [7] con un *folder* (*Leave-One-Out*, LOO), 5 y 10 *folders*.
- Este tipo de validación implica:
  - La estimación en todas las localizaciones de muestreo
  - La comparación de las estimaciones con los valores observados en dichas localizaciones

# Cross-Validation Kriging: OK

	ME	MAE	MSE	RMSE	MSRE	RVar	EF	E
Linear, 1	0,0015	0,6013	0,5676	0,7534	1,0445	0,1849	0,1485	14,8480
Linear, 5	0,0061	0,5987	0,5655	0,7520	1,0280	0,1829	0,1517	15,1655
Linear, 10	0,0024	0,6022	0,5713	0,7559	1,0454	0,1860	0,1429	14,2889
Exponential, 1	0,0018	0,6019	0,5682	0,7538	1,0386	0,1884	0,1476	14,7597
Exponential, 5	0,0001	0,5995	0,5688	0,7542	1,0257	0,1814	0,1467	14,6728
Exponential, 10	0,0016	0,6072	0,5781	0,7604	1,0509	0,1897	0,1327	13,2695
Spherical, 1	0,0021	0,6019	0,5684	0,7539	1,0426	0,1894	0,1473	14,7261
Spherical, 5	0,0047	0,5975	0,5660	0,7524	1,0249	0,1869	0,1509	15,0863
Spherical, 10	-0,0034	0,5957	0,5595	0,7480	1,0209	0,1814	0,1606	16,0606

	MSNE	RMNSE	MDE	COP1	COP2	CPR1	CPR2	EV
Linear, 1	1,0445	1,0220	2,8363	0,3877	0,3877	-0,0459	-0,0459	0,1485
Linear, 5	1,0280	1,0139	2,8363	0,3912	0,3912	-0,0396	-0,0396	0,1517
Linear, 10	1,0454	1,0224	2,8363	0,3813	0,3813	-0,0539	-0,0539	0,1429
Exponential, 1	1,0386	1,0191	2,8363	0,3870	0,3870	-0,0509	-0,0509	0,1476
Exponential, 5	1,0257	1,0128	2,8363	0,3852	0,3852	-0,0440	-0,0440	0,1467
Exponential, 10	1,0509	1,0251	2,8363	0,3701	0,3701	-0,0703	-0,0703	0,1327
Spherical, 1	1,0426	1,0211	2,8363	0,3868	0,3868	-0,0524	-0,0524	0,1473
Spherical, 5	1,0249	1,0124	2,8363	0,3907	0,3907	-0,0451	-0,0451	0,1509
Spherical, 10	1,0209	1,0104	2,8363	0,4015	0,4015	-0,0267	-0,0267	0,1606

# Cross-Validation Kriging: UK

	ME	MAE	MSE	RMSE	MSRE	RVar	EF	E
Linear, 1	-0,0022	0,6050	0,5757	0,7588	1,0531	0,2146	0,1364	13,6359
Linear, 5	0,0115	0,6147	0,5915	0,7691	1,0682	0,2162	0,1127	11,2653
Linear, 10	0,0045	0,6153	0,6003	0,7748	1,0885	0,2228	0,0995	9,9454
Exponential, 1	-0,0052	0,6043	0,5745	0,7579	1,0407	0,2132	0,1382	13,8207
Exponential, 5	-0,0092	0,6044	0,5836	0,7640	1,0459	0,2189	0,1245	12,4475
Exponential, 10	-0,0057	0,5989	0,5650	0,7516	1,0179	0,2115	0,1525	15,2470
Spherical, 1	-0,0041	0,6041	0,5746	0,7580	1,0462	0,2137	0,1381	13,8071
Spherical, 5	-0,0034	0,6026	0,5857	0,7653	1,0552	0,2174	0,1214	12,1418
Spherical, 10	0,0008	0,6067	0,5777	0,7601	1,0452	0,2156	0,1334	13,3377

	MSNE	RMNSE	MDE	COP1	COP2	CPR1	CPR2	EV
Linear, 1	1,0531	1,0262	2,8363	0,3788	0,3788	-0,0909	-0,0909	0,1364
Linear, 5	1,0682	1,0335	2,8363	0,3538	0,3538	-0,1180	-0,1180	0,1128
Linear, 10	1,0885	1,0433	2,8363	0,3414	0,3414	-0,1376	-0,1376	0,0995
Exponential, 1	1,0407	1,0201	2,8363	0,3806	0,3806	-0,0875	-0,0875	0,1382
Exponential, 5	1,0459	1,0227	2,8363	0,3671	0,3671	-0,1077	-0,1077	0,1246
Exponential, 10	1,0179	1,0089	2,8363	0,3958	0,3958	-0,0697	-0,0697	0,1525
Spherical, 1	1,0462	1,0228	2,8363	0,3805	0,3805	-0,0880	-0,0880	0,1381
Spherical, 5	1,0552	1,0272	2,8363	0,3633	0,3633	-0,1098	-0,1098	0,1214
Spherical, 10	1,0452	1,0223	2,8363	0,3758	0,3758	-0,0951	-0,0951	0,1334

# Cross-Validation Kriging: EDK NDVI

	ME	MAE	MSE	RMSE	MSRE	RVar	EF	E
Linear, 1	0,0023	0,5847	0,5432	0,7370	0,9729	0,2545	0,1851	18,5086
Linear, 5	0,0036	0,5853	0,5513	0,7425	0,9775	0,2607	0,1730	17,3001
Linear, 10	0,0098	0,5875	0,5438	0,7374	0,9694	0,2577	0,1842	18,4169
Exponential, 1	0,0016	0,5841	0,5437	0,7374	0,9644	0,2556	0,1843	18,4325
Exponential, 5	0,0127	0,5931	0,5577	0,7468	0,9755	0,2432	0,1633	16,3300
Exponential, 10	0,0048	0,5867	0,5436	0,7373	0,9586	0,2607	0,1846	18,4569
Spherical, 1	0,0021	0,5838	0,5429	0,7368	0,9681	0,2555	0,1856	18,5579
Spherical, 5	0,0019	0,5880	0,5454	0,7385	0,9593	0,2410	0,1818	18,1792
Spherical, 10	0,0054	0,5880	0,5459	0,7388	0,9672	0,2557	0,1811	18,1095

	MSNE	RMNSE	MDE	COP1	COP2	CPR1	CPR2	EV
Linear, 1	0,9729	0,9864	2,8363	0,4357	0,4357	-0,0762	-0,0762	0,1851
Linear, 5	0,9775	0,9887	2,8363	0,4247	0,4247	-0,0944	-0,0944	0,1730
Linear, 10	0,9694	0,9846	2,8363	0,4354	0,4354	-0,0800	-0,0800	0,1843
Exponential, 1	0,9644	0,9820	2,8363	0,4351	0,4351	-0,0781	-0,0781	0,1843
Exponential, 5	0,9755	0,9877	2,8363	0,4124	0,4124	-0,0883	-0,0883	0,1635
Exponential, 10	0,9586	0,9791	2,8363	0,4361	0,4361	-0,0825	-0,0825	0,1846
Spherical, 1	0,9681	0,9839	2,8363	0,4363	0,4363	-0,0767	-0,0767	0,1856
Spherical, 5	0,9593	0,9795	2,8363	0,4306	0,4306	-0,0667	-0,0667	0,1818
Spherical, 10	0,9672	0,9835	2,8363	0,4319	0,4319	-0,0815	-0,0815	0,1811

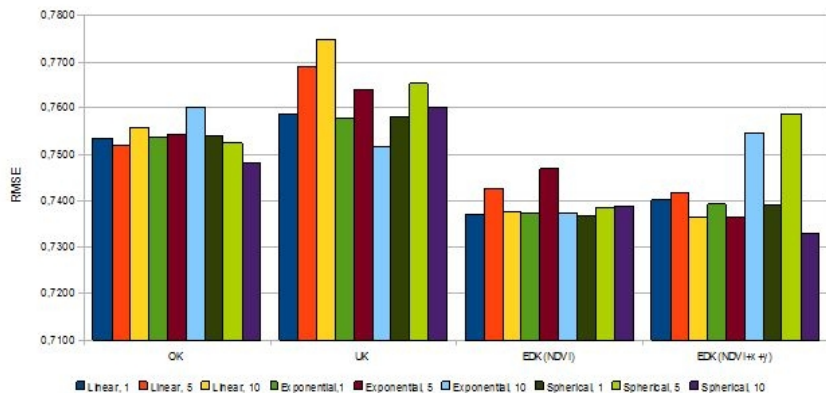
# Cross-Validation Kriging: EDK NDVI+x+y

	ME	MAE	MSE	RMSE	MSRE	RVar	EF	E
Linear, 1	-0,0040	0,5861	0,5477	0,7400	0,9806	0,2889	0,1784	17,8416
Linear, 5	0,0036	0,5806	0,5502	0,7418	0,9718	0,3070	0,1746	17,4603
Linear, 10	-0,0057	0,5821	0,5422	0,7363	0,9624	0,2887	0,1866	18,6638
Exponential, 1	-0,0077	0,5853	0,5466	0,7393	0,9686	0,2850	0,1800	18,0000
Exponential, 5	-0,0068	0,5790	0,5422	0,7363	0,9441	0,2800	0,1866	18,6611
Exponential, 10	-0,0073	0,5938	0,5693	0,7545	0,9996	0,2871	0,1460	14,5976
Spherical, 1	-0,0063	0,5848	0,5462	0,7390	0,9732	0,2860	0,1806	18,0631
Spherical, 5	0,0086	0,5996	0,5757	0,7588	1,0089	0,2733	0,1363	13,6348
Spherical, 10	-0,0017	0,5767	0,5371	0,7329	0,9523	0,2920	0,1943	19,4290

	MSNE	RMNSE	MDE	COP1	COP2	CPR1	CPR2	EV
Linear, 1	0,9806	0,9903	2,8363	0,4347	0,4347	-0,1134	-0,1134	0,1784
Linear, 5	0,9718	0,9858	2,8363	0,4346	0,4346	-0,1315	-0,1315	0,1746
Linear, 10	0,9624	0,9810	2,8363	0,4424	0,4424	-0,1053	-0,1053	0,1867
Exponential, 1	0,9686	0,9842	2,8363	0,4356	0,4356	-0,1085	-0,1085	0,1801
Exponential, 5	0,9441	0,9716	2,8363	0,4410	0,4410	-0,0978	-0,0978	0,1867
Exponential, 10	0,9996	0,9998	2,8363	0,4042	0,4042	-0,1424	-0,1424	0,1461
Spherical, 1	0,9732	0,9865	2,8363	0,4363	0,4363	-0,1088	-0,1088	0,1807
Spherical, 5	1,0089	1,0044	2,8363	0,3919	0,3919	-0,1408	-0,1408	0,1365
Spherical, 10	0,9523	0,9759	2,8363	0,4500	0,4500	-0,1007	-0,1007	0,1943



# Cross-Validation Kriging: RMSE







- Los resultados obtenidos muestran que los métodos de kriging son útiles en tareas de estimación de datos del área forestal tales como la biomasa, combustibles forestales, etc.

- Consideramos de interés la realización de estudios más precisos y detallados de este tipo de técnicas para el problema de la estimación de la biomasa y otros problemas de este área:
  - Utilización de herramientas computacionales para el ajuste del variograma experimental al variograma teórico: SVM
  - Utilización de nuevas variables auxiliares como factor externo de ayuda al suavizado del kriging.
    - Índice de superficie foliar (*Leaf Area Index*, LAI)
    - Altura de la localización
    - Tipo de vegetación existente en la zona de estudio
  - Utilización y análisis comparativo de otros métodos de interpolación.
  - Análisis más detallado de las medidas del error y, en general, de las técnicas de validación del kriging.





Gracias por su atención



# Referencias bibliográficas I

-  S. BASU AND T. THIBODEAU.  
Analysis of spatial correlation in house prices.  
*Journal of Real Estate Finance and Economics* **17**, 61–85 (1998).
-  R. BRUNO AND G. RASPA.  
“Geostatistics Tróia '92”, chapter Towards a direct structural analysis of an IRF-K, page 49–59.  
Kluwer Academic Publishers, Dordrecht (1993).
-  E. CHUVIECO.  
“Fundamentos de Teledetección Espacial”.  
Madrid, 3 edition (2000).
-  N. CRESSIE.  
“Statistics for Spatial Data”.  
Wiley-Interscience, 1 edition (1993).

# Referencias bibliográficas II

-  M.D. ECKER AND A.E. GELFAND.  
Bayesian modeling and inference for geometrically anisotropic spatial data.  
*Mathematical Geology* **31**(1), 67–83 (1999).
-  J.G. FLORES.  
“Modeling the spatial variability of forest fuel arrays”.  
PhD thesis, Dept. of Forest Sciences. Colorado State University (2001).
-  PIERRE GOOVAERTS.  
“Geostatistics for Natural Resources Evaluation”.  
Oxford University Press, USA (September 1997).
-  PETER GUTTORP, WENDY MEIRING, AND PAUL D. SAMPSON.  
A space-time analysis of ground-level ozone data.  
*Environmetrics* **5**, 241–254 (1994).



TOMISLAV HENGL.

A practical guide to geostatistical mapping of environmental variables.  
Technical Report, EUR 22904 EN (2007).



E.H. ISAAKS AND R.M. SRIVASTAVA.

“An introduction to applied geostatistics” .  
Oxford University Press, Nueva York (1989).



D. KELKER AND W. LANGENBURG.

Ellipsoid estimation in coal reflectance anisotropy.  
*Mathematical Geology* **29**(2), 185–198 (1997).



P.K. KITANIDIS.

Parameter uncertainty in estimation of spatial functions: Bayesian analysis.

*Water Resources Research* **22**(4), 499–507 (1986).

## Referencias bibliográficas IV



E.B. KNIPLING.

Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation.

*Remote Sensing of Environment* **1**, 155-159 (1970).



P. KRAJEWSKI, A. MOLIŃSKA, AND K. MOLIŃSKI.

Elliptical anisotropy in practice - a study of air monitoring data.

*Environmetrics* **7**, 291-298 (1996).



JIN LI AND ANDREW D. HEAP.

A review of spatial interpolation methods for environmental scientists.

Technical Report, GeoCat 68229, 2008/23 (2008).



G. MATHERON.

La théorie des variables régionalisées et ses applications.

*Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau Fascicule* **5**, 212 p. (1970).





G. RITVO, J.B. DIXON, W.H. NEILL, T.M. SAMOCHA, AND A.L. LAWRENCE.

The effect of controlled soil sulfur concentration on growth and survival of *Litopenaeus vannamei*.

*Journal of the World Aquaculture Society* **31**, 381-389 (2000).



SHAHROKH ROUHANI AND DONALD MYERS.

Problems in space-time kriging of geohydrological data.

*Mathematical Geology* **22**(5), 611-623 (July 1990).



J.H. SCHUENEMEYER AND H.C. POWER.

Uncertainty estimation for resource assessment: an application to coal.

*Mathematical Geology* **32**(5), 521-541 (2000).



COMPTON J. TUCKER.

Red and photographic infrared linear combinations for monitoring vegetation.

*Remote Sensing of Environment* **8**(2), 127–150 (May 1979).



H. WACKERNAGEL.

“Multivariate Geostatistics”.

Springer (2003).



R. WEBSTER AND M.A. OLIVER.

Sample adequately to variograms of soil properties.

*J. Soil Sci* **43**, 177–192 (1992).

## Definition

La **región** se refiere al espacio en el cual existe y se estudia el fenómeno natural.

## Definition

Una **localización** es el punto de una región en la cual se define una variable aleatoria regionalizada.

## Definition

El **soporte geométrico** viene determinado por el elemento físico sobre el cual se realiza la determinación de la variable aleatoria regionalizada. Esto es, la muestra unitaria sobre la cual estudiaremos el atributo de interés.

## Definition

Un **dato espacial** es un objeto o entidad que resulta de una abstracción del espacio geográfico real y que puede ser un rasgo o característica natural, una obra humana o alguna abstracción numérica derivada del tratamiento de cifras relacionadas con tal objeto o entidad.

- Su característica distintiva es la referencia espacial en una, dos o tres dimensiones generalmente.
- Se considerarán sinónimos de dato espacial los siguientes términos: dato geográfico, dato geoespacial, entidad geográfica, objeto espacial, rasgo geográfico, características geográficas y demás que correspondan con esta definición.

## Definition

Una muestra o realización de una función aleatoria  $Z(s)$  es una función espacial discreta o una **variable regionalizada** que puede ser vista como una variable estructurada, cuando se tiene en cuenta únicamente su comportamiento a gran escala, y errática o aleatoria cuando tenemos en cuenta el comportamiento a pequeña escala.

- Esto es, según el modelado espacial, una variable regionalizada posee dos componentes (aparentemente contradictorias):
  - Una componente estructurada, a nivel global, que refleja las características generales del fenómeno y puede ser caracterizada mediante una función determinista.
  - Una componente aleatoria, a nivel local, asociada a las variaciones erráticas e impredecibles de la variable.

- La representación de la variación espacial de una variable regionalizada viene dada por la suma de tres componentes:

$$Z(s) = \mu + \varepsilon'(s) + \varepsilon''(s)$$

donde:

- $Z(s)$  es el valor de la variable en la posición  $s$
- $\mu$  representa la función determinista (la media o cualquier tendencia espacial)
- $\varepsilon'(s)$  es la componente correlacionada espacialmente pero localmente variable
- $\varepsilon''(s)$  es un término de error residual (generalmente, normal), o ruido aleatorio espacialmente independiente con una media nula y varianza  $\sigma^2(s)$

# Momentos de una variable regionalizada

- Los momentos o estimadores de una variable regionalizada son parámetros de resumen que caracterizan a las poblaciones estadísticas.
- Los momentos se utilizan en la inferencia indirecta de propiedades de esa población.
- El momento más usado es la media, el parámetro que describe la tendencia central de una población.



## Definition

El momento de primer orden de una función aleatoria  $Z(s)$  conocido como **valor medio** o **media** se define como:

$$E[Z(s)] = m(s)$$

- Es insuficiente para caracterizar enteramente a una población ya que la media no proporciona información sobre el número total de individuos ni sobre las variaciones entre ellos.

# Segundo momento de una variable regionalizada: Varianza, covarianza y variograma

## Definition

Los momentos de segundo orden de una función aleatoria  $Z(s)$  son:

- 1 La **varianza** de  $Z(s)$  que se define como:

$$\sigma^2(s) = \text{var} [Z(s)] = E \left[ (Z(s) - m(s))^2 \right]$$

- 2 La **covarianza** de  $Z(s)$ , también conocida como *covariograma*, definida como:

$$C(s_i, s_j) = \text{cov} [Z(s_i), Z(s_j)] = E [(Z(s_i) - m(s)) (Z(s_j) - m(s))]$$

## Segundo momento de una variable regionalizada: Varianza, covarianza y variograma

### Definition (cont.)

- 1 El **variograma** de  $Z(s)$ , también conocido como función de semivarianza o correlograma, definido como:

$$2\gamma(s_i, s_j) = \text{var} [Z(s_i) - Z(s_j)] = E \left[ (Z(s_i) - Z(s_j))^2 \right]$$

A la función  $\gamma(\cdot)$  se le denomina **semivariograma** [10].

- Se dice que una función aleatoria  $Z(s)$  es **estrictamente estacionaria**<sup>2</sup> si la función de distribución de probabilidades de las variables aleatorias regionalizadas  $Z(s_i)$  es independiente de la localización  $s_i$ , o lo que es lo mismo es invariante a cualquier traslación respecto a un vector  $h$ .
- Requiere que los momentos de distinto orden para cada variable aleatoria regionalizada sean completamente independientes de la localización.

---

<sup>2</sup>A esta condición también se le denomina estacionariedad fuerte.

## Definition

El proceso  $\{Z(s) : s \in D\}$  se dice que es **estrictamente estacionario** (estacionariedad fuerte) dentro de  $D$ , si sus distribuciones finitas dimensionales son invariantes bajo la traslación  $h$ . Esto es, si las familias  $\{Z(s_1), \dots, Z(s_n)\}$  y  $\{Z(s_1 + h), \dots, Z(s_n + h)\}$  tienen la misma función de distribución conjunta multivariada para todo  $h \in D$ :

$$F_{Z(s_1+h), \dots, Z(s_n+h)}(z_1, \dots, z_n) = F_{Z(s_1), \dots, Z(s_n)}(z_1, \dots, z_n) \quad \forall s_i, s_i + h \in D$$

- Resulta práctico limitar la hipótesis de estacionariedad a los primeros momentos ya que la estacionariedad estricta es una condición muy fuerte y poco habitual que establece que las distribuciones de probabilidad conjunta permanezcan invariables ante una traslación.

- Una condición menos exigente que la estacionariedad estricta es la estacionariedad de segundo orden, más frecuente en la práctica, que conlleva que la esperanza sea constante y que la función de covarianza sea invariante a la traslación.
  - Es decir, una función aleatoria es estacionaria de segundo orden si sus momentos de primer y de segundo orden no dependen de la posición.

## Definition

El proceso  $\{Z(s) : s \in D\}$  se dice **estacionario de segundo orden** en el campo o dominio  $D$ , si

$$E[Z(s)] = m \quad \forall s \in D$$

$$C(s, s+h) = \text{cov}[Z(s), Z(s+h)] = \text{cov}[h]$$

para todo  $s, s+h \in D$ , asumiendo que  $\text{var}[Z(s)] = \sigma^2 < \infty, \forall s \in D$ .

- La función de covarianza de un proceso estacionario se puede expresar en función del vector de diferencia o de traslación entre dos localizaciones.
  - Es decir, la función de covarianza  $cov [Z (s) , Z (s + h)]$  depende únicamente de  $h$ .



# Estacionariedad de segundo orden

- Esta hipótesis requiere estacionariedad sólo para la media y para la función de covarianza de la variable aleatoria regionalizada.
- La segunda condición implica, estacionariedad de la varianza y del variograma.
- La función de covarianza  $C(\cdot)$ , o covariograma, se puede especificar también en términos del variograma:

$$2\gamma(h) = 2\gamma(s, s+h) = E \left[ (Z(s+h) - Z(s))^2 \right]$$

- La estacionariedad de segundo orden, también se conoce como estacionariedad débil, o como proceso con estructura de covarianza estacionaria.

- Cuando la función aleatoria no es estacionaria pero las diferencias  $Z(s+h) - Z(s)$  sí son estacionarias de segundo orden, decimos que el proceso tiene estacionariedad intrínseca.
  - Es decir, la varianza de las diferencias entre las variables en dos puntos depende únicamente del vector que los separa  $h$ .

## Definition

El proceso  $\{Z(s) : s \in D\}$  se dice **íntrinsecamente estacionario** dentro de  $D$ , si cumple que:

$$E[Z(s+h) - Z(s)] = m \quad \forall s$$

y la varianza de la diferencia es:

$$\text{var}[Z(s+h) - Z(s)] = 2\gamma(h) \quad \forall s$$

donde  $2\gamma(h)$  es el **variograma** del proceso estocástico.

- Esta condición es más débil que la estacionariedad de segundo orden y se emplea habitualmente en el modelado geoestadístico.
  - Esta condición se encuentra con bastante frecuencia en la naturaleza, pues existen muchos procesos que no tienen varianza finita y sin embargo, poseen una función variograma finita.
- También se dice que, cuando se cumple esta condición, la función aleatoria  $Z(s)$  es homogénea.
- La estacionaridad de segundo orden, siempre implica la condición intrínseca (homogeneidad), sin embargo la relación inversa no siempre se cumple.

## Definition

Dada una variable regionalizada  $Z(s)$  que cumpla la hipótesis intrínseca entonces existe la función semivarianza o semivariograma teórico y se define como:

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

- La estacionariedad de segundo orden está definida en términos del covariograma mientras que la estacionariedad intrínseca está definida en términos del semivariograma.
  - El semivariograma es un estimador insesgado que, sin embargo, se ve afectado cuando hay observaciones atípicas.

- En el caso de un proceso estacionario, el variograma y el semivariograma pueden representarse como una función del vector diferencia entre las posiciones,  $h = s_j - s_i$  conocido como *lag* [4]:

$$\gamma(s_i, s_j) = \gamma(s_j - s_i)$$

- Si el proceso estacionario es además isotrópico, el variograma y el semivariograma pueden representarse como función solamente de la distancia entre dos posiciones  $s$  y  $s + h$  [4]:

$$\gamma(s, s + h) = \gamma(h)$$

- En resumen, el semivariograma es una función que relaciona la semivarianza con la separación en distancia y dirección o *lag*  $h$  de cualquier par de localizaciones  $Z(s)$  y  $Z(s + h)$ .
- Las condiciones necesarias para el variograma son:
  - 1 Condicional definido negativo:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j 2\gamma(s_i - s_j) \leq 0$$

- 2 Hipótesis intrínseca: cuando  $\|h\| \rightarrow \infty$ , tenemos que:

$$\frac{2\gamma(h)}{\|h\|^2} \rightarrow 0$$



- Los parámetros del semivariograma caracterizan los tres elementos más importantes de la variabilidad de un atributo:
  - La discontinuidad en el origen (existencia del efecto de pepita)
  - El valor máximo de la variabilidad (meseta)
  - El área de influencia de la correlación (alcance)

- El **rango** o **alcance** de un proceso espacial es la distancia  $h_s$  a la que se alcanza la meseta:

$$\gamma(h) \approx c_s \quad \forall h > h_s$$

- Es decir, es la distancia en la cual la diferencia del variograma y de su silla es prácticamente despreciable:  $\gamma(h) \approx c_s$ .
- Se dice aproximadamente ( $\approx$ ) porque algunos procesos alcanzan la correlación cero sólo asintóticamente para  $\|h\| = \infty$ ; mientras que otros, tienen un rango finito.
- No obstante, la posición exacta, en donde la diferencia es "despreciable" es bastante difícil de detectar.

- Un semivariograma crece con la distancia, dando lugar al fenómeno de que el proceso es similar en puntos próximos hasta que se estabiliza en un valor denominado **silla**, **umbral**, **alféizar** o **meseta** (*sill*) que expresa la variabilidad entre puntos distantes.
- Es decir, es el límite del variograma cuando las distancias  $h$  tienden a infinito (la asíntota superior del semivariograma si el variograma es una función acotada):

$$\lim_{h \rightarrow \infty} \gamma(h) = c_s > 0$$

- Únicamente, los procesos estacionarios de segundo orden tiene silla.
  - En estos casos, la silla es  $c_s = C(0)$ .
- Puede ser obtenido trazando una línea recta entre los primeros puntos del semivariograma empírico y extender ésta hasta que se intercepte con el eje  $y$ .
  - Si esta intersección ocurre por debajo de cero, el valor asumido por este efecto es cero, pues valores negativos de  $\gamma(0)$  no tienen significado y no es una situación muy común.

- El rango de un proceso espacial puede ser visto como la distancia a la cual los puntos ya no se consideran correlacionados o la distancia  $h$  para la cual las variables  $Z(s)$  y  $Z(s + h)$  son independientes.
  - Los puntos separados por una distancia inferior al rango se consideran espacialmente correlacionados.
- Por otra parte, observaciones espaciadas por más que el rango, se consideran independientes.

- El semivariograma, por definición, es nulo en el origen, pero en la práctica las funciones obtenidas pueden presentar discontinuidad en el origen.
- El **efecto pepita**, término extraído de las aplicaciones en la minería (pepita, *nugget*), es la situación en la que el variograma no tiende a 0 al acercarse al origen.

$$\lim_{h \rightarrow 0} \gamma(h) = c_0 > 0$$

- Esta discontinuidad en el origen puede deberse a mediciones erróneas y/o microvariabilidades (variaciones a muy pequeña escala).

- Todos los *estimadores kriging* son variaciones del estimador de regresión lineal  $Z^*(s)$  [7] definido como:

$$Z^*(s) - m(s) = \sum_{i=1}^{n(s)} \lambda_i(s) \cdot [Z(s_i) - m(s_i)]$$

donde  $s$  y  $s_i$  son la posición de interés y sus posiciones vecinas,  $n(s)$  es el número de puntos del vecindario local,  $m(s)$  y  $m(s_i)$  son los valores medios de las variables aleatorias  $Z(s)$  y  $Z(s_i)$  respectivamente y  $\lambda_i(s)$  es el *peso de kriging*<sup>3</sup> asignado al dato  $Z(s_i)$  en la posición de interés  $s$ .

---

<sup>3</sup>De esta forma, el mismo dato puede tener diferentes pesos para diferentes posiciones de estimación.

- Es el método más básico de interpolación espacial geoestadística también denominado *kriging lineal con valores esperados conocidos*.
- Requiere estacionariedad de segundo orden y considera que la media es constante y conocida:

$$E [Z (s)] = m (s) = \mu \text{ conocida}$$

- Este método es difícil de aplicar ya que requiere demasiado conocimiento de la variable,  $Z (s)$ .



- El estimador de kriging simple es:

$$Z_{SK}^*(s_0) = \mu + \sum_{i=1}^n \lambda_i^{SK}(s_0) \cdot [Z(s_i) - \mu]$$

- La varianza del kriging simple es:

$$\sigma_{SK}^2(s) = C(0) - \sum_{i=1}^n \lambda_j^{SK}(s_0) \cdot C(s_i - s_0)$$

y la varianza del error es:

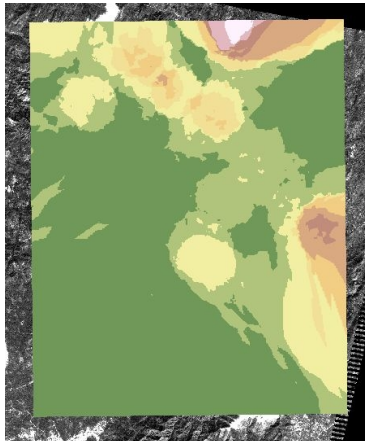
$$\sigma_E^2(s_0) = \text{var}[Z(s_0)] - \sum_{i=1}^n \lambda_i \cdot C(s_i - s_0)$$

esto es, la varianza del estimador es menor que la varianza de la variable a estudiar no regionalizada.

# IDW con ArcMap



# Kriging con ArcMap: Ordinary Kriging



# Kriging con ArcMap: Universal Kriging



- Sean  $p_i$  y  $o_i$  los valores estimados y las observaciones donde  $n$  es el número observaciones.
- Sean  $p_{si}$  y  $o_{si}$  los valores estimados y las observaciones estandarizados.
- Sea  $s$  es la desviación estándar del error de la estimación.
- Sea  $\bar{o}$  es la media de los valores observados.
- Sean  $p'_i = p_i - \bar{o}$  y  $o'_i = o_i - \bar{o}$ .

- 1 Error medio de la predicción o *Mean Prediction Error* (ME) o *Bias* (*Mean Bias Error*, MBE):

$$ME = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)$$

- 2 Promedio absoluto del error o *Mean Absolute Error* (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i|$$

- 3 Promedio del error cuadrático o *Mean Square Error* (MSE) o *Mean Square Prediction Error* (MSPE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2$$

- 4 Precisión o *Root Mean Square Error* (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2}$$

- 5 Error cuadrático medio reducido o *Mean Square Reduced Error* (MSRE):

$$MSRE = \frac{1}{n} \sum_{i=1}^n \frac{(p_i - o_i)^2}{s^2}$$

- 6 Error medio estándar o *Mean Standardized Error* (MSE2):

$$MSE2 = \frac{1}{n} \sum_{i=1}^n (p_{si} - o_{si})$$



- 7 Error cuadrático medio estándar o *Root Mean Square Standardized Error* (RMSSE):

$$RMSSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_{si} - o_{si})^2}$$

- 8 Error medio estándar o *Averaged Standard Error* (ASE):

$$ASE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[ p_i - \left( \frac{1}{n} \sum_{i=1}^n p_i \right)^2 \right]}$$

- 9 Willmott's D:

$$D = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (|p_i| - |o_i|)^2}$$

- 10 Cociente de las varianzas de los valores estimados y de los valores observados (RVar):

$$RVar = \frac{Var [p]}{Var [o]}$$

- 11 Eficiencia del modelo o *Model Efficiency* (EF) o R2:

$$EF = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2}$$

- 12 Efectividad de la Predicción (E):

$$E = 100 \cdot EF$$

- 13 Error cuadrático medio normalizado o *Mean Square Normalized Error* (MSNE):

$$MSNE = \frac{1}{n} \sum_{i=1}^n \left[ \frac{p_i - o_i}{s_i} \right]^2$$

- 14 Raíz del error cuadrático medio normalizado o *Root Mean Normalized Square Prediction Error* (RMNSE) o *Mean Squared Deviation Ratio* (MSDR):

$$RMNSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[ \frac{p_i - o_i}{s_i} \right]^2} = \sqrt{MSNE}$$

- 15 Error de la mediana o *Median Error* (MDE):

$$MDE = \text{median}(o_i)$$

- 16 *Correlation Observed and Predicted (COP):*

$$COP = cor [o_i, p_i]$$

o

$$COP = cor [o_i, o_i - (p_i - o_i)]$$

- 17 *Correlation Predicted and Residual Error (CPR):*

$$CPR = cor [p_i, p_i - o_i]$$

o

$$CPR = cor [o_i - (p_i - o_i), p_i - o_i]$$

- 18 EV:

$$EV = 1 - \frac{var [p_i - o_i]}{var [o_i]}$$

19 R:

$$R = \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n \sigma_K^2(x_i)}$$

## 1 *Relative Improvement* (RI):

$$RI = 100 \cdot \frac{RMSE(mejor) - RMSE(metodo)}{RMSE(mejor)}$$

compara el valor del mejor RMSE obtenido con el proporcionado por el método a comparar donde

$$RMSE(mejor) = \min(RMSE)$$