

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

---

## **Estadística Espacial:** Primera entrega

---

Nicolás Benhur Cardona Giraldo  
Deivid Alejandro Coronado Sarmiento  
Luisa Fernanda Guantiva Vargas  
John Dagoberto Moyano Duarte

Universidad Nacional de Colombia  
Facultad de Ciencias  
Departamento de Estadística  
Diciembre 2021

# Tabla de contenido

---

<b>1. Introducción</b>	<b>2</b>
<b>2. Planteamiento del Problema</b>	<b>4</b>
<b>3. Objetivos</b>	<b>5</b>
<b>4. Metodología</b>	<b>6</b>
4.1. Descripción de la base de datos . . . . .	6
4.2. Estacionariedad en media . . . . .	7
4.3. Semivariograma empírico . . . . .	9
4.4. Semivariograma teórico . . . . .	11
4.4.1. Estimación por mínimos cuadrados ordinarios . . . . .	11
4.4.2. Estimación por mínimos cuadrados ponderados . . . . .	11
4.4.3. Estimación por mínimos cuadrados de Cressie . . . . .	12
4.4.4. Estimación por máxima verosimilitud . . . . .	12
4.4.5. Estimación por máxima verosimilitud restringida . . . . .	12
<b>5. Conclusiones</b>	<b>14</b>

# Introducción

---

El desastre de Chernóbil es uno de los accidentes más graves en la historia de la industria nuclear. La explosión que se produjo el 26 de abril de 1986 en el reactor cuatro de la central nuclear situada en Ucrania, 17 km al sur de la frontera con Bielorrusia. El estallido y el incendio posterior liberaron a la atmósfera grandes cantidades de yodo radiactivo y cesio que las corrientes de aire dispersaron por el cielo de Bielorrusia, Rusia y Ucrania. Las autoridades soviéticas intentaron ocultar el desastre, decidieron no avisar a los habitantes de la ciudad más cercana a la planta. Las primeras reacciones llegaron en la tarde del día siguiente, cuando los trabajadores de la planta y los cerca de 50.000 habitantes de Prípiat fueron evacuados. Pero los dos millones de residentes de Kiev, hoy la capital de Ucrania, seguían sin tener idea de qué estaba pasando France24 (2021).



Figura 1.1: Nube tomado de NatGeo and Alcalde (2021)

La primera alerta de los altos niveles de radiación llegó desde el norte de Suecia dos días después euronews (2011), desde la planta de Forsmark, a más de 1.100 kilómetros de Chernóbil dado que uno de los empleados de la planta de energía nuclear notó los niveles alterados “Gracias a nues-

tra detección temprana, pudimos informar a las autoridades suecas, quienes luego informaron al mundo sobre la contaminación radioactiva proveniente del desastre en la Unión Soviética”, dijo Claes-Göran Runermark, el gerente de operaciones a cargo en ese momento. Finlandia y Alemania también reportaron altos niveles de radioactividad. Y solo después de eso, el mismo 28 de abril, las autoridades soviéticas informaron escuetamente a través de los canales oficiales de televisión lo que había ocurrido. “Hubo un accidente en la central de energía de Chernóbil y uno de los reactores resultó dañado”, rezaba el comunicado France24 (2021).



Figura 1.2: Mapa tomado de ICRIN (2000)

Durante mucho tiempo y en la actualidad muchas regiones están contaminadas con material radiactivo procedente de Chernobyl. Por esta razón desde la fecha del accidente se han tomado diferentes muestras de la concentración de isótopos y partículas de aerosol radiactivas en distintos lugares, fechas y horas para hacer un análisis de la contaminación del aire y sus posibles consecuencias en diferentes regiones para prevenir y saber cuales regiones pueden ser peligrosas, ya que en su momento no se tuvo prevención y hubo consecuencias graves desde intoxicaciones por yodo, deformaciones, y peligrosas enfermedades que en la actualidad pueden tener efectos secundarios graves como Cáncer de tiroides, leucemia, cataratas y depresión. NatGeo and Alcalde (2021)

# Planteamiento del Problema

---

A partir de los datos registrados en las estaciones nucleares queremos hacer un análisis espacial de la concentración de las partículas de Yodo-131 (I-131), Cesio-134 (Cs-134) y Cesio-137 (Cs-137) para entender cómo se esparció por todo el continente europeo y América del Norte (cienciaplus (2018)). Varios países europeos se vieron muy afectados por la explosión debido a una transferencia atmosférica masiva de material radiactivo y dado que, hasta el día de hoy, el impacto real de este evento sigue siendo objeto de controversia.

El análisis geoespacial es un recurso útil e interesante para poder identificar las cantidades de estos materiales radiactivos, por lo cual elegimos el reactivo más potente en fechas y horas específicas para poder observar el comportamiento, poder realizar un análisis de los datos observados y validar si existe o no correlación espacial.

## Objetivos

---

- Realizar un análisis descriptivo de la concentración del material radiactivo Cesio-137 en el aire del territorio europeo.
- Explorar las distintas funciones de semivarianza para poder compararlas entre sí para ver la relación existente entre la ubicación geográfica y la concentración de Cesio-137.

## 4.1. Descripción de la base de datos

La base de datos sobre el material radioactivo expulsado en el accidente de Chernobyl se extrajo de la pagina *Kaggle* Chemkaeva (2019). En esta base se encuentran 3 variables Cesio-134, Yodo-131 y Cesio-137 las cuales son medidas en  $Bq/m^3$ . El bequerelio por metro cúbico (Bq) es la unidad internacional de radiactividad igual a una desintegración nuclear por segundo. ehu (2021)

Dado que el Yodo-131 y Cesio-134 tienen una vida media de desintegración de 8 días y 2.06 años, respectivamente, en comparación con el isótopo Cesio-137 que presenta una vida media aproximada de 30.17 años, Okumura (2003), elegimos éste último como nuestra variable de interés. La siguiente tabla presenta la variable elegida que incluye el país y la ciudad que registró la información, así como la latitud y longitud donde se encuentra la estación.

País	Ciudad	Latitud	Longitud	CS_137
Dinamarca	HANNOVER	9,73	52,38	0,000
Dinamarca	NEUHERBERG	11,58	48,13	2,100
Dinamarca	NORDERNEY	7,15	53,71	0,089
Francia	CADARACHE	5,75	43,71	0,410
Francia	CHINON	0,25	47,15	0,650
Francia	CHOOZ	4,81	50,10	3,000
Francia	CRUAS	4,80	44,63	0,095
Francia	FESSENHEIM	7,56	47,91	1,500
Francia	FLAMANVILLE	-1,90	49,55	0,000
Francia	GRAVELINES	2,31	51,00	0,000
Francia	GRENOBLE	5,72	45,18	0,163
Francia	MARCOULE	4,80	44,13	0,999
Francia	MONACO	7,42	43,73	0,180
Francia	ORSAY	2,18	48,68	3,230
Francia	PARIS	2,36	48,85	1,490
Francia	SACLAY	2,16	48,73	1,400
Francia	TRICASTIN	4,73	44,33	0,170
Francia	VERDUN	5,38	49,15	5,400
Italia	BOLOGNA	11,33	44,48	2,237
Inglaterra	HARWELL	-1,30	51,62	0,002
Belgica	MOL	5,12	51,18	0,970
R.Checa	FRIBOURG	7,12	46,49	1,147
Noruega	BERGEN	10,66	59,90	0,000

Figura 4.1: Tabla de los países y ciudades con sus respectivas mediciones de material radioactivo Cesio-137 del día 02 de Mayo de 1986 a las 09:00 A.M.

El objetivo es conocer la concentración de Cesio-137 en Europa luego de lo sucedido en el accidente de la planta nuclear de Chernobyl. Esta variable se filtra en el día 02 de Mayo de 1986 a las 09:00 A.M. de la cual podemos observar 23 observaciones, elegimos esta fecha y hora estratégicamente ya que es la fecha más cercana al accidente que cuenta con más información.

## 4.2. Estacionariedad en media

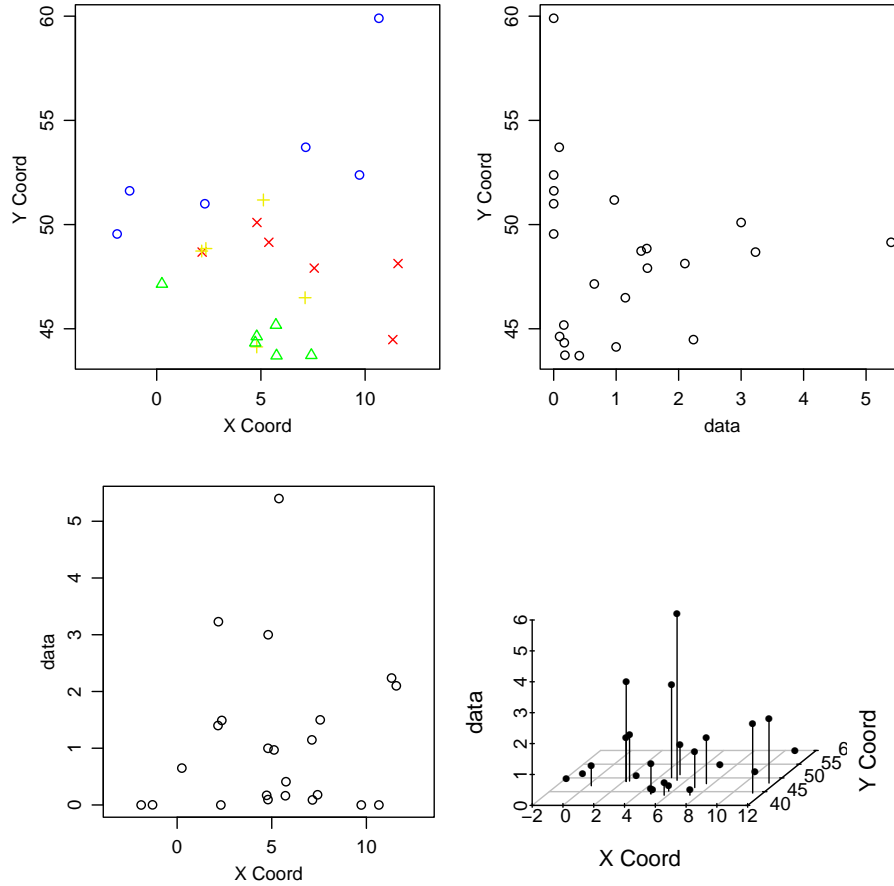


Figura 4.2: Análisis gráfico de la estacionariedad en media

En el gráfico (4.2) podemos ver que parece existir tendencia tanto para la coordenada en X como para la coordenada en Y; por ejemplo, en ésta última los datos se agrupan alrededor de cero. Como existe tendencia, no podemos asumir una media constante, por lo cual, se ajustan diferentes modelos de regresión y se escoge el que tenga mayor Coeficiente de determinación ajustado ( $R_{adj}^2$ ). En la tabla (4.1) se muestran los valores obtenidos. De acuerdo al  $R_{ajus}^2$  se escoge el cuarto modelo incluyendo hasta un término cúbico en la coordenada Y. En el gráfico 4.3 podemos ver el cambio de los puntos principalmente en la coordenada y.



Modelo		$R^2_{ajus}$	Valor P
(1)	$\mu = \beta_0 + \beta_1x + \beta_2y$	0	0.83
(2)	$\mu = \beta_0 + \beta_1x + \beta_2y + \beta_3xy$	0	0.7
(3)	$\mu = \beta_0 + \beta_1x + \beta_2y + \beta_3y^2$	0.08	0.21
(4)	$\mu = \beta_0 + \beta_1x + \beta_2y + \beta_3y^2 + \beta_4y^3$	0.23	0.07

Cuadro 4.1: Resumen valores  $R^2_{ajus}$  y valores-P para los modelos de regresión ajustados

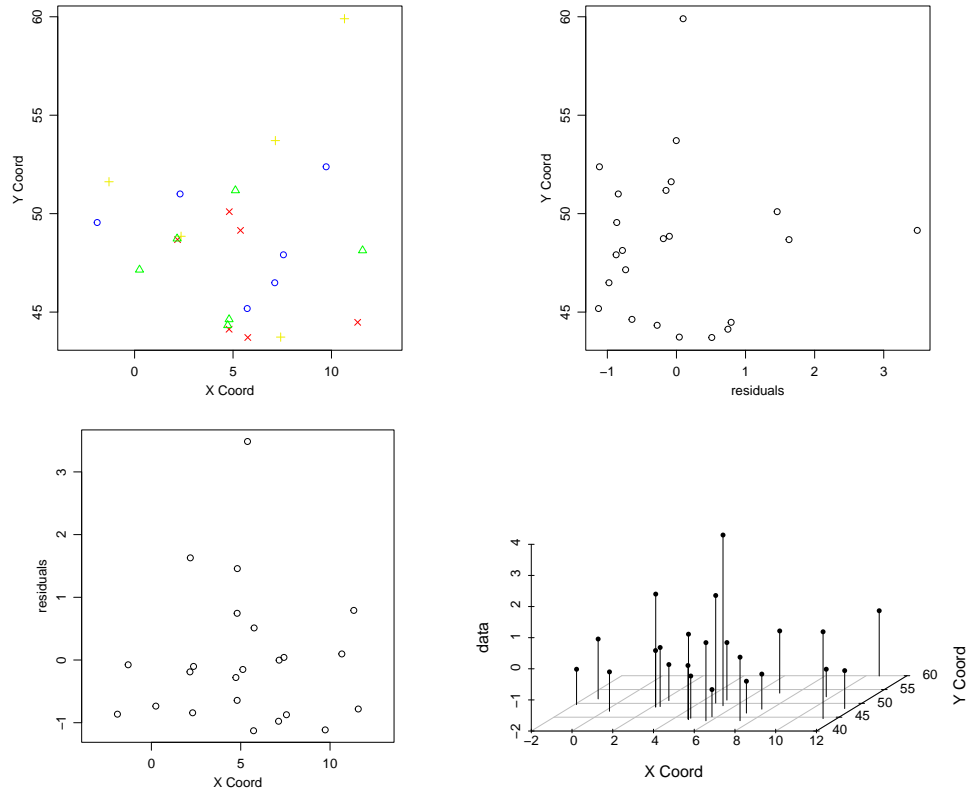


Figura 4.3: Análisis gráfico de la estacionariedad en media sin tendencia

En contraste con la gráfica 4.2, podemos ver un cambio en la distribución de los puntos en la conjunta de X y Y. Podemos ver que, por ejemplo, los triángulos verdes, así como las demás figuras dejan de estar agrupadas de manera que se elimina la tendencia de la estacionariedad en media que presentaban los datos.

### 4.3. Semivariograma empírico

En la figura (4.4) se muestran los gráficos del semivariograma empírico a 0, 45 y 90 grados, así como el omnidireccional para detectar la posible isotropía del proceso.

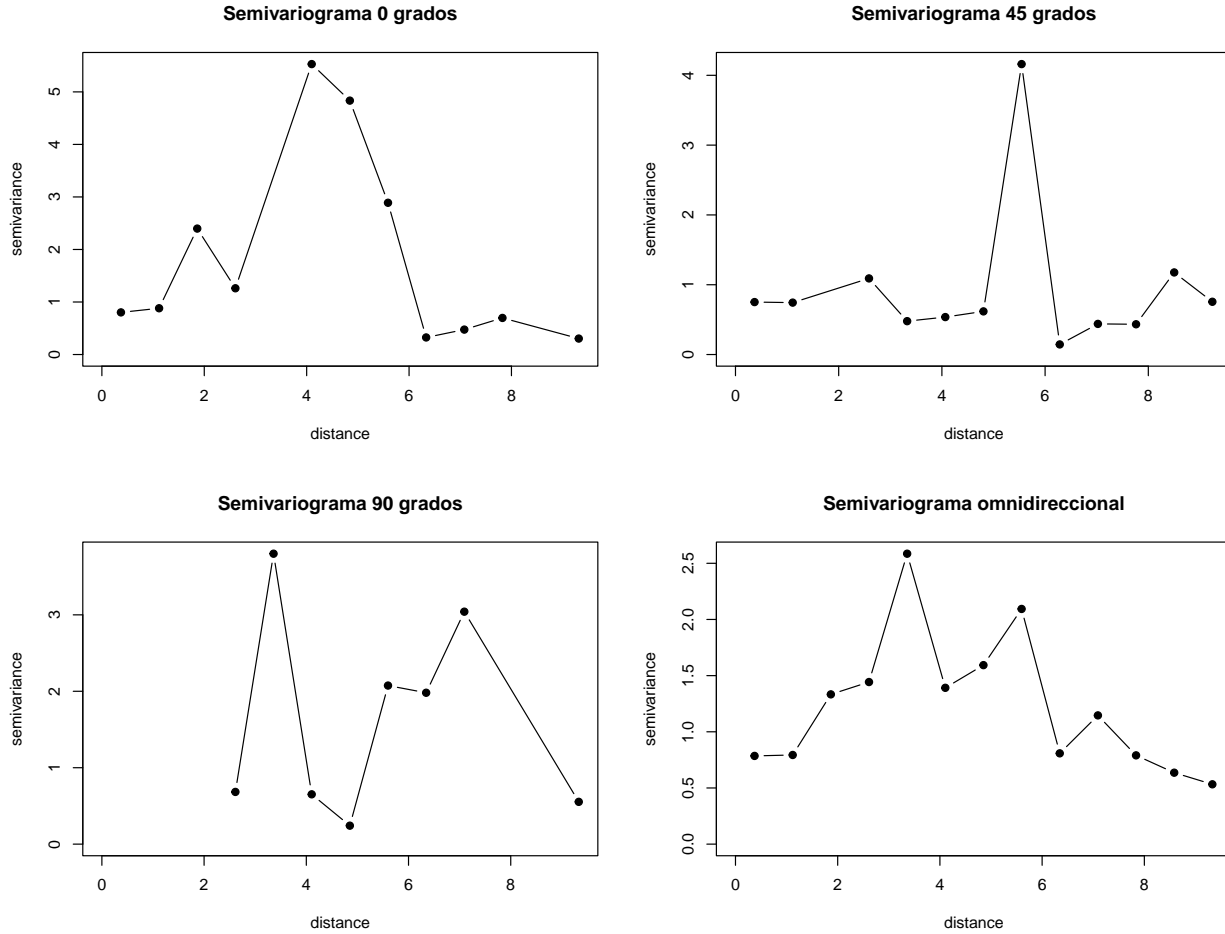


Figura 4.4: Semivariogramas en diferentes direcciones

Se puede observar que en cada dirección tienen comportamientos ligeramente distintos. En 0 grados, hay un pico alto en  $Y = 5$  y decrece a una distancia de 6. En contraste, a 45 grados se observa un comportamiento estable con un pico de corta duración al rededor de 6; mientras que a 90 grados, existen dos picos: uno al rededor de 3 con  $y=4$ , y otro entre 6 y 8. De modo que el semivariograma omnidireccional refleja estos resultados.

Como sabemos el semivariograma empírico es muy útil para determinar cuál puede ser el modelo que sigue el proceso para detectar su dependencia espacial así como los parámetros iniciales. Lo que se suele hacer es agrupar los puntos por intervalos de rezagos espaciales y en cada intervalo, revisar cuantos pares de datos hay y calcular la semivarianza espacial. En este trabajo utilizamos estimación tendencia de primer y segundo orden pero decidimos elegir el semivariograma empírico de los residuales del modelo de regresión escogido en la sección anterior (modelo de regresión 4), por los datos que podemos observar en la tabla (4.1).

#	Número de parejas	Distancias	Semivariograma
1	11	0.65	0.77
2	16	1.95	1.74
3	29	3.25	1.74
4	42	4.55	1.56
5	38	5.86	1.73
6	37	7.16	0.95
7	18	8.46	0.69
8	30	9.76	0.43
9	8	11.06	0.58
10	6	12.36	1.18
11	7	13.66	0.49
12	4	14.97	0.35
13	6	16.27	0.23

Cuadro 4.2: Relación del número de pares de puntos con respecto a la semivarianza y la distancia

Según el Cuadro 4.2 podemos observar que a partir del punto 9 el número de parejas es significativamente menor que con respecto a los demás puntos. Se evidenció que cuando se grafican todos los números de parejas se genera un comportamiento muy errático y poco confiable. «La estimación del semivariograma es una mezcla de ciencia y arte» Waller and Gotway (2004) Por esta razón, en principio, hemos decidido incluir los puntos hasta una distancia de 9,76.

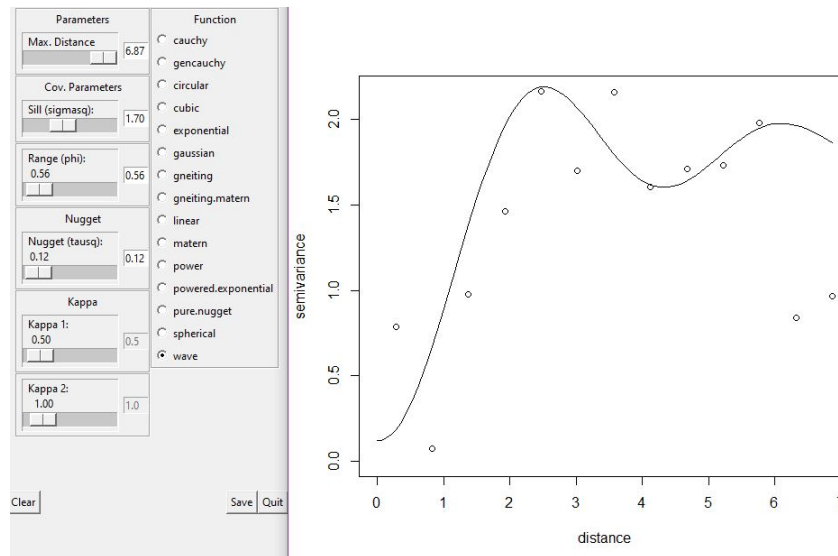


Figura 4.5: Eyefit: Estimación a Sentimiento

Así, generamos los puntos del semivariograma empírico y estimamos los parámetros de un modelo teórico con la herramienta *eyefit* del paquete *geoR* de (R Development Core Team (2007)). Luego de varios intentos y comparaciones escogemos el modelo "Wave", con una distancia máxima de 7.16 y parámetros silla de 1.7, pepita de 0.12 y rango 0.56 como se muestra en la figura 4.5

## 4.4. Semivariograma teórico

Usaremos los métodos de estimación teórica del semivariograma tomando los valores del semivariograma empírico obtenido en la sección anterior como variable respuesta en función de la distancia. Bohorquez (2021)

Se presentan los resultados obtenidos con los métodos de mínimos cuadrados ordinarios, mínimos cuadrados ponderados, mínimos cuadrados de Cressie, máxima verosimilitud y máxima verosimilitud restringida. Al final de la sección, se realiza la comparación entre ellos y un gráfico con las curvas de estimación teóricas del semivariograma.

### 4.4.1. Estimación por mínimos cuadrados ordinarios

Para la estimación por mínimos cuadrado ordinarios es necesario minimizar la expresión dada por:

$$\sum_{t=1}^k (\gamma(\hat{h}_t) - \gamma(h_t; \theta))^2$$

A partir de lo anterior, los parámetros obtenidos fueron los siguientes:

Parámetros	Estimaciones
$a$	0.5
$\sigma^2$	1.2915691
$\phi$	0.8082762

Cuadro 4.3: Parámetros de los mínimos cuadrados ordinarios

### 4.4.2. Estimación por mínimos cuadrados ponderados

Para la estimación por mínimos cuadrado ponderados es necesario minimizar la expresión dada por:

$$[(2\gamma(h) - 2\gamma(\hat{h}_i; \theta)]W^{-1}[2\gamma(\hat{h}) - 2\gamma(h_i; \theta)]$$

A partir de lo anterior, los parámetros obtenidos fueron los siguientes:

Parámetros	Estimaciones
$a$	0.5
$\sigma^2$	1.4133380
$\phi$	0.4870028

Cuadro 4.4: Parámetros de los mínimos cuadrados ponderados

#### 4.4.3. Estimación por mínimos cuadrados de Cressie

Para la estimación por mínimos cuadrado de Cressie es necesario minimizar la expresión dada por:

$$\sum_{i=1}^k n_k \left[ \frac{\hat{\gamma}_k - \gamma_k(\theta)}{\gamma_k(\theta)} \right]^2$$

A partir de lo anterior, los parámetros obtenidos fueron los siguientes:

Parámetros	Estimaciones
a	0.5
$\sigma^2$	1.5459703
$\phi$	0.5383552

Cuadro 4.5: Parámetros de los mínimos cuadrados de Cressie

#### 4.4.4. Estimación por máxima verosimilitud

Para la estimación por máxima verosimilitud es necesario minimizar la expresión dada por:

$$\frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (Z - X\beta)^t \Sigma(\theta)^{-1} (Z - X\beta)$$

A partir de lo anterior, los parámetros obtenidos fueron los siguientes:

Parámetros	Estimaciones
a	0.12
$\sigma^2$	1.088
$\phi$	0.618

Cuadro 4.6: Parámetros de los mínimos cuadrados ordinarios

#### 4.4.5. Estimación por máxima verosimilitud restringida

Para la estimación por máxima verosimilitud restringida es necesario minimizar la expresión dada por:

$$\frac{n-1}{2} \log(2\pi) + \frac{1}{2} \log |A^T \Sigma(\theta) A| + \frac{1}{2} Z^T A (A^T \Sigma(\theta) A)^{-1} A^T Z$$

A partir de lo anterior, los parámetros obtenidos fueron los siguientes:

Parámetros	Estimaciones
a	0.12
$\sigma^2$	1.16
$\phi$	0.6205

Cuadro 4.7: Parámetros de la máxima verosimilitud restringida

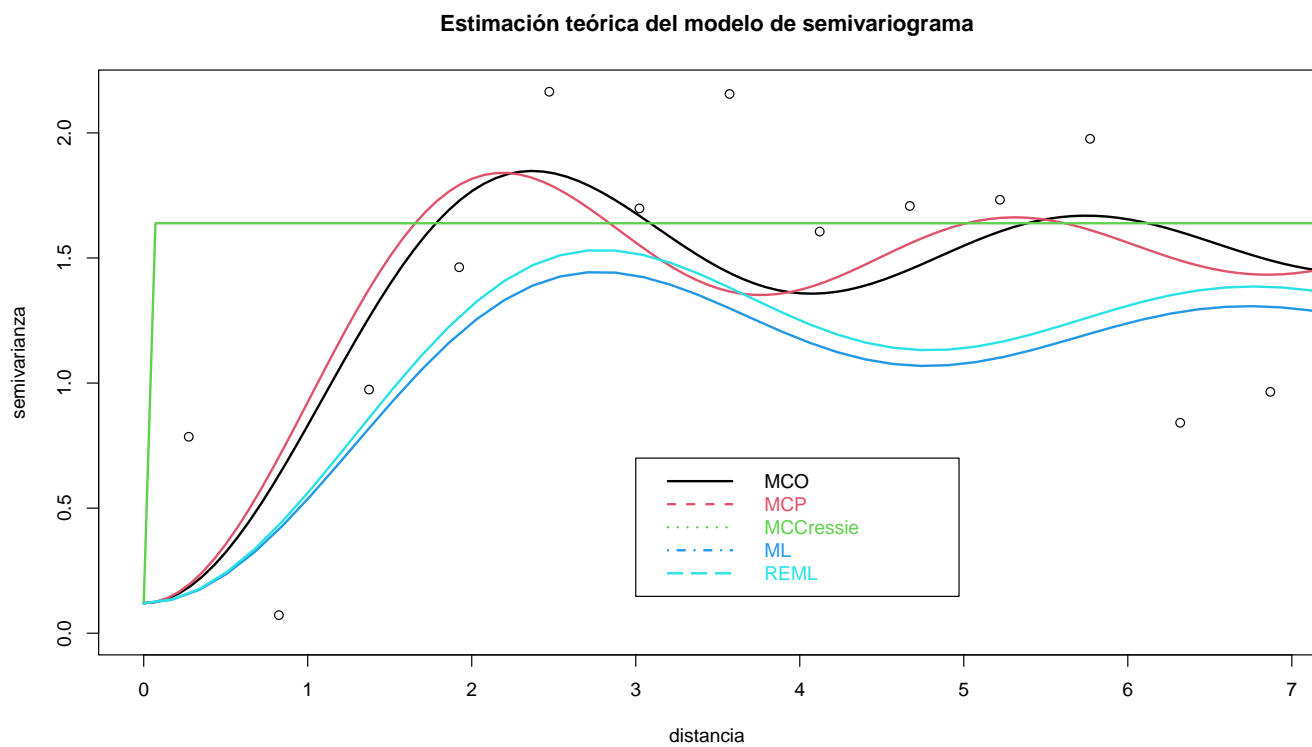


Figura 4.6: Estimación teórica del Semivariograma

## Conclusiones

---

- Como en un inicio el proceso parecía tener tendencia sobre todo en el eje y, optamos por trabajar con los residuales obtenidos del modelo cúbico mostrado en la sección 4.2
- Respecto al semivariograma, mediante la función `eyefit` de R escogimos el modelo «Wave» con una distancia máxima de 7.16 y parámetros silla de 1.7, pepita de 0.12 y rango 0.56. Luego se realiza estimación teórica mediante 5 métodos diferentes encontrando que Mínimos cuadrados ponderados es el que mejor se adecua a nuestros datos.
- Como conclusión extra, podemos observar en la figura 5.1 el mapa de la izquierda que ubica cada uno de los puntos de nuestra muestra tomada de la base de datos de chernóbil y el punto en rojo del desastre; en la imagen de la derecha podemos observar el gráfico de interpolación del cual se podría decir que con base en los datos registrados de la concentración del aire con cesio 137 se distribuye de forma horizontal hacia el oeste de Europa.

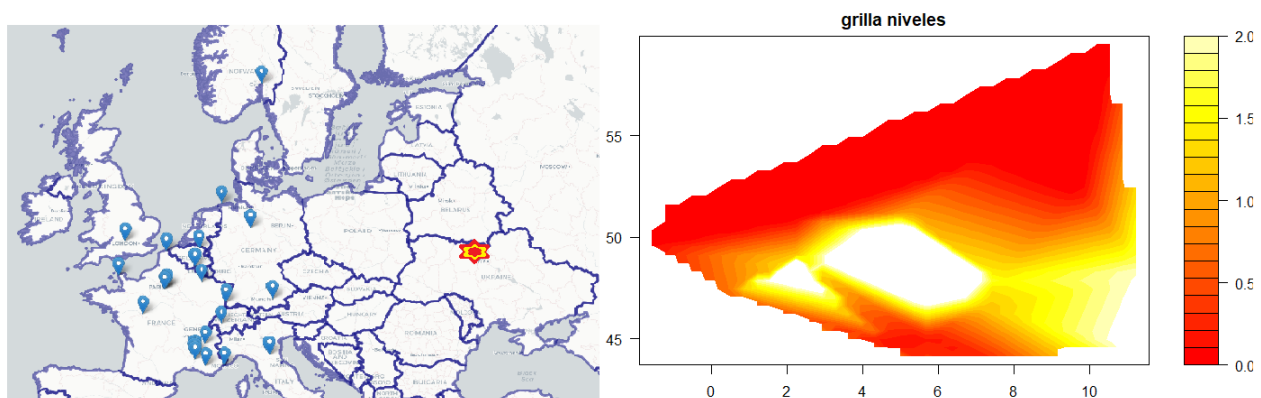


Figura 5.1: Gráficos de interpolación

## Referencias

---

- Bohorquez, M. (2021). *Estadística Espacial y Espacio-Temporal para Campos Aleatorios Escalares y Funcionales. Notas de clase*. Universidad Nacional de Colombia. Departamento de Estadística.
- Chemkaeva, D. (2019). Chernobyl air concentration. <https://www.kaggle.com/lsind18/chernobyl-data-air-concentration>.
- cienciaplus (2018). Chernóbil se desató por una explosión nuclear, seguida de otra de vapor. <https://www.europapress.es/ciencia/cambio-climatico/noticia-chernobil-desato-explosion-nuclear-seguida-otra-vapor-20171117171506.html>.
- ehu (2021). Unidades de radioactividad. <http://www.ehu.eus/biomoleculas/isotopos/unidades.htm#:~:text=La%20unidad%20de%20radioactividad%20del,es%20una%20desintegraci%C3%B3n%20por%20segundo>.
- euronews (2011). La explosión en el reactor número cuatro de chernóbil. [https://www.youtube.com/watch?v=\\_tYSF3D9Ejg](https://www.youtube.com/watch?v=_tYSF3D9Ejg).
- France24 (2021). Chernóbil: la vida 35 años después del peor accidente nuclear de la historia. <https://www.france24.com/es/europa/20210425-chernobil-35-anos-despues-peor-accidente-nuclear>.
- ICRIN (2000). Preface: The chernobyl accident. <https://web.archive.org/web/20110723112745/http://chernobyl.info/en-US/Home/History-of-Chernobyl-Disaster/The-Accident.aspx>.
- NatGeo and Alcalde, S. (2021). Las enfermedades causadas por el accidente nuclear de chernóbil. [https://www.nationalgeographic.com.es/ciencia/enfermedades-causadas-por-accidente-nuclear-chernobil\\_14365](https://www.nationalgeographic.com.es/ciencia/enfermedades-causadas-por-accidente-nuclear-chernobil_14365).
- Okumura, T. (2003). The material flow of radioactive cesium-137 in the u.s. 2000. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=875E9C73ACFB282C520A11F123686033?doi=10.1.1.476.4830&rep=rep1&type=pdf>.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Waller, L. and Gotway, C. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley Series in Probability and Statistics. Wiley.