



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Licenciatura en Estadística

INVESTIGACIÓN DE METODOLOGÍAS PARA EL ANÁLISIS ESPACIO-TEMPORAL DEL EVENTO GRANIZO EN URUGUAY

Lucía Rijo - Florencia Santiñaque

Tutores:

Laura Nalbarte

Ramón Álvarez

Co-tutor:

Carlos Rivera

Montevideo, Marzo 2011.

Agradecimientos

A nuestras familias que estuvieron en todo momento incondicionalmente y por brindarnos todo su apoyo durante este largo proceso.

A nuestros tutores: Ramón Álvarez, Laura Nalbarte y Carlos Rivera por su apoyo, dedicación y sugerencias realizadas.

A Compañía Cooperativa de Seguros SURCO por dar inicio a este proyecto y por el aporte de datos para el mismo.

A la Oficina de Estadísticas Agropecuarias (DIEA) por aportes y sugerencias realizadas.

A Carmen Mesones que nos brindó sus conocimientos y su voluntad para enseñarnos.

A Andrés Castrillejo por su aporte en un tema de importancia para nuestro trabajo.

A Diego Vázquez Melo por enseñanza en el área meteorológica.

A Dirección Nacional de Meteorología por su aporte de base de datos.

A Mario Bidegain por aporte de datos y sugerencias desde Facultad de Ciencias.

A nuestros compañeros de trabajo y amigos por el apoyo incondicional y buena onda.

Índice general

Índice general	2
I Introducción	6
1. Introducción	7
2. Objetivo	9
3. Antecedentes	10
II Marco Teórico	12
4. El Granizo y sus efectos	13
4.1. El granizo	13
4.2. Efectos	14
5. Territorio uruguayo	15
5.1. Geografía	15
5.1.1. Superficie	15
5.1.2. Relieve	15
5.1.3. Puntos extremos	16
5.1.4. Clima	17
5.2. Estaciones meteorológicas	18
III Marco Metodológico	19
6. Interpolación por Kriging	20
6.1. Introducción	20
6.2. Geoestadística	20
6.2.1. Datos georeferenciados	21

ÍNDICE GENERAL

6.2.2. Coordenadas geográficas	21
6.2.3. Proyección geográfica	22
6.3. El modelo espacial general	23
6.3.1. Procesos estacionarios	23
6.4. Tendencia espacial	25
6.4.1. Estimación de la tendencia	25
6.4.2. Implicancias de la existencia de tendencia	26
6.5. Interpolación espacial	27
6.6. Interpolación por Kriging	27
6.7. Kriging Ordinario	29
6.7.1. Ecuaciones de Kriging	29
6.7.2. Varianza de Kriging	31
6.7.3. Kriging Residual	31
6.8. Variograma	32
6.8.1. Variograma teórico	32
6.8.2. Variograma experimental	34
6.8.3. Modelos de variograma	35
6.8.4. Variograma bajo estructura de dependencia espacial	35
6.9. Validación Kriging	36
7. Análisis de Grupos Difusos	38
7.1. Análisis de Grupos	38
7.2. Análisis de Grupos Difusos	39
7.3. Incorporación de información espacio-temporal	41
7.3.1. Información espacial	41
7.3.2. Información temporal	42
IV Resultados y conclusiones	44
8. Resultados	45
8.1. Análisis exploratorio de los datos	45
8.1.1. Análisis descriptivo	47
8.2. Interpolación por Kriging Ordinario	50
8.2.1. Análisis de Tendencia	50
8.2.2. Kriging Residual	52
8.2.3. Estimación del Variograma	52
8.2.4. Estimación por Kriging	55
8.2.5. Validación Kriging	60
8.2.6. Conclusiones	62

ÍNDICE GENERAL

8.3. Análisis de Grupos Difusos	63
8.3.1. Características de los grupos	64
8.3.2. Análisis de coeficientes de pertenencia	66
8.3.3. Conclusiones	69
9. Conclusiones generales	71
9.1. Conclusiones	71
9.2. Limitaciones	71
9.3. Futuras investigaciones	72
Bibliografía	73
V Apéndice Metodológico	76
A. Interpolación por Kriging Ordinario	77
A.1. Análisis de tendencia	77
A.2. Obtención de las ecuaciones de Kriging	78
A.3. Variograma	78
A.3.1. Interpretación del variograma	78
A.3.2. Estimación del variograma	79
A.3.3. Variograma Omnidireccional	80
A.3.4. Variograma Cloud	81
A.3.5. Mapa de Variograma	81
A.3.6. Modelos de variograma	81
A.4. Covariograma	85
A.5. Anisotropía	86
B. Análisis de Grupos Difusos	87
B.1. Distancia	87
B.2. Algoritmos de clasificación	87
B.3. Reglas de detención	88
VI Apéndice Resultados	90
C. Interpolación por Kriging Ordinario	91
C.1. Análisis de cuatro escenarios	91
C.2. Análisis de la tendencia	95
C.3. Estimación del Variograma	97
C.4. Estimación por Kriging	97

D. Análisis de Grupos	99
D.1. Análisis de Grupos Jerárquico	99
D.2. Análisis de Grupos Difusos	100
D.2.1. Coeficientes de pertenencia	101
D.3. Análisis Instantáneo y Longitudinal	101
VII Apéndice Código R	104
E. Código en R	105
E.1. Análisis descriptivo	105
E.2. Interpolación por Kriging Ordinario	107
E.3. Análisis de Grupos	114
E.3.1. Análisis de Grupos Jerárquico	114
E.3.2. Análisis de Grupos Difusos	116
Índice de figuras	122
Índice de cuadros	124

Parte I

Introducción

Capítulo 1

Introducción

El aumento de la actividad agrícola tanto como la actividad aseguradora en Uruguay, y de la variabilidad climática en dicho territorio, impone la necesidad de contar con mejores herramientas que permitan conocer las zonas que son más o menos vulnerables a los diversos fenómenos climáticos para poder:

- Determinar el riesgo climático de las diversas zonas del país en relación a determinado evento.
- Mejorar la calidad de coberturas de seguro, incorporando un equilibrio racional entre riesgo a cubrir y costo del seguro para cada zona, por parte de las empresas aseguradoras.
- Generar información zonificada para eventos climáticos que pueda ser insumo confiable en la determinación de políticas públicas.

En Uruguay el subsector agropecuario representa el 8,1 % del PBI en forma directa y el 12,2 % si se considera la fase de industrialización de productos agropecuarios. En términos de exportaciones, los productos vegetales de la agricultura representan cerca del 27 % exportado anualmente [15]. El sector agropecuario es uno de los motores del desarrollo del país; adicionalmente, se ha producido en los últimos años un gran crecimiento del sector agrícola de granos y oleaginosos. Este crecimiento viene acompañado de la necesidad de mitigación de riesgos agro-climáticos. La contratación de seguros agrícolas como medida de prevención, ha seguido creciendo de manera que un alto porcentaje de las hectáreas cultivadas hoy, contratan alguna de las coberturas de seguros ofrecidas por las aseguradoras o integran algún sistema de aseguramiento mutual: de acuerdo al último relevamiento realizado por la Oficina de Programación y Política Agropecuaria (OPYPA) el 100 % de la superficie sembrada con cereales de invierno, y el 50 % de los cereales y oleaginosos de verano está amparada por algún tipo de cobertura.

La motivación de este trabajo, radica en la necesidad de contar con herramientas que permitan obtener un modelo de distribución espacio-temporal de eventos climáticos en todo el territorio uruguayo. Los eventos climáticos de granizo son en gran medida causa de eventos adversos, emergencias o desastres en todo el mundo. Dado que su ocurrencia no puede ser evitada se toman medidas preventivas y/o predictivas para reducir el riesgo de sus efectos a nivel económico.

El presente documento se encuentra estructurado en siete partes, nueve capítulos y tres apéndices. La primer parte de este trabajo contiene el detalle del objetivo del mismo así como el resumen de las investigaciones realizadas previamente. La segunda parte introduce la definición del evento climático granizo y sus efectos, así como la descripción del espacio geográfico a analizar: el territorio uruguayo. En la tercer parte se describen las técnicas investigadas: Interpolación por Kriging Ordinario y Análisis de Grupos Difusos, en la cuarta parte se detallan los resultados y conclusiones obtenidos, mientras que las partes quinta, sexta y séptima del documento contienen los apéndices metodológicos, de resultados y el código generado en el software utilizado para realizar los cálculos estadísticos, respectivamente.

Capítulo 2

Objetivo

El objetivo del presente trabajo es la revisión y análisis de metodologías estadísticas para la obtención de un mapa de riesgo climático en todo el territorio uruguayo, de manera de obtener zonas de comportamiento homogéneas para el fenómeno meteorológico de interés.

El fenómeno climático en estudio es el granizo. Se analiza el comportamiento espaciotemporal de dicho fenómeno desde dos perspectivas:

1. Realizar un análisis espacial del fenómeno. La importancia del espacio dentro del estudio estadístico de variables climáticas es incuestionable. En particular se desea analizar el problema de la dependencia espacial del evento climático granizo en una determinada zona geográfica, de manera de poder estimar dicho fenómeno en zonas donde se carece de información. La complejidad del problema aumenta debido a las interdependencias multidireccionales que puedan surgir en el espacio de estudio.
2. Estudiar la conformación de zonas homogéneas según la caída de granizo y otras variables correlacionadas al mismo. Asimismo se desea analizar el comportamiento temporal del evento y sus repercusiones en dicho análisis.

Capítulo 3

Antecedentes

Algunos intentos de zonificar riesgos han merecido cuestionamientos diversos: desde la falta de series confiables de información hasta lo discutible de la metodología estadística utilizada.

En Uruguay existen escasos desarrollos en estudios similares.

La empresa española Agroseguros en conjunto con el Ministerio de Ganadería, Agricultura y Pesca (MGAP), desarrollaron en 2008 un estudio para la generación de mapas de riesgos para varios fenómenos meteorológicos en Uruguay [17]. Se utiliza una metodología basada en análisis multivariante e interpolación por Kriging. A través del análisis de grupos jerárquico se procedió a estudiar tanto el fenómeno granizo así como los eventos meteorológicos lluvia y temperatura. Se generan mapas de riesgo mediante la aplicación de la metodología de Kriging a través del software Surfer 8 [19]. Fue uno de los primeros estudios realizados en nuestro país de forma interdisciplinaria. Aun así los resultados no son utilizados por ninguna empresa aseguradora. Esto puede deberse a que los resultados finales fueron desarrollados a partir de cinco estaciones meteorológicas del país, arrojando grandes errores de estimación.

Otro trabajo encontrado recientemente [1], llevado a cabo por la Dirección Nacional de Meteorología (DNM) realiza un breve informe sobre las distintas metodologías de interpolación espacial, entre ellas *Interpolación por Kriging Ordinario*. El desarrollo se realiza con datos sobre el fenómeno meteorológico precipitaciones, obteniendo un mapa de riesgo para dicho evento. Se comparan las diferencias y errores cometidos según los distintos tipos de interpolación.

En Argentina se han encontrado antecedentes de un estudio espacial del evento granizo denominado "Distribución espacio-temporal del granizo en Argentina" [6] desarrollado por el Instituto del Clima y Agua, Centro de Investigación de Recursos

Naturales e Instituto Nacional de Tecnología Agropecuaria. En el mismo se presenta la metodología *kriging* como herramienta geoestadística para obtener la distribución temporal y espacial de los eventos de granizo en Argentina.

Los mismos autores realizan un trabajo [7] de similares características en el que realizan énfasis en estudiar la variabilidad estacional del fenómeno granizo en ese mismo país.

En Ecuador se ha realizado un estudio [8] para clasificar musáceas en zonas bananeras en ese país. Se aplica la metodología de interpolación por Kriging en cuatro localidades pertenecientes al centro y sur del territorio ecuatoriano. En cada una de ellas se detallan las características encontradas según los distintos modelos utilizados.

Parte II
Marco Teórico

El Granizo y sus efectos

4.1. El granizo

El granizo es la precipitación atmosférica de agua congelada y amorfa. Está constituido por un núcleo central de hielo esponjoso y una envoltura gruesa, compuesta por capas alternas de hielo transparente y opaco, de forma esférica irregular [14]. Dicha precipitación se origina en nubes convergentes, como las cumulonimbus, en forma de glóbulos o trozos irregulares de hielo. Generalmente el granizo tiene un diámetro que oscila entre 5 y 50 milímetros¹.

Para que se produzca la precipitación de granizo deben existir ciertas condiciones atmosféricas, como ser: la atmósfera debe ser inestable, es decir, deben existir condiciones de temperatura y humedad que permitan el desarrollo de tormentas eléctricas, con fuertes corrientes ascendentes. En una atmósfera inestable, una cierta cantidad de aire en ascenso continúa acelerándose en la misma dirección. A veces, la aceleración prosigue hasta que el aire alcanza la base de la estratósfera. De este modo, pueden formarse tormentas muy grandes y violentas, que producen intensas lluvias, rayos, y a veces, granizo.

El granizo se origina dentro de nubes cumulonimbus de tormenta, de gran desarrollo vertical (desde 2000 a 4000 metros), que presenta poderosas corrientes ascendentes de aire húmedo, y en cuya parte superior la temperatura es menor a cero grado. Cuando el aire caliente se encuentra por encima de las masas más frías, comienza el enfriamiento y concomitante condensación del vapor de agua en gotitas de agua. Esta condensación calienta el aire circundante por el calor latente, haciendo avanzar el ascenso de las masas de aire. Así una gota de agua dentro de la nube, va ascendiendo conforme transcurre el tiempo, ganando dimensión, hasta congelarse dentro, y conformarse de varias capas de hielo y caer en forma de granizadas o tormentas de

¹Cifras provenientes del Sitio Web de DNM

granizo.

Las bolas de granizo suelen ser pequeñas, de algunos milímetros de diámetro. Sin embargo, de vez en cuando se originan bolas mucho mayores, de varios centímetros de diámetro debido a que en la circulación ciclónica de la tormenta, las pequeñas bolas ascienden y descienden varias veces formándose distintas capas de hielo, unas sobre otras. Debido a su proceso de formación todas ellas tienen una estructura en forma de capas de cebolla.

Los lugares típicos de gran formación de estas nubes se encuentran, en las zonas templadas, alrededor de una línea de frente frío, cerca de los océanos donde la brisa marina puede proveer energía a la tormenta, o en montañas en las laderas de barlovento donde el viento se ve forzado a elevarse ocasionando que el aire más caliente (menos denso) ascienda dando origen a fuertes precipitaciones y tormentas. Dicho evento ocurre tanto en verano como en invierno, aunque el caso es más frecuente cuando está presente la canícula, días del año en los que es más fuerte el calor. Los cumulonimbos sirven para equilibrar, dentro de la zona intertropical, las pequeñas áreas de inestabilidad que se originan por la insolación. Siempre generan su propia energía por la acumulación de calor en un área mucho más extendida que la propia base del cumulonimbo.

4.2. Efectos

Dependiendo de su tamaño, las bolas de granizo pueden no ser más que una pequeña molestia, causar daños en automóviles y estructuras de cristal, pudiendo causar heridas a las personas. En el ámbito rural puede causar mucha destrucción, pudiendo producir pérdidas totales o parciales de cultivos.

Una particularidad del granizo muy importante es que este fenómeno jamás se realiza en forma simultánea sobre grandes áreas territoriales, sino que las granizadas se presentan sobre pequeñas áreas a la vez. Esta peculiaridad hace que este fenómeno sea más difícil de observar y consecuentemente más difícil de predecir.

Territorio uruguayo

5.1. Geografía

5.1.1. Superficie

Uruguay posee una superficie terrestre de 176.215 km^2 (de la que 175.016 km^2 es la suma total de los departamentos y 1.199 km^2 comprende la suma de los lagos artificiales del río Negro).

5.1.2. Relieve

El relieve está vinculado en la parte sur a las tierras pampeanas y está constituido por vastas llanuras onduladas y surcadas por colinas de escasa elevación llamadas cuchillas. Las más importantes son las que pertenecen a la Cuchilla de Haedo y a la Cuchilla Grande.

- El punto más elevado de Uruguay es el Cerro Catedral de unos 514 msnm^1 . ($513,66 \text{ msnm}$) ubicado en la Sierra Carapé, al norte del departamento de Maldonado.
- El punto más bajo del país es sobre la costa al nivel del mar.

¹Metros sobre el nivel del mar

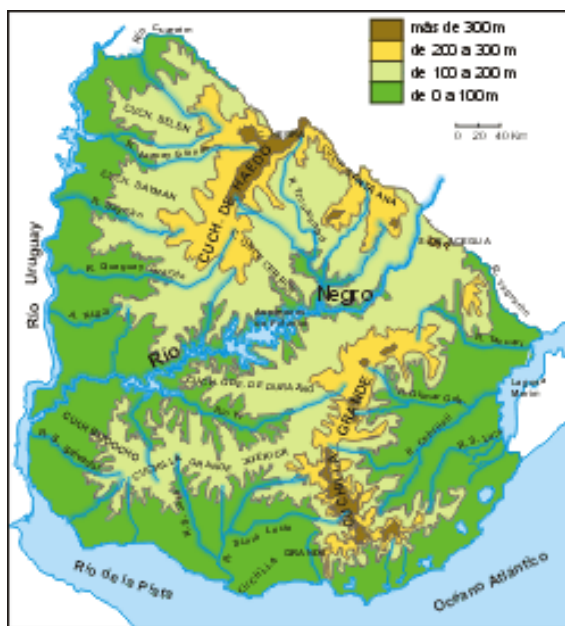


Figura 5.1: Mapa físico de Uruguay

5.1.3. Puntos extremos

Uruguay es el país de posición más meridional en América, junto con Argentina y Chile, que se prolongan mucho más al sur. Su área territorial está comprendida aproximadamente entre los paralelos 30° y 35° de latitud sur y los meridianos 53° y 58° de longitud oeste. Si se incluye el mar territorial, las coordenadas son 30° y 38° de latitud sur y 50° y 58° de longitud oeste.

Latitud y longitud

- El punto más septentrional de Uruguay se encuentra sobre el río Cuareim, próximo al arroyo Yacot (Departamento de Artigas). Sus coordenadas son: 30°05'08" S (-30.08556) 56°57'06" O (-56.95167).
- El punto más austral del país se encuentra en el extremo sur de la península de Punta del Este (Departamento de Maldonado). Sus coordenadas son: 34°58'27" S (-34.97417) 54°57'07" O (-54.95194).

La distancia entre el punto más al norte y el punto más al sur es de 571 km.

- El punto más oriental se encuentra sobre la desembocadura del río Yaguarón en la Laguna Merín (Departamento de Cerro Largo). Sus coordenadas son: 32°39'14" S (-32.65389) 53°10'58" O (-53.1827).

- El punto más occidental de Uruguay se encuentra al sur del río San Salvador (Departamento de Soriano). Sus coordenadas son: 33°31'30" S (-33.525) 58°26'01" O (-58.43361).

La distancia entre el punto más al este y el punto más al oeste es de 500 km.

5.1.4. Clima

El clima en Uruguay es templado y húmedo (promedio 17 °C), con veranos cálidos y precipitaciones más o menos homogéneas durante todo el año. En Uruguay, donde la influencia marítima como también continental se hacen notar, la distribución de lluvias presenta una doble estación lluviosa, con un máximo principal en otoño y un máximo secundario en primavera. Por su latitud, entre 30°S y 35°S, las cuatro estaciones están claramente diferenciadas por la temperatura. La zona situada en el extremo noroeste del país (Artigas, Salto, Rivera) es considerablemente más cálida con una media de aproximadamente 19 °C y un promedio de precipitaciones de unos 1.400 mm anuales. El Sur y Este (Montevideo, Maldonado, Rocha, Lavalleja) en cambio son más frescos con una media de alrededor de 16 °C y 1.000 mm anuales. Siendo un país bajo, el clima está determinado por la latitud y la influencia de las corrientes marinas del Océano Atlántico. La corriente cálida del Brasil incrementa la temperatura de las costas del Atlántico desde fines de enero hasta principios de mayo; la corriente fría de las islas Malvinas enfrían sus aguas desde junio a setiembre. El efecto de ambas determina una temperatura promedio del mar a nivel superficial (Punta del Este) entre 8 °C y 23 °C según la época del año. De febrero a abril, la temperatura del océano es muy agradable y generalmente sensiblemente superior a la que se registra desde junio hasta fines de diciembre, aunque existe importante variación interanual durante el verano.

El frío es por lo general bastante húmedo, muy ventoso con días nublados, el calor no es demasiado seco, más bien húmedo y pesado en la zona sur y más seco en el norte.

El fenómeno climático más frecuente son las heladas meteorológicas, que afectan mayormente a las zonas centro-sur y centro-norte del país. A modo de ejemplo de la variabilidad climática del Uruguay, en los 31 días de un mes de julio pueden registrarse 25 días de heladas como en la ciudad de Florida, 34.1° S 56.2° W, a 54 msnm, a escasos 90 km de Montevideo, (así ocurrió en julio de 2007) o tan solo 6 días (en julio de 2006), esto evidencia una gran variación entre años en la época fría, el verano a diferencia del invierno es más uniforme. El fenómeno de La Niña (año 2007) provoca un invierno uniformemente frío y sequías prolongadas (Florida,

media registrada de julio de 2007 6.8 °C), en tanto que el fenómeno del Niño provoca lluvias e inviernos benignos (Florida, media registrada de julio de 2006 13.1 °C).

El clima del Uruguay es propicio para la producción ganadera a partir de las pasturas naturales. Éstas tienen por lo general una estacionalidad con un pico muy importante en primavera, debido a una combinación óptima de humedad y temperatura y un déficit muy marcado en invierno para cubrir las necesidades nutricionales, debido a la afectación de la calidad y volumen de las pasturas por las heladas meteorológicas.

5.2. Estaciones meteorológicas

En Uruguay existen tres tipos de estaciones de captura de datos de fenómenos climáticos: las estaciones meteorológicas, las estaciones pluviométricas y las estaciones agronómicas.

Existen veinte estaciones meteorológicas, ubicadas en todo el territorio uruguayo. Éstas también suelen denominarse estaciones meteorológicas-pluviométricas. Las mismas se destacan por poseer personal específicamente preparado y calificado para la obtención de los datos de los diferentes fenómenos meteorológicos.

Existen alrededor de doscientas estaciones pluviométricas en todo el territorio uruguayo. En comparación con las anteriores, las mismas no poseen personal calificado para la captura y recolección de datos climáticos, haciendo que la confiabilidad y calidad de los mismos sea baja.

En cuanto a las estaciones agronómicas, existen cinco en todo el país situadas en el área de las estaciones de INIA. Éstas poseen la desventaja de no capturar datos del fenómeno granizo.

Parte III

Marco Metodológico

Interpolación por Kriging

6.1. Introducción

Kriging es un método de interpolación espacial cuya idea principal es la construcción de nuevos puntos partiendo del conocimiento de un conjunto discreto de puntos [4]. G. Matheron (1930-2000) denominó este método de predicción espacial como el método lineal óptimo luego de D.G. Krige, un ingeniero de minas sudafricano quien en 1950, desarrolló métodos empíricos para determinar la verdadera distribución de oro en un mina a partir de datos de localizaciones de oro de una muestra dada. Los aportes dados por Wold (1938), Kolmogorov (1941) y Wiener (1949) contienen nociones de predicciones óptimas lineales que reflejaban que las observaciones cercanas en el tiempo al punto de predicción deberían tener valores de ponderadores altos. En geoestadística¹ se fueron desarrollando estas mismas ideas en términos de distribución espacial.

A continuación se realiza una breve descripción de los conceptos relacionados con la geoestadística.

6.2. Geoestadística

Se le llama *geoestadística* [16] al conjunto de métodos y modelos para el estudio de datos espaciales con las siguientes características:

- Los valores de la variable se observan en un conjunto discreto de localizaciones muestrales dentro de un área determinada.
- Cada observación es una medida de un fenómeno espacial en las correspondientes localizaciones muestrales.

¹Ver definición en sección 6.2

En líneas generales, se puede afirmar que el análisis geoestadístico consta de cuatro etapas:

1. Análisis exploratorio de los datos
2. Análisis estructural (Análisis de dependencia espacial)
3. Modelamiento espacial
4. Estimaciones

6.2.1. Datos georeferenciados

Los datos espaciales son medidas u observaciones que tienen asociadas una localización específica y de forma general se puede referir a ellos como *datos georeferenciados*. Los mismos involucran coordenadas de posicionamiento como la latitud y la longitud.

6.2.2. Coordenadas geográficas

El sistema de coordenadas geográficas es un sistema de referencia que utiliza las dos coordenadas angulares latitud (norte-sur) y longitud (este-oeste) para determinar los ángulos laterales de la superficie terrestre. Estas coordenadas se suelen expresar en grados sexagesimales.

En el presente trabajo se analizan un conjunto de datos georeferenciados a través de coordenadas geográficas. Para ello es necesario tomar en cuenta la forma de la tierra caracterizando la misma según un modelo *Esférico* o un modelo *Elipsoidal*. Dependiendo del modelo utilizado podrán variar las definiciones de longitud y latitud.

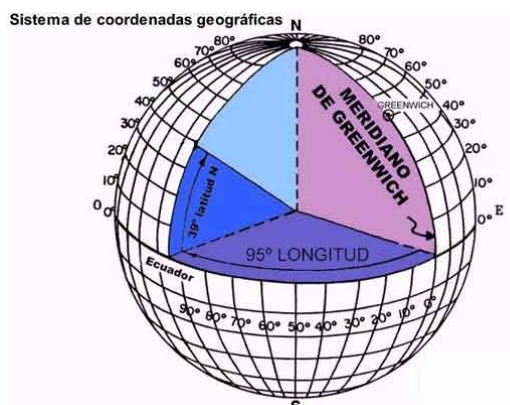


Figura 6.1: Sistema de coordenadas geográficas

Modelo esférico

Las líneas de latitud se llaman paralelos y son círculos paralelos al ecuador en la superficie de la Tierra. Se miden en grados, minutos y segundos, desde 0° en el Ecuador hasta 90° norte o 90° sur. Se consideran grados positivos al norte y negativos al sur.

Las líneas de longitud son círculos máximos que pasan por los polos y se llaman meridianos. Se miden en grados, minutos y segundos desde 0° hasta 180° al este o al oeste de Greenwich, considerando positivas al este y negativas al oeste.

Modelo elipsoidal

Latitud Elipsoidal o geodésica de un punto en la superficie del elipsoide, es el ángulo que la normal al elipsoide en ese punto forma con su proyección ecuatorial. Se mide a partir del plan ecuatorial, positiva en el hemisferio norte y negativa en el sur.

Longitud elipsoidal o Geodésica se define igual que la longitud esférica.

En el presente trabajo se trabaja con un sistema de coordenadas esféricas.

6.2.3. Proyección geográfica

La proyección geográfica es un sistema de representación gráfico que establece una relación ordenada entre los puntos de la superficie curva de la Tierra y los de una superficie plana. Estos puntos se localizan auxiliándose en una red de meridianos y paralelos, en forma de malla.

En un sistema de coordenadas proyectadas, los puntos se identifican por las coordenadas (x, y) en una malla cuyo origen depende de los casos. Este tipo de coordenadas se obtienen matemáticamente a partir de las coordenadas geográficas (longitud y latitud), que son no proyectadas.

6.3. El modelo espacial general

Cualquier proceso espacial puede caracterizarse mediante un proceso estocástico [4]:

$$\{Z(s) : s \in D\}$$

Donde:

- $s \in \mathbb{R}^d$ es una posición genérica localizada en el espacio d-dimensional.
- $Z(s)$ un valor aleatorio localizado en s .
- D un subconjunto de \mathbb{R}^d

De manera general, el modelo $Z(s)$ se descompone en una parte determinística y otra residual de la forma [13]:

$$Z(s) = \mu(s) + \epsilon(s) \tag{6.1}$$

Donde $\mu(s)$, conocida como función de tendencia, da cuenta de la variabilidad a gran escala. A su vez, $\epsilon(s)$ representa el comportamiento local o evolución a pequeña escala, siendo un proceso de media cero, varianza constante y con función de covarianza que depende de la diferencia entre las posiciones en cuestión.

6.3.1. Procesos estacionarios

Se dice que el proceso $\{Z(s) : s \in D\}$ es *estrictamente estacionario* en D si la función de distribución conjunta

$$F_{Z(s_1), \dots, Z(s_n)}(z_1 \dots z_n) = P(Z(s_1) \leq z_1, \dots, Z(s_n) \leq z_n)$$

es invariante ante traslaciones; es decir que dado un conjunto de n puntos $\{s_1, \dots, s_n\}$ y para algún vector h , se cumple que:

$$F_{Z(s_1), \dots, Z(s_n)}(z_1 \dots z_n) = F_{Z(s_1+h), \dots, Z(s_n+h)}(z_1 \dots z_n)$$

Lo que significa que si cambia la configuración de puntos en una determinada dirección, la distribución múltiple no se altera. De otra manera, se puede definir *estacionariedad espacial* cuando las propiedades del modelo a estudiar no cambian sobre

la región de interés D .

En la práctica se utilizan las definiciones de estacionariedad de segundo orden y estacionariedad intrínseca ya que las mismas parten de supuestos más débiles relacionados con los momentos de primer y segundo orden.

Estacionariedad de segundo orden

$\{Z(s) : s \in D\}$ se dice que es estacionario de segundo orden si cumple con las siguientes hipótesis:

1. El momento de primer orden existe y es constante para todo s , o sea:

$$E(Z(s)) = \mu \quad \forall s \in D$$

2. Para cualquier par de variables, $Z(s)$ y $Z(s+h)$ existe la covarianza y sólo depende de la distancia h :

$$\text{Cov}(Z(s), Z(s+h)) = C(h) \tag{6.2}$$

Donde la función $C(\cdot)$ se denomina *covariograma*².

La condición 6.2 implica que la varianza existe, es finita y no depende de s , pues:

$$\sigma^2 = C(0) = \text{Var}(Z(s))$$

Estacionariedad intrínseca

Se dice que $\{Z(s) : s \in D\}$ es intrínsecamente estacionario si satisface:

1. El valor esperado de las diferencias cumple que:

$$E(Z(s+h) - Z(s)) = 0$$

2. La varianza de las diferencias es finita y está dada por:

$$\text{Var}(Z(s+h) - Z(s)) = 2\gamma(h)$$

La función $\gamma(\cdot)$ se denomina *Variograma*³

Por lo anterior, si un proceso espacial tiene varianza finita, eso implica que sus diferencias también tendrán varianza finita; es decir, que si un proceso es estacionario de segundo orden también será estacionario intrínseco.

²Ver definición en apéndice A.4

³Ver sección 6.8

6.4. Tendencia espacial

Los diferentes métodos de interpolación espacial se fundamentan, en la mayoría de los casos, bajo el supuesto de estacionariedad de segundo orden del proceso en cuestión. La presencia de tendencia espacial del fenómeno en estudio indica el no cumplimiento de dicho supuesto, es decir, cuando:

$$E(Z(s)) = \mu(s) \quad \forall s \in D$$

En la presente sección se estudia la estimación de la tendencia en un conjunto de datos espaciales. Una primera aproximación al problema consiste en proponer modelos paramétricos de la tendencia y estimar los coeficientes de la misma planteando un problema de regresión lineal.

6.4.1. Estimación de la tendencia

Sea $\{Z(s_1), \dots, Z(s_n)\}$ un conjunto de n datos conocidos. Los mismos son realizaciones de un proceso espacial y se pueden descomponer como fue visto en el modelo 6.1:

$$Z(s_i) = \mu(s_i) + \epsilon(s_i) \quad \forall i = 1, \dots, n$$

La componente determinística asociada a la tendencia del modelo puede ser modelizada a través de una combinación lineal de ciertas variables de interés:

$$\mu(s_i) = \sum_{j=0}^k \beta_j x_j(s_i) \quad k \ll n$$

Donde $\beta = (\beta_0, \dots, \beta_k)^t$ es un vector desconocido de coeficientes y $\{x_j, j = 0, \dots, k\}$ un conjunto de variables conocidas.

Por lo tanto, el proceso espacial es posible expresarlo de la siguiente manera:

$$Z(s_i) = \sum_{j=0}^k \beta_j x_j(s_i) + \epsilon(s_i) \quad \forall i = 1, \dots, n$$

Puede entonces escribirse el sistema de ecuaciones correspondiente de manera matricial como:

$$Z = X\beta + \epsilon$$

Siendo $Z = \{Z(s_1), \dots, Z(s_n)\}$, X la matriz de diseño, $\beta = (\beta_0, \dots, \beta_k)^t$ y $\epsilon = (\epsilon(s_1), \dots, \epsilon(s_n))^t$.

Suponiendo que los errores son incorrelacionados, de media cero, y varianza constante σ^2 , el mejor estimador lineal insesgado de β , conocido como el estimador por mínimos cuadrados ordinarios (mco) es:

$$\hat{\beta}_{mco} = (X^t X)^{-1} X^t Z \quad (6.3)$$

No obstante, si los errores no tienen varianza constante o no son incorrelacionados, como ocurre en una situación de dependencia espacial, es posible encontrar un mejor estimador de β .

Supóngase entonces que los errores tienen matriz de covarianza $\sigma^2 \Sigma$, siendo Σ la matriz de correlación de los errores. En lugar de la ecuación 6.3, se tiene que la mejor estimación de β es:

$$\hat{\beta}_{mcg} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Z \quad (6.4)$$

El estimador de β en este caso, se conoce como el estimador por mínimos cuadrados generalizados (mcg).

Por lo tanto, se obtiene la estimación de la tendencia de la siguiente manera:

$$\hat{\mu} = X \hat{\beta} = \hat{Z}$$

6.4.2. Implicancias de la existencia de tendencia

La existencia de dependencia espacial, como es el caso de la tendencia, viola el supuesto de estacionariedad de segundo orden. Esto provoca la necesidad de transformar el proceso considerado inicialmente ($Z(s)$), buscando encontrar un nuevo proceso que cumpla con las condiciones de estacionariedad requeridas. Es por ello que se plantea el siguiente modelo:

$$\epsilon(s) = Z(s) - \mu(s)$$

Donde $\epsilon(s)$ representa un proceso estacionario, con esperanza nula y variograma conocido.

Para obtener dicho proceso, es necesario estimar la tendencia modelizada, como se vió anteriormente, y sustraerla a los valores observados de la variable de interés. Una vez estimada la misma, se pueden obtener las estimaciones de los errores de la siguiente manera:

$$e(s_i) = Z(s_i) - \hat{\mu}(s_i) \quad \forall i = 1, \dots, n$$

Donde el proceso de residuos $\{e(s)\} = \{\hat{e}(s)\}$ será el proceso a estimar mediante el procedimiento de interpolación espacial.

6.5. Interpolación espacial

El objetivo de la interpolación espacial [4] es estimar el comportamiento de una variable aleatoria Z de interés, en zonas donde no existen datos a partir de una muestra dada $\{z(s_1), z(s_2), z(s_3), \dots, z(s_n)\}$ en localizaciones conocidas $\{s_1, s_2, s_3, \dots, s_n\}$.

Se estima a (6.1) mediante el siguiente estimador:

$$Z^*(s) = \sum_{i=1}^n \lambda_i Z(s_i) \quad (6.5)$$

Donde:

- $Z^*(s)$ es el valor estimado para la posición s .
- $\{Z(s_i), i = 1, \dots, n\}$ son los valores de que toma Z en las posiciones $\{s_i, i = 1, \dots, n\}$.
- $\{\lambda_i, i = 1, \dots, n\}$ son los ponderadores.

Existen diversos métodos de interpolación espacial como ser: *Estimación del vecino más cercano*, *Estimación del inverso de la distancia*. Estas metodologías son sencillas de aplicar, pero la asignación de los ponderadores sólo depende del criterio de cercanía. El método de *Interpolación por Kriging* incorpora información sobre continuidad espacial y anisotropía, mediante la utilización del *Variograma*.

6.6. Interpolación por Kriging

La interpolación por kriging es un método geoestadístico de estimación de puntos que utiliza un modelo de variograma para la obtención de datos, es decir, parte del principio que puntos próximos al espacio tienden a tener valores más parecidos de los puntos más distantes. La técnica de Kriging asume que los datos recogidos de una determinada población se encuentran correlacionados en el espacio.

El objetivo de la técnica de Kriging es minimizar la varianza del error de predicción sujeto a la restricción de insesgadez del estimador, es decir, se minimiza:

$$\text{Var} (Z^*(s) - Z(s))$$

sujeto a

$$E (Z^*(s) - Z(s)) = 0$$

Así, el estimador Kriging es el **mejor estimador lineal e insegado**.

- *Lineal* porque es una combinación lineal ponderada de los datos.
- *Insegado* porque el error de estimación tiene una media igual a 0, es decir $E(Z^*(s)) = \mu$
- *Mejor* en el sentido del error de varianza mínima para un modelo dado de covarianza/variograma.

El estimador de Kriging, visto en 6.5, depende del modelo que se adopte para la componente determinística:

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

Como consecuencia existen varios tipos de Kriging, que difieren en los supuestos que se apliquen a la tendencia $\mu(s)$:

- *Kriging Simple*: $E(Z(s)) = \mu$ con μ conocida en todo D .
- *Kriging Ordinario*: $E(Z(s)) = \mu$ con μ desconocida.
- *Kriging Universal*: $E(Z(s)) = \mu(s)$. La tendencia $\mu(s)$ modela mediante combinaciones lineales de ciertas funciones de las coordenadas $x_j(s)$:

$$\mu(s) = \sum_{j=0}^k \beta_j x_j(s)$$

Los coeficientes β_j se suponen desconocidos.

A continuación se desarrolla la técnica de *Kriging Ordinario*.

6.7. Kriging Ordinario

Las principales hipótesis que tiene el método de *Kriging Ordinario* son las siguientes:

- La media poblacional de la variable aleatoria de interés es **desconocida**.
- Se conoce el *Variograma*.

El estimador propuesto para $Z(s)$ es el presentado previamente en la ecuación 6.5.

6.7.1. Ecuaciones de Kriging

Para determinar los pesos $\{\lambda_i, i = 1, \dots, n\}$ se examinan las condiciones de insesgamiento y de varianza mínima del estimador.

Condición de insesgamiento

Para que $E(Z^*(s))$ sea insesgado se debe cumplir que:

$$E(Z^*(s)) = E\left(\sum_{i=1}^n \lambda_i Z(s_i)\right) = \sum_{i=1}^n \lambda_i E(Z(s_i)) = \sum_{i=1}^n \lambda_i \mu = \mu$$

Entonces:

$$\sum_{i=1}^n \lambda_i = 1$$

Condición de varianza mínima

La varianza del error de estimación se expresa en función de la covarianza:

$$V(Z^*(s) - Z(s)) = V(Z^*(s)) + V(Z(s)) - 2Cov(Z^*(s), Z(s)) \quad (6.6)$$

$$= V\left(\sum_{i=1}^n \lambda_i Z(s_i)\right) + V(Z(s)) - 2Cov\left(\sum_{i=1}^n \lambda_i Z(s_i), Z(s)\right) \quad (6.7)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov(Z(s_i), Z(s_j)) + C(0) - 2 \sum_{i=1}^n \lambda_i Cov(Z(s), Z(s_i)) \quad (6.8)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(s - s_i) \quad (6.9)$$

Donde:

- $C(\cdot)$ se denomina *covariograma*.
- $C(0) = Cov(Z(s_i), Z(s_i))$
- $C(s_i - s_j) = Cov(Z(s_i), Z(s_j))$

Por lo tanto los ponderadores (λ_i) se obtienen de manera que se resuelva el siguiente problema de optimización:

$$\begin{cases} \min V(Z(s) - Z^*(s)) \\ \text{sujeto a } \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

Este tipo de problemas de minimización con restricciones, se resuelve utilizando la técnica de Multiplicadores de Lagrange ⁴.

De esta manera las ecuaciones de kriging quedan determinadas de la siguiente forma:

- $-\sum_{j=1}^n \lambda_j \gamma(s_i - s_j) + \gamma(s - s_i) - \beta = 0 \quad i = 1, \dots, n$
- $\sum_{i=1}^n \lambda_i = 1$

Donde $\gamma(\cdot)$ es la función de Variograma.

Las ecuaciones anteriores se pueden escribir de forma matricial:

$$\begin{pmatrix} 0 & \gamma(s_1 - s_2) & \cdots & \gamma(s_1 - s_n) & 1 \\ \gamma(s_2 - s_1) & 0 & \cdots & \gamma(s_2 - s_n) & 1 \\ \vdots & & & & \\ \gamma(s_n - s_1) & \gamma(s_n - s_2) & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \beta \end{pmatrix} = \begin{pmatrix} \gamma(s_1 - s) \\ \gamma(s_2 - s) \\ \vdots \\ \gamma(s_n - s) \\ 1 \end{pmatrix}$$

Los valores óptimos para $\lambda_1, \dots, \lambda_n$ pueden obtenerse de la siguiente manera:

$$\hat{\lambda} = A^{-1}b$$

⁴Ver apéndice A.2

Donde:

- $\hat{\lambda} \equiv (\hat{\lambda}_1, \dots, \hat{\lambda}_n, \hat{\beta})'$
- $b \equiv (\gamma(s - s_1), \dots, \gamma(s - s_n), 1)'$
- $A \equiv \begin{cases} \gamma(s_i - s_j) & i = 1, \dots, n \quad j = 1, \dots, n \\ 1 & i = n + 1 \quad j = 1, \dots, n \\ 0 & i = n + 1 \quad j = n + 1 \end{cases}$

6.7.2. Varianza de Kriging

El error de estimación, denominado *Varianza de Kriging* se expresa de la siguiente manera:

$$\sigma^2(s) = V(Z(s) - Z^*(s)) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(s - s_i)$$

Si se considera la siguiente relación $C(h) = C(0) - \gamma(h)$ lo anterior se puede escribir de la siguiente manera:

$$\begin{aligned} \sigma^2(s) &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (C(0) - \gamma(s_i - s_j)) + C(0) - 2 \sum_{i=1}^n \lambda_i (C(0) - \gamma(s - s_i)) \\ &= \underbrace{\sum_{i=1}^n \lambda_i}_{=1} \underbrace{\sum_{j=1}^n \lambda_j}_{=1} C(0) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(s_i - s_j) + C(0) - 2 \underbrace{\sum_{i=1}^n \lambda_i}_{=1} C(0) + 2 \sum_{i=1}^n \lambda_i \gamma(s - s_i) \end{aligned}$$

Por lo tanto, la varianza de Kriging se expresa de la siguiente manera:

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

6.7.3. Kriging Residual

Se denomina Kriging Residual a la técnica de Kriging Ordinario bajo el no cumplimiento del supuesto de estacionariedad de segundo orden del proceso, es decir ante la presencia de tendencia espacial en los datos.

Si se asume el modelo 6.1, es importante observar que:

$$\epsilon(s) = Z(s) - \mu(s)$$

Donde $\epsilon(s)$ es un proceso estacionario de segundo orden con media cero y por lo tanto se puede utilizar la técnica de kriging Ordinario en los mismos.

Por lo mencionado anteriormente, la estimación de $Z(s)$ resulta de añadir a los residuos interpolados ($\hat{\epsilon}(s)$) la tendencia estimada.

6.8. Variograma

El variograma es una herramienta que permite analizar el comportamiento espacial de una variable sobre una zona dada. Esta herramienta permite, entre otras cosas:

- Detectar zonas de anisotropía⁵.
- Detectar el entorno de interpolación (asociado a la correlación espacial).
- Determinar la variabilidad espacial en función de la distancia.

6.8.1. Variograma teórico

El variograma teórico se define como:

$$2\gamma(h) = \text{Var} (Z(s + h) - Z(s)) \quad \forall s \in \mathfrak{R}^d, \forall h \in \mathfrak{R}^d$$

Si además, el proceso es intrínsecamente estacionario:

$$2\gamma(h) = E (Z(s + h) - Z(s))^2 \quad \forall s \in \mathfrak{R}^d, \forall h \in \mathfrak{R}^d$$

A la expresión anterior se la puede interpretar como el valor promedio de la diferencia al cuadrado de los valores que toma la variable de interés en dos puntos separados por una distancia h .

El término $2\gamma(\cdot)$ se denomina *variograma*, mientras que el término $\gamma(\cdot)$ es denominado *semivariograma*. Para poder implementar la técnica de *kriging* solo es necesario la función $\gamma(\cdot)$, por ello algunos autores han llamado variograma a $\gamma(\cdot)$.

El variograma se obtiene graficando $\gamma(\cdot)$ en función de los *lags* o h que se obtengan para una dirección dada o para todas las direcciones (omnidireccional).

⁵Ver definición en apéndice A.5

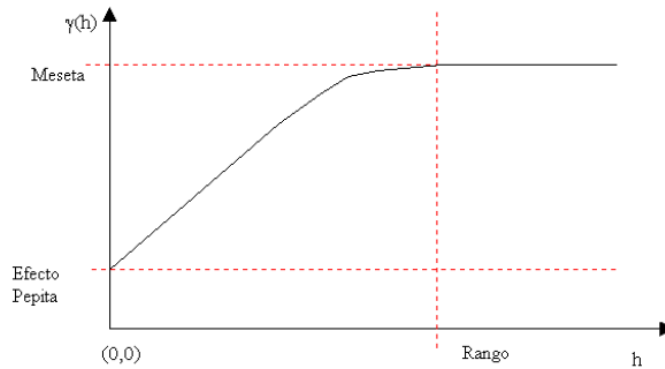


Figura 6.2: Ejemplo de Variograma

Propiedades del variograma

1. $\gamma(\cdot)$ es independiente de la localización s
2. $\gamma(\cdot)$ depende del módulo y dirección de h
3. $\gamma(0) = 0$ pero puede presentar salto en $h = 0$ denominado *efecto pepita*
4. $\gamma(-h) = \gamma(h) \quad \forall h$
5. $\gamma(h) \geq 0 \quad \forall h$
6. Condicionalmente semidefinido negativo:

Para $n \geq 1$ y $\forall \{a_i\}_{i=1}^n \subset \Re$ y $\sum a_i = 0$:

$$\sum \sum a_i a_j \gamma(s_i - s_j) \leq 0$$

7. $a\gamma(\cdot)$ es un semivariograma en $\Re^d \quad \forall a \geq 0$
8. $\gamma_1 + \gamma_2$ es semivariograma en \Re^d
9. El *rango* es la distancia a la cual el variograma se estabiliza o, dicho de otra forma, distancia a partir de la cual no hay correlación.
10. El *sill* o *meseta* es el valor constante que toma el variograma en distancias mayores al rango (varianza de la variable aleatoria Z), o sea, si el variograma es una función acotada entonces:

$$\gamma(h) \rightarrow \sigma^2 \quad \text{si } \|h\| \rightarrow \infty$$

El variograma puede presentar diferentes comportamientos según se lo estudie a pequeñas distancias o a grandes distancias, es por ello que el mismo puede dar lugar a distintas interpretaciones⁶.

6.8.2. Variograma experimental

El variograma experimental es la estimación del variograma teórico. El mismo es construido a partir de los datos muestrales.

Bajo el supuesto de que el proceso aleatorio $Z(s)$ tiene media constante (supuesto de proceso estacionario), el estimador del variograma que se obtiene por el método de los momentos es el siguiente:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{|s_i - s_j| = |h|} (z(s_i) - z(s_j))^2$$

donde

$$N(h) = \{(s_i, s_j) : |s_i - s_j| = h; \quad i, j = 1, \dots, n\}$$

Siendo $N(h)$ el número de pares de coordenadas que distan a una distancia h .

Notar que $N(-h) \neq N(h)$, pero $2\hat{\gamma}(-h) = 2\hat{\gamma}(h)$ preservando las propiedades teóricas del variograma.

Notar también que no es necesario estimar μ . Por esta razón este estimador se conoce como estimador *clásico*.

Además, si $Z(\cdot)$ es intrínsecamente estacionaria entonces el estimador $2\hat{\gamma}(h)$ es insesgado.

Para la obtención del variograma experimental se deben considerar los siguientes puntos:

- Escoger una distancia (lag/rezago) h
- Escoger una dirección θ
- Calcular $\gamma(h)$ para los valores de $h, 2h, 3h, \dots, nh$
- Graficar $\gamma(h)$ versus los valores de $h, 2h, 3h, \dots, nh$

⁶Ver interpretación variograma en apéndice A.3.1

Existen diferentes formas de determinar las distancias h y las direcciones θ , para ello existen las clases de distancias y las clases de direcciones⁷.

6.8.3. Modelos de variograma

El variograma experimental no puede ser evaluado en distancias o direcciones intermedias. Una interpolación entre los puntos del variograma experimental no garantiza la existencia y unicidad de la solución del sistema de Kriging, ya que las funciones matemáticas de los modelos deben permitir que la matriz de varianzas y covarianzas pueda ser invertida para determinar los pesos para la interpolación por Kriging. Además, el variograma experimental no satisface las condiciones que todo variograma debe satisfacer.

Es por ello que al variograma experimental se le ajusta un modelo de variograma teórico adecuado. Los modelos de variograma isotrópicos más comunes son: *Efecto pepita puro*, *Esférico*, *Exponencial*, *Gaussiano*, *Cúbico*, *Seno cardinal*, *Potencia*, entre otros⁸.

Se selecciona el modelo de variograma tal que se minimice el error de estimación de Kriging.

6.8.4. Variograma bajo estructura de dependencia espacial

Los modelos de variogramas vistos anteriormente son construídos bajo el cumplimiento del supuesto de estacionariedad de segundo orden del proceso en cuestión.

La existencia de dependencia espacial, como es el caso de tendencia del proceso, provoca que los variogramas experimentales deban ser construídos con el proceso de los residuos visto en sección 6.4.2.

Es por esta razón que el variograma teórico de los residuos es definido de la siguiente manera:

$$2\gamma(h) = E(\epsilon(s+h) - \epsilon(s))^2 \quad \forall s \in \mathfrak{R}^d, \forall h \in \mathfrak{R}^d$$

Por lo tanto, la estimación del variograma de los residuos, es decir el variograma experimental, puede ser expresado:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{|s_i - s_j| = |h|} (e(s_i) - e(s_j))^2$$

⁷Ver apéndice A.3.2

⁸Ver apéndice A.3.6

donde

$$N(h) = \{(s_i, s_j) : |s_i - s_j| = h; \quad i, j = 1, \dots, n\}$$

6.9. Validación Kriging

La validación del método de Kriging es un proceso en el cual se busca optimizar las estimaciones resultantes, es decir minimizar el error de estimación. Dicho proceso consiste en analizar, para cada modelo de variograma, los errores cometidos por dicho método. Para ello, existen diferentes técnicas:

- **Cross Validation (leaving one out):** se estima sucesivamente mediante Kriging cada observación conocida considerando las observaciones restantes.
- **Jack-Knife:** se divide la muestra inicial en dos partes y se estima una parte a partir de la otra.

Luego, se realiza un análisis de los errores cometidos para saber si la estimación por kriging fue satisfactoria, es decir, buena precisión, poco sesgo.

Existen ciertos criterios que pueden colaborar a decidir cuál es el mejor estimador de Kriging, es decir, el que posee mayor precisión.

Los mismos son:

- *Medias de los errores:* deben ser cercanos a cero, esto proporciona un estimador sin sesgo.
- *Varianza de los errores:* debe ser la más baja posible, esto proporciona un estimador preciso.
- *Nube de dispersión entre valores reales y estimados:* los puntos deben acercarse a la diagonal $x = y$.

En el presente trabajo se aplica el método de *cross validation o leaving one out*. El mismo es el que se detalla a continuación.

Para calcular el error de estimación cometido por el método se utiliza el siguiente procedimiento:

Dadas las n observaciones, $Z(s_1), Z(s_2), \dots, Z(s_n)$, en cada punto i se procede de la siguiente manera:

6.9. Validación Kriging

1. Se excluye la observación i y se estima con el conjunto de las $n-1$ observaciones restantes.
2. Se calcula el error de estimación cometido en cada punto:

$$e(i) = (Z^*(s_i) - Z(s_i))^2 \quad \forall i = 1, \dots, n$$

Se estima así el error cuadrático medio cometido por el método de la siguiente manera:

$$\hat{E} = \sqrt{\frac{\sum_{i=1}^n e(i)}{n}}$$

\hat{E} sirve como indicador para medir la exactitud del método. Se calcula \hat{E} para cada modelo de variograma, buscando optimizar los resultados obtenidos.

Análogamente, si se trabaja bajo dependencia espacial, el análisis de cross-validation se realiza con los residuos:

$$r(i) = (\epsilon^*(s_i) - \epsilon(s_i))^2 \quad \forall i = 1, \dots, n$$

Por lo tanto,

$$\hat{E} = \sqrt{\frac{\sum_{i=1}^n r(i)}{n}}$$

Donde $r(i)$ representa el error de estimación de Kriging para la observación i .

Análisis de Grupos Difusos

7.1. Análisis de Grupos

El análisis de grupos [2] tiene por objetivo conformar grupos de acuerdo a ciertas características que pueden ser de interés, de manera de que dentro cada grupo los elementos sean lo más homogéneos posible. La homogeneidad (y heterogeneidad) estará medida a través de una distancia pre-definida.

La clasificación obtenida dependerá de las variables seleccionadas, pudiendo variar la estructura de grupos en el caso de que se consideren otras variables diferentes.

Los métodos de clasificación pueden ser *jerárquicos* o *no jerárquicos*. En los métodos jerárquicos, la agrupación se realiza mediante un proceso de agrupación (o desagrupación) sucesiva, cuyo resultado final es una jerarquía de unión completa en la que cada grupo se une (o separa) en una determinada fase. En cambio, los métodos no jerárquicos, permiten la reasignación de elementos, es decir que las observaciones pueden variar de grupo en cada iteración. Éstos se caracterizan por necesitar el número de grupos a priori.

Ambos análisis no son competitivos, sino que se complementan, ya que el análisis de grupos jerárquico puede pensarse como una primer exploración de la estructura de los datos, mientras que el análisis de grupos no jerárquico se verá como una afinación de dichos resultados.

Dentro de los métodos no jerárquicos, se pueden encontrar diferentes técnicas como ser *k-means* ó *Fuzzy*¹, entre otras.

La gran diferencia entre las metodologías no jerárquicas mencionadas anteriormente,

¹La palabra *Fuzzy* es una palabra de origen inglés que se puede traducir como *Difuso/a*

radica en que *Fuzzy* posee la característica de que una observación puede pertenecer a varios grupos a la vez, a través de un *grado de pertenencia* a cada uno de los grupos. Esta característica hace que la técnica *Fuzzy* sea de gran interés.

En el presente trabajo se desarrolla la metodología de *Análisis de Grupos Difusos* (*Fuzzy Clustering Analysis*), pues se considera de gran interés el aporte que brindan los grados de pertenencia de las observaciones a los grupos, lo cual le asigna al análisis la posible superposición entre los grupos.

7.2. Análisis de Grupos Difusos

La técnica de Grupos Difusos puede verse como una extensión del análisis de Grupos Jerárquico, en el sentido que las observaciones pueden pertenecer a varios grupos al mismo tiempo, con diferentes grados de pertenencia, mientras que el análisis de grupos jerárquico asigna cada observación a un único grupo.

Sea $\{x_1, x_2, \dots, x_n\}$ un conjunto de n datos, donde $x_k = (x_{1k}, x_{2k}, \dots, x_{nk})^t$ con $x_k \in \mathfrak{R}^n$, siendo $k = 1, \dots, K$ el número de variables que participan del análisis.

La matriz de datos se conforma de la siguiente manera :

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{pmatrix}$$

El algoritmo de cluster fuzzy, divide a las observaciones de X en C grupos difusos, formando una partición de los datos que puede ser representada por una matriz U , cuyos elementos $u_{iv} \in [0, 1]$ representan el grado de pertenencia de la observación i al grupo v . Así, la i -ésima fila de U contiene los valores de los coeficientes de pertenencia correspondientes de la v -ésima partición.

El algoritmo tiene por objetivo minimizar la siguiente función:

$$J(X; U, V) = \frac{\sum_{v=1}^C \sum_{i,j} (u_{iv})^m (u_{jv})^m d(i, j)}{2 \sum_j (u_{jv})^m}$$

Donde:

- $V = [v_1, v_2, \dots, v_C]$ con $v_s \in \mathfrak{R}^K$ centroide del grupo s con $s = 1, \dots, C$.
- $m \in (1, \infty)$ es un ponderador llamado *fuzzifier*², que determina el grado de superposición que se desea de los grupos.
- $d(i, j)$ es la distancia³ entre las observaciones i y j .

El problema de optimización es el siguiente:

$$\begin{cases} \min J(X; U, V) \\ \text{sujeto a } \sum_{v=1}^C u_{iv} = 1, \forall i \end{cases}$$

El algoritmo de iteración comienza cuando se le proporciona un número a priori de grupos C , así como el grado de superposición que se desea de los grupos (m) y una determinada regla de detención del algoritmo. Al inicio, en el momento $t = 0$, se genera una partición aleatoria U^0 . En el segundo paso del algoritmo, dados los grados de pertenencia del paso anterior, se calculan los centroides de los grupos v_c^t para $c = 1, \dots, C$ de la siguiente forma:

$$v_c^t = \frac{\sum_{i=1}^n (u_{ic}^t)^m x_i}{\sum_{i=1}^n (u_{ic}^t)^m}$$

Dados los nuevos centroides v_c^t , se recalculan los coeficientes de pertenencia:

$$u_{ic}^{t+1} = \frac{1}{\sum_{v=1}^C \left(\frac{d^2(x_i, v_c^t)}{d^2(x_i, v_v^t)} \right)^{\frac{2}{m-1}}}$$

El proceso se detiene cuando se cumple la regla de detención $|U^{t+1} - U^t| \leq \epsilon$, o cuando el algoritmo llega a un número predeterminado de pasos.

Notar que, cuando $m \rightarrow 1$ los grupos conformados tienden a solparse, ya que los grados de pertenencia son más parecidos entre sí. Mientras que cuanto más grande sea m , los grupos se conforman con observaciones con altos grados de pertenencia.

²*Fuzzifier* es una palabra de origen inglés que puede ser traducida como *Difusor*

³Ver apéndice B.1

7.3. Incorporación de información espacio-temporal

Sin perder de vista el objetivo planteado, el presente análisis toma mayor interés si se incorpora tanto información espacial como temporal entre las observaciones.

Los aspectos más relevantes a tener en cuenta son:

- La información espacial, dada por la naturaleza espacial de las observaciones a clasificar.
- La información temporal, proveniente de las trayectorias de las series de tiempo de cada observación.

7.3.1. Información espacial

La información espacial es analizada mediante una matriz de proximidad espacial o geográfica y la autocorrelación espacial entre las observaciones. La misma es incorporada al análisis mediante la introducción de un adecuado factor a la función objetivo, de manera que incorpore a la técnica la distancia espacial entre las observaciones en cuestión.

La forma analítica de incorporar la información geográfica es sumar a la función objetivo mencionada anteriormente un factor de penalización. Cuando una observación i pertenece a un grupo c con un alto grado de pertenencia, entonces este factor ‘forza’ a que las observaciones vecinas tengan también un alto grado de pertenencia a dicho grupo c , tanto como sea posible, aunque no asegura que dichas observaciones finalmente terminen estando en un mismo grupo.

El término de penalización es el siguiente:

$$J(X; U, P) = \frac{\beta}{2} \sum_{i=1}^n \sum_{v=1}^C (u_{ic})^m \sum_{j=1}^n \sum_{c \in v_c} p_{ij} (u_{jc})^m$$

Donde:

- P es una matriz cuyos elementos son de la forma p_{ij} , que toman valor 1 si la obs. i es vecina a la obs. j , y 0 en otro caso.
- v_c es el centroide del grupo c .

7.3.2. Información temporal

La información temporal, es decir, la información proporcionada por las trayectorias de las diferentes observaciones, es incorporada al análisis a través de la utilización de una medida de disimilaridad apropiada como son las distancias propuestas por Carlier [3].

Carlier propuso el cálculo de distancias entre trayectorias de observaciones en distintas ocasiones o momentos. Dentro de esas distancias se detallan con más importancia tres de ellas: la distancia instantánea (o transversal), la distancia longitudinal (o evolutiva) y la distancia mixta, la cual toma en cuenta tanto las distancias transversales como longitudinales entre trayectorias.

A continuación se detallan los cálculos de cada una de las distancias mencionadas anteriormente:

Distancia de Carlier Instantánea

$$d_{il}^2(\text{inst}) = \sum_{t=1}^T \sum_{j=1}^J (x_{ijt} - x_{ljt})^2$$

Distancia de Carlier Longitudinal

$$d_{il}^2(\text{long}) = \sum_{t=2}^T \sum_{j=1}^J [(x_{ijt} - x_{lj(t-1)}) (x_{ljt} - x_{ij(t-1)})]^2$$

Donde para las dos clases de distancias i indica la observación, j la variable, t el tiempo y x_{ijt} indica el valor que toma la variable j en la estación i para el tiempo t .

Distancia de Carlier Mixta

$$d_{il}^2 = \alpha_1 d_{il}^2(\text{inst}) + \alpha_2 d_{il}^2(\text{long})$$
$$\alpha_1 + \alpha_2 = 1$$

Las críticas principales a la distancia mixta son:

- La elección de los pesos es subjetiva.

7.3. Incorporación de información espacio-temporal

- Se consideran T sumandos en la distancia instantánea, mientras que en la longitudinal se consideran $T-1$ sumandos.

En el presente trabajo se implementa la técnica de Análisis de Grupos Difusos, adecuando esta metodología a los aspectos de interés anteriormente mencionados.

Parte IV

Resultados y conclusiones

Capítulo 8

Resultados

8.1. Análisis exploratorio de los datos

A continuación se realiza una breve descripción de los datos utilizados en el presente trabajo.

Se dispone de una base de datos que contiene información de ocurrencia de granizo anual a nivel de 19 estaciones meteorológicas de Uruguay en un período comprendido desde el año 1998 a 2008 inclusive¹, así como otras variables asociadas a este fenómeno. En el siguiente cuadro se detallan las estaciones meteorológicas en estudio y los departamentos a los cuales pertenecen.

Estación Meteorológica	Departamento
Artigas	Artigas
Bella Unión	Artigas
Carrasco	Montevideo
Colonia	Colonia
Durazno	Durazno
Melilla	Montevideo
Melo	Cerro Largo
Mercedes	Soriano
Paso de los Toros	Tacuarembó
Paysandú	Paysandú
Prado	Montevideo
Punta del Este	Maldonado
Rivera	Rivera
Rocha	Rocha
Salto	Salto
Tacuarembó	Tacuarembó
Treinta y Tres	Treinta y Tres
Trinidad	Flores
Young	Río Negro

Cuadro 8.1: Estaciones meteorológicas en estudio

¹Base otorgada por SURCO Seguros y DNM

8.1. Análisis exploratorio de los datos

Dichas estaciones se encuentran irregularmente espaciadas en todo el territorio; doce se encuentran ubicadas en la zona litoral oeste del país, tres de ellas se encuentran posicionadas en la zona metropolitana y solo cuatro cubren el área este del país. A continuación se detallan gráficamente las ubicaciones geográficas de las estaciones meteorológicas.

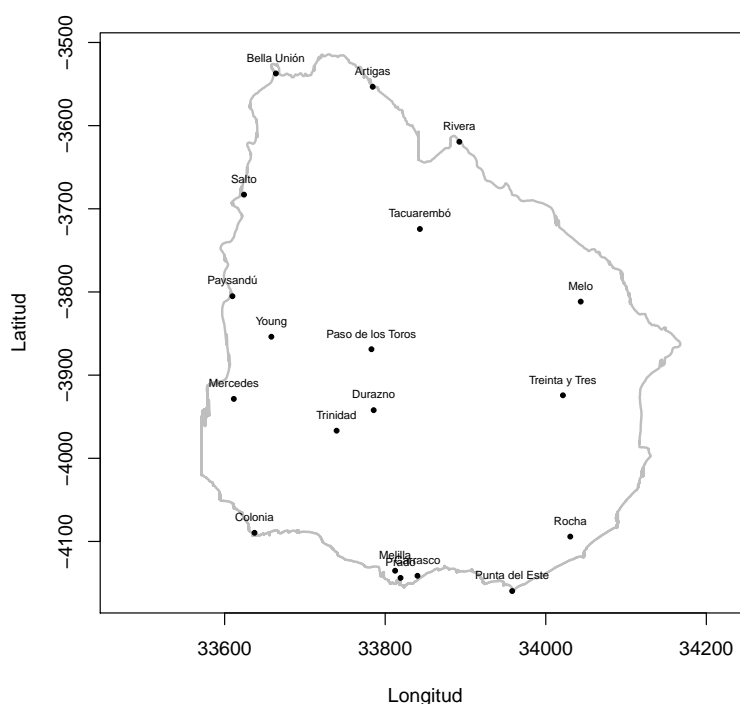


Figura 8.1: Ubicación geográfica estaciones meteorológicas

La distancia máxima se observa entre las estaciones meteorológicas de Bella Unión y Punta del Este, siendo la misma de 688 km. La menor distancia es de 10.9 km, observada entre las estaciones de Melilla y Prado.

Para la primera metodología, *Interpolación por kriging*, se utiliza como información el promedio anual de granizadas en el período mencionado, para cada estación meteorológica, así como sus coordenadas geográficas de longitud y latitud para localizarlas espacialmente.

Para la metodología de *Análisis de Grupos Difusos*, se trabaja con información de la frecuencia acumulada de granizada en el período mencionado y tamaño de piedra de granizo máximo, para las 19 estaciones meteorológicas. Con dicha información se construye una matriz de disimilaridad utilizando las distancias de *Carlisle In-*

stantánea, Longitudinal y Mixta, las cuales son el insumo para realizar el presente análisis.

Cabe destacar que las variables longitud y latitud están inicialmente expresadas en coordenadas geográficas², donde las mismas se encuentran expresadas en grados decimales. No obstante, son proyectadas a una superficie plana mediante la proyección esférica de *Mercator* [11], obteniendo coordenadas cartesianas.

8.1.1. Análisis descriptivo

A efectos descriptivos se analizan en primer lugar las correlaciones lineales entre las variables en estudio y las variables de posicionamiento geográfico, longitud y latitud. En el siguiente cuadro se observan dichas correlaciones.

	Longitud	Latitud	Tamaño máx piedra	Promedio granizo
Longitud	1.00	-0.25	-0.28	-0.09
Latitud		1.00	0.58	0.56
Tamaño máx piedra			1.00	0.72
Promedio granizo				1.00

Cuadro 8.2: *Matriz de correlaciones lineales*

La correlación más alta se detecta entre las variables promedio de granizo anual y tamaño máximo de piedra. Lo anterior es razonable ya que el tamaño máximo de piedra es consecuencia de la ocurrencia del granizo. Asimismo, se observa una alta correlación entre el promedio anual de granizadas y la variable latitud, con un valor de 0.56. La correlación entre longitud y el granizo es baja y se encuentra en el entorno de cero, tomando un valor de -0.09.

Asimismo, se presentan las medidas de resumen de las variables de interés en el presente estudio:

	Tamaño máximo de piedra	Promedio anual de granizo
Mínimo	3.91	0.636
1er cuartil	6.80	1.137
Mediana	10.18	1.636
Media	10.50	1.708
3er cuartil	13.81	2.136
Máximo	19.36	3.091

Cuadro 8.3: *Medidas de resumen variables de interés*

²Mencionadas en sección 6.2.2

8.1. Análisis exploratorio de los datos

La variable promedio anual de granizo presenta un mínimo de 0.636 a lo largo del período analizado. Dicho mínimo fue observado en la estación meteorológica de Melilla. El máximo de dicha variable de 3.091 granizadas anuales promedio en el mismo período y fue observado en la estación de Paso de los Toros.

En cuanto al tamaño máximo de piedra, su valor mínimo de 3.91 mm fue observado en la localidad de Punta del Este y el máximo observado de 19.36 mm está localizado en Colonia.

Se calculan los cuartiles de la variable promedio anual de granizo y los mismos se detallan en el siguiente cuadro:

Cuartiles				
0 %	25 %	50 %	75 %	100 %
0.636	1.1365	1.636	2.1365	3.091

Cuadro 8.4: *Cuartiles del promedio anual de caída de granizo*

A continuación se presenta el gráfico de las estaciones meteorológicas en función de los cuartiles del promedio anual de granizadas.

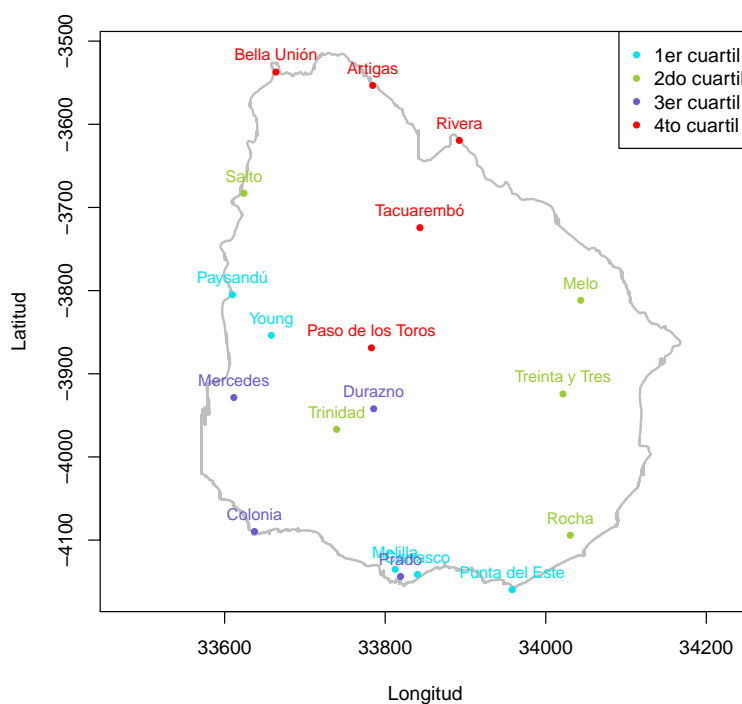


Figura 8.2: *Promedio anual de granizada según cuartiles*

8.1. Análisis exploratorio de los datos

Se observa que en promedio la mayor caída de granizo ocurrió en la región centro-norte del país, mientras que en las zonas cercanas a la costa oceánica se observa menor ocurrencia del fenómeno estudiado.

8.2. Interpolación por Kriging Ordinario

A continuación se detallan los resultados obtenidos mediante la aplicación de la metodología de Interpolación por Kriging. Se analizan cuatro escenarios, desarrollándose en detalle el primero de ellos por ser el que mejor se ajusta a los datos³.

En primer lugar, se analiza si el proceso es estacionario, y para ello se estudia la posible existencia de tendencia en los datos. Se construye el variograma experimental y luego se ajusta el modelo de variograma óptimo para este caso. La información mencionada anteriormente se utiliza como insumo para la estimación de Kriging.

En el presente estudio $Z(s_i)$ representa el promedio anual de granizadas en el período de 1998 a 2008 para la estación meteorológica i localizada geográficamente en las coordenadas s_i , con $s \in \mathbb{R}^2$, siendo $s = (x, y)$ con $x = longitud$ e $y = latitud$.

8.2.1. Análisis de Tendencia

Se desea testear la posible existencia de dependencia espacial del promedio anual de caída de granizo en el territorio uruguayo. El análisis de tendencia se implementa utilizando como información principal a las variables relacionadas con la ubicación espacial de las estaciones meteorológicas. Es decir, se analiza si existe tendencia espacial de granizo a medida que nos desplazamos en dirección sur-norte ó en dirección oeste-este.

Es así como se realiza un *Modelo de Regresión Lineal* donde se explique el promedio de caída de granizo anual en función de las variables latitud (indica dirección sur-norte) y longitud (indica dirección oeste-este).

En una primera instancia se consideran diferentes modelos de regresión de primer y de segundo orden en las variables explicativas, contemplando las distintas combinaciones entre las mismas. Se analiza en cada caso la significación del modelo, así como la significación de las variables utilizadas⁴. Finalmente se decide utilizar un modelo de regresión lineal de primer orden sin coeficiente independiente:

$$Z_i = \beta_1 Longitud_i + \beta_2 Latitud_i + \epsilon_i \quad \forall i = 1, \dots, 19$$

Donde Z_i representa el promedio anual de granizo para la estación i .

³Ver distintos escenarios en apéndice C.1

⁴Ver apéndice A.1

8.2. Interpolación por Kriging Ordinario

Se realizan las estimaciones de los parámetros por mínimos cuadrados ordinarios y los resultados son los que se detallan a continuación:

	Estimación	Error estándar	Estadístico t	p-valor
Longitud	0.0003	0.0001	3.43	0.0032
Latitud	0.0019	0.0007	2.79	0.0126

Cuadro 8.5: *Estimaciones, errores estándar y estadístico de prueba modelo regresión*

Mediante la aplicación de la prueba de significación de parámetros, se concluye que ambos parámetros son significativos al 5 %, pues su p-valor es menor que el nivel de significación $\alpha=0.05$. Esto da la pauta de existencia de tendencia en ambas direcciones, o sea, tanto para dirección sur-norte como para dirección oeste-este.

Al realizar la prueba de significación del modelo resulta que el mismo es significativo al 5 %, siendo su p-valor menor que el nivel de significación predeterminado (5 %). El R^2 ajustado es el mayor de entre todos los modelos testeados de primer y segundo orden, con un valor de 90.14 %.

Se chequean los supuestos de homocedasticidad, incorrelación y normalidad de los residuos del modelo. Mediante la aplicación de las pruebas específicas de cada supuesto, se concluye que los residuos del modelo son homoscedásticos, incorrelacionados y provienen de una distribución normal⁵.

El modelo estimado que cumple con los supuestos anteriormente mencionados es el siguiente:

$$\hat{\mu}_i = \hat{Z}_i = 0,0003Longitud_i + 0,0019Latitud_i$$

Por lo tanto, mediante la aplicación de la técnica anteriormente mencionada se concluye la existencia de tendencia espacial del promedio anual de granizadas tanto para la dirección sur-norte como para la dirección oeste-este, es decir, tanto en latitud como en longitud.

En el siguiente gráfico se observa la representación de la superficie de tendencia espacial cuando se ajusta la regresión lineal estimada a una grilla de valores. Dicha grilla abarca toda la superficie del territorio uruguayo y es la misma que se aplica para la interpolación por Kriging. El detalle de la construcción de la misma se expone en sección 8.2.4.

⁵Ver resultados en apéndice C.2

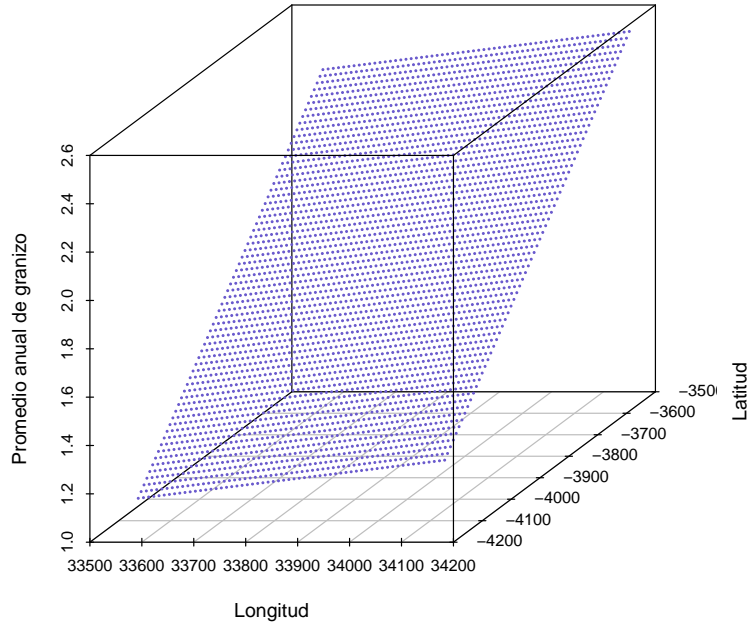


Figura 8.3: Representación de la superficie de tendencia

8.2.2. Kriging Residual

Bajo lo anteriormente estudiado, la presencia de tendencia espacial viola los supuestos de estacionariedad del proceso $Z(s)$. En tal sentido, Kriging Ordinario es implementado con el proceso estacionario $\hat{\epsilon}(s)$, es decir que utilizan los valores residuales del modelo planteado anteriormente y los residuos interpolados se añaden a la tendencia para calcular los valores estimados para $Z(s)$.

De otra manera, el proceso que se utiliza para realizar la interpolación por Kriging es el siguiente:

$$e(s_i) = \epsilon(\hat{s}_i) = Z(s_i) - \mu(\hat{s}_i) \quad \forall i = 1, \dots, 19$$

8.2.3. Estimación del Variograma

Se estima el variograma teórico mediante el estimador clásico. Se construye dicho variograma a partir de los residuos $\{e(s_1), \dots, e(s_{19})\}$, tal como se menciona en sección 6.8.4.

8.2. Interpolación por Kriging Ordinario

Las distancias máxima y mínima entre las estaciones meteorológicas es de 688 y 10.9 km respectivamente como se menciona en sección 8. Como primer medida se trabaja con un intervalo de distancia entre estaciones de 0 a 400 km, es decir que se acota el intervalo inicial mencionado anteriormente. Para determinar la distancia h , se divide el intervalo escogido en aproximadamente 13 segmentos iguales, por lo cual $h = 30$.

Para definir la tolerancia Δh se debe tener en cuenta que las estaciones se encuentran irregularmente espaciadas⁶, por lo tanto puede ocurrir que no existan valores del promedio anual de granizo a una distancia $h = 30$. Para solucionar este inconveniente se consideran las *clases de distancias*⁷. El valor de Δh se debe escoger de manera que no queden superpuestas las clases de distancia y de que no hayan valores de la variable que queden por fuera de las clases. La manera óptima es tomar a Δh es el 50% del valor de h .

Por otro lado, el variograma realizado es el omnidireccional⁸, pues en los residuos no se detecta anisotropía⁹.

A continuación se presenta el gráfico obtenido del variograma experimental de los residuos.

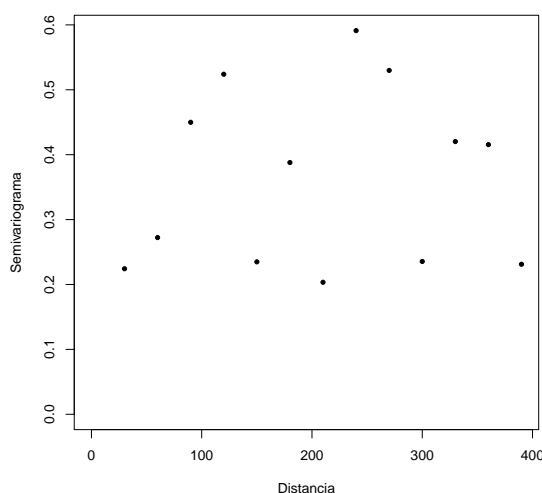


Figura 8.4: Variograma experimental

⁶Ver apéndice A.3.2

⁷Ver apéndice A.3.2

⁸Ver apéndice A.3.3

⁹Ver apéndice A.5

Una vez construido el variograma experimental se procede a escoger un modelo de variograma teórico óptimo que se ajuste al estimado. Para ello se calcula el error cuadrático medio cometido por el método de Kriging Ordinario según los distintos modelos de variogramas teóricos y se selecciona aquel modelo que minimice dicho error. Para calcular los errores cometidos se procede a aplicar la técnica conocida como *cross-validation*¹⁰.

El modelo de variograma que logra obtener el menor error de estimación de Kriging es el seno cardinal. El mismo posee un comportamiento cuadrático en el origen con un rango estimado¹¹ igual a 22.96 km y meseta estimada igual a 0.29, siendo su fórmula la siguiente:

$$\gamma(h) = 0,29 \left(1 - \frac{\text{sen}\left(\frac{|h|}{22,96}\right)}{\frac{|h|}{22,96}} \right)$$

En el siguiente gráfico se presenta el variograma experimental y el modelo de variograma ajustado:

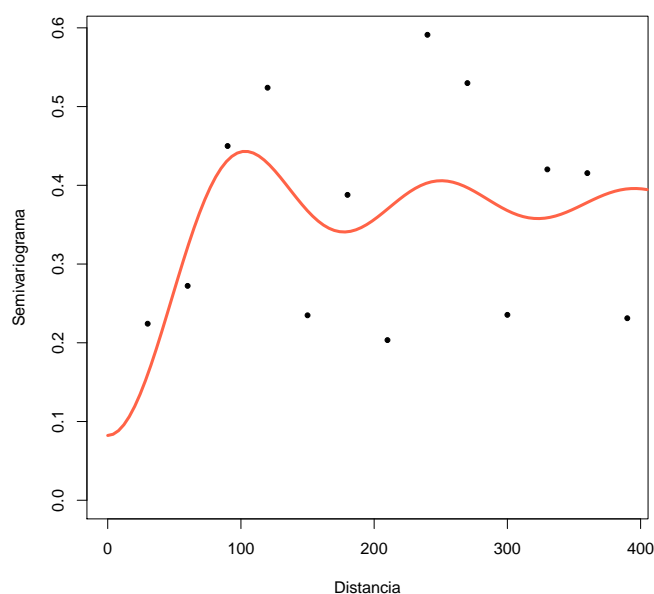


Figura 8.5: Variograma experimental ajustado con modelo seno cardinal

¹⁰Ver sección 8.2.5

¹¹Las estimaciones tanto del rango como de la meseta se realizan mediante mínimos cuadrados ordinarios

Se observa la presencia de un efecto pepita, lo que correspondería a posibles errores de medida en los datos¹².

Ver gráficos de los restantes modelos de variogramas ajustados al experimental en apéndice C.3.

8.2.4. Estimación por Kriging

Una vez escogido el modelo de variograma, se resuelven las ecuaciones de kriging y se obtienen los valores estimados de los residuos en las zonas donde se carece de información.

Luego, se estima el comportamiento del promedio anual de caída de granizo, sumándole a los residuos interpolados la tendencia estimada. De esta manera se obtiene un mapa zonificado del comportamiento del promedio de ocurrencia anual de granizo en Uruguay.

Para ello, previamente se debe de contruir una malla o retícula que abarque la superficie que se desea estimar. En tal sentido, se toma el rango de los vectores longitud y latitud y se dividen en p segmentos iguales.

Se construyen diferentes tamaños de grillas, que difieren en la la cantidad de particiones que se toman tanto para la longitud como para la latitud. Se realiza Kriging Ordinario con cada una de ellas, y se estudia el error de Kriging o varianza de Kriging. Finalmente, se decide trabajar con una grilla con $p = 10$, la cual obtiene la menor varianza de estimación en promedio de 0.3881.

¹²Ver apéndice A.3.6

8.2. Interpolación por Kriging Ordinario

La retícula contruida es la siguiente:

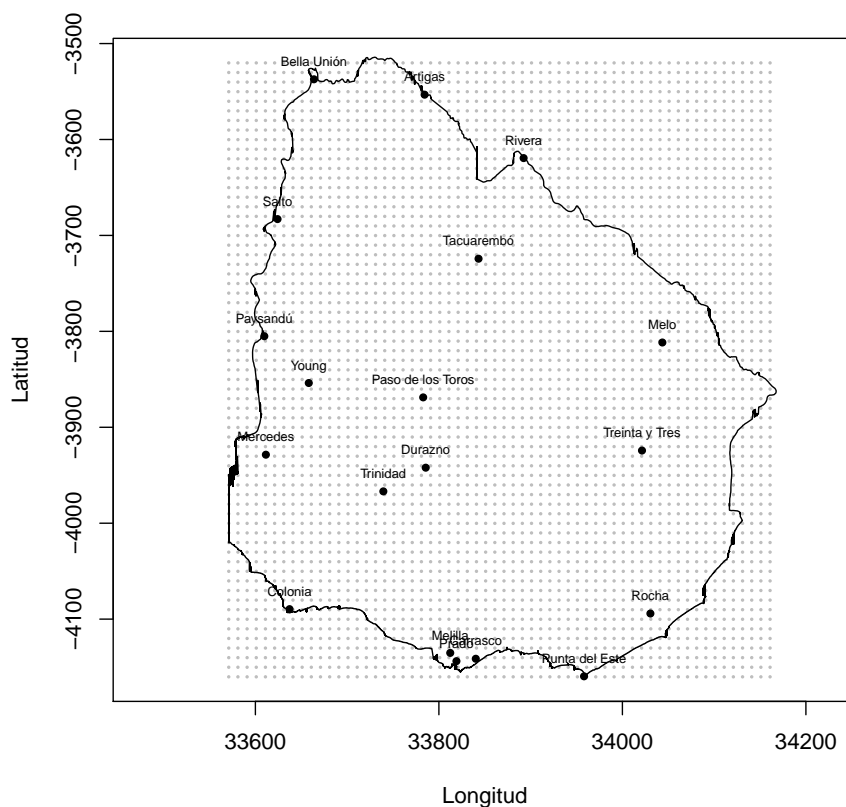


Figura 8.6: Grilla de interpolación

Una vez construida la grilla, y utilizando la información proveniente del modelo de variograma, la interpolación por Kriging se encarga de estimar cada punto de la misma, obteniendo como resultado un mapa estimado del promedio de caída de granizo anual en el territorio uruguayo.

8.2. Interpolación por Kriging Ordinario

A continuación se observan las curvas de nivel para los valores estimados según kriging.

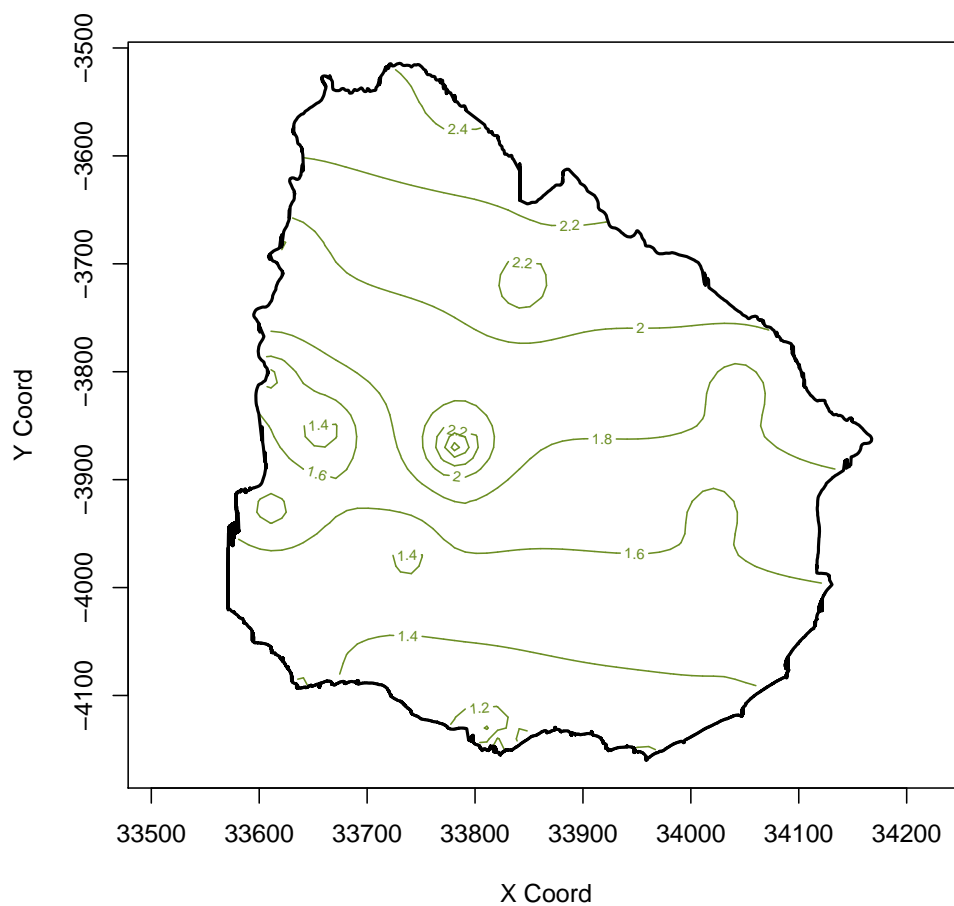


Figura 8.7: Curvas de nivel estimación por Kriging Ordinario

A continuación se presenta el gráfico con la estimación del comportamiento del promedio de granizadas anuales en Uruguay.

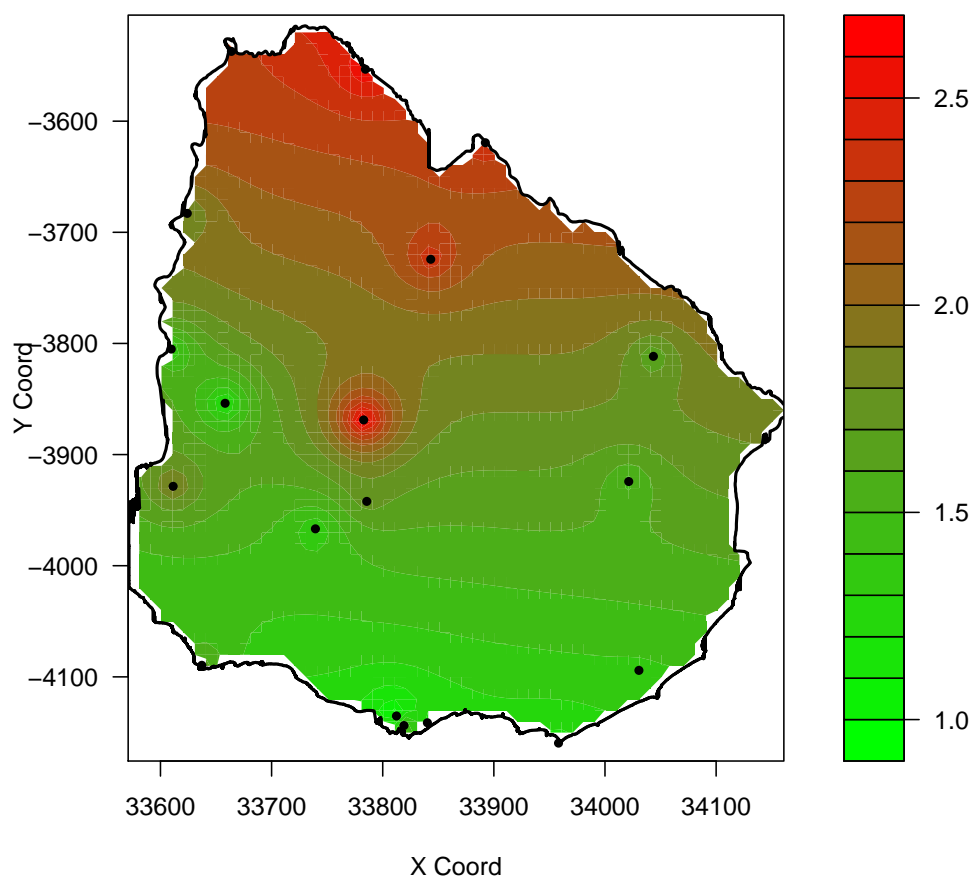


Figura 8.8: Estimación por Kriging Ordinario

Las estimaciones obtenidas reflejan que en la región centro-norte del país se presenta una mayor caída de granizo anual en promedio, oscilando sus valores entre 2 y 2.7. Los valores reales en dichas zonas superan los valores estimados, por lo que dicho método está sub-estimando el comportamiento de la caída de granizo.

En la zona sur y zonas aledañas a la costa oceánica del país se encuentran las estaciones meteorológicas con menor caída anual de granizo en promedio en términos reales. Kriging estima a estas zonas con bajos valores de caída granizo, aun así en estas zonas se ha sobre-estimado el comportamiento de dicha variable. En las zonas mencionadas los valores estimados oscilan entre 0.9 y 1.5 caídas anuales de granizo en promedio.

8.2. Interpolación por Kriging Ordinario

En el siguiente grafico se presentan los desvíos estándar de las estimaciones cometidos por el método.

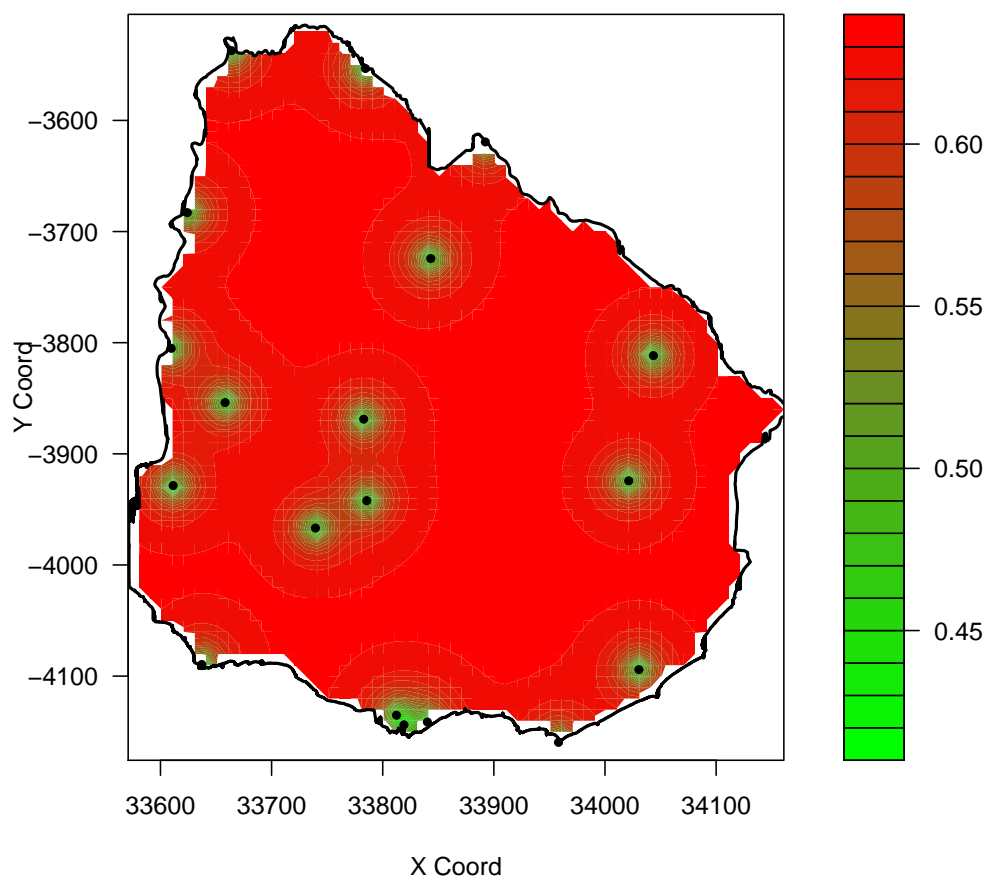


Figura 8.9: *Desviación estándar estimación por Kriging Ordinario*

Se observa que en las zonas próximas a las estaciones meteorológicas es donde se comete el menor error de estimación, aumentando de forma sustancial a medida que la distancia del punto a estimar con dichas estaciones aumenta.

En el apéndice C.4 se exhiben otros gráficos análogos a los vistos anteriormente.

8.2.5. Validación Kriging

La validación de Kriging consiste en:

- Validar el modelo de variograma
- Validar grilla de interpolación
- Analizar errores de estimación en toda la superficie de interpolación

En una primera instancia, para validar el modelo de variograma a utilizar, para cada uno de los modelos de variogramas se aplica la técnica de *cross-validation*. La misma se detalla a continuación:

Se consideran los residuos de cada una de las estaciones meteorológicas $e(s_1), e(s_2), \dots, e(s_{19})$ y en cada punto se procede de la siguiente manera:

1. Se excluye la estación s_i y se realiza la estimación por Kriging con el conjunto de las 18 restantes.
2. Se calcula el error de estimación cometido en cada punto:

$$r(i) = (e^*(s_i) - e(s_i))^2 \quad \forall i = 1, \dots, 19$$

Se estima así el error cuadrático medio cometido por el método de la siguiente manera:

$$\hat{E} = \sqrt{\frac{\sum_{i=1}^{19} r(i)}{19}}$$

Se calcula \hat{E} para cada modelo de variograma, buscando optimizar los resultados de las estimaciones.

El modelo de variograma que optimiza las estimaciones de Kriging es el seno cardinal, como fue detallado en sección 8.2.3.

Una vez ajustado el modelo de variograma, se busca optimizar los valores de interpolación mediante la construcción de la grilla. Para ello, se modifica el tamaño de las distintas retículas propuestas y se analizan las varianzas de kriging obtenidas para cada una ellas. Finalmente, la grilla que hace que las estimaciones de los residuos obtengan una menor varianza es la vista en sección 8.2.4.

En tercer lugar, se analizan los errores de estimación cometidos por la interpolación según ciertos criterios como son el análisis de la media y la nube de dispersión entre

los valores predichos y los valores reales.

A continuación se presentan gráficamente los residuos predichos por la interpolación versus los residuos observados.

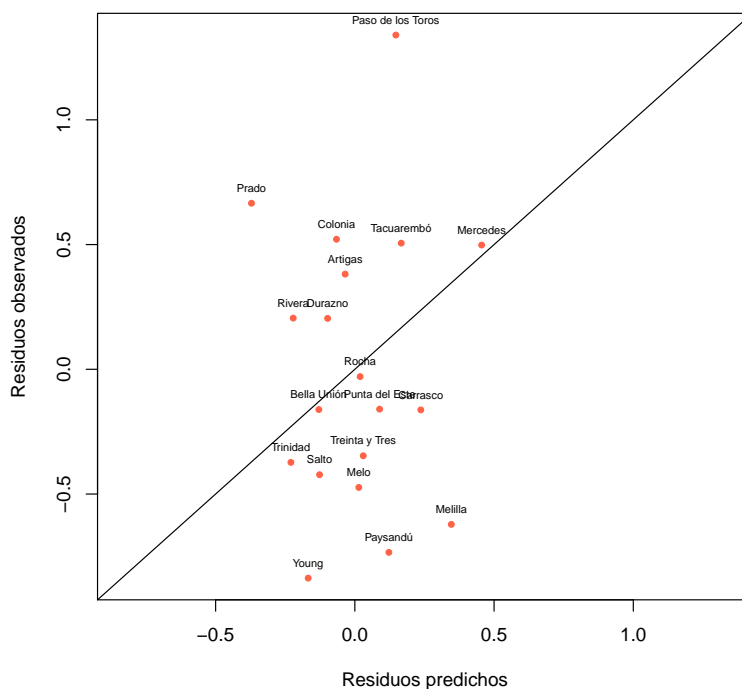


Figura 8.10: *Residuos predichos versus residuos observados*

En el gráfico anterior se puede observar que los valores que toman los residuos observados se encuentran en el rango de valores (-1,1) mientras que los valores de los residuos predichos se encuentran acotados en un entorno menor de (-0.5,0.5). Lo anterior da la pauta que dicho modelo está subestimando y sobre-estimando algunos datos.

Las estaciones de Bella Unión, Mercedes, Rocha y Trinidad son las que obtienen una mejor estimación por parte de Kriging respecto a sus valores reales. Mientras que el mayor error de estimación se comete para las estaciones de Paso de los Toros, Young, Paysandú, Melilla y Prado. En términos generales, aquellas observaciones que se encuentran por encima de la recta $x = y$ son sub-estimadas, mientras que las encuentran por debajo son sobre-estimadas.

8.2.6. Conclusiones

Las principales conclusiones surgidas de implementar la metodología de Interpolación por Kriging Ordinario son las siguientes:

1. Mediante el análisis de la posible existencia de anisotropía en el proceso en estudio, se encuentra que el comportamiento del promedio de granizo anual depende linealmente de la latitud y de la longitud. La existencia de tendencia en ambas direcciones hace que se deban interpolar los residuos y luego obtener los valores estimados de granizo para cada punto de la grilla sumándole a los mismos la tendencia estimada en cada localización.
2. Se estima que la mayor caída de granizo en promedio anual ocurre en la región centro-norte del país, aumentando su frecuencia a medida que se desplaza en dirección sur-norte. En las zonas cercanas a la costa oceánica se estima menor caída del fenómeno estudiado, resultado que coincide con la naturaleza del fenómeno estudiado.
3. Las estaciones de Paso de los Toros, Prado, Colonia, Tacuarembó, Rivera y Durazno son sub-estimadas por el método. Las estaciones mencionadas anteriormente son las que presentan los mayores valores de la variable en estudio, donde los mismos pertenecen al tercer y cuarto cuartil de la distribución.
4. Las estaciones meteorológicas de Young, Paysandú, Melilla, Carrasco, Melo, Treinta y Tres, Punta del Este y Salto son sobre-estimadas por el método de Kriging. Las mismas presentan valores pequeños de la variable de interés, ubicándose en el primer y segundo cuartil de la distribución.
5. Las estaciones de Mercedes, Rocha, Bella Unión y Trinidad son las que obtienen las mejores estimaciones de Kriging respecto a sus valores reales.

8.3. Análisis de Grupos Difusos

Con el objetivo de encontrar tipologías entre las estaciones meteorológicas en función de variables relacionadas a la caída de granizo en un período determinado, se implementa la técnica de *Análisis de Grupos Difusos*. La idea de incorporar la difusión entre los grupos conformados se debe a que, bajo el contexto del problema en estudio, es posible encontrar zonas 'grises', donde no esté clara la pertenencia de determinadas estaciones meteorológicas a un solo grupo.

De manera de incorporar información temporal a dicho análisis, se contruye una matriz de disimilaridad en base a las distancias entre las trayectorias temporales de las estaciones, utilizando las distancias propuestas por Carlier¹³ vistas en sección 7.3.2.

Se estudia la conformación de grupos bajo tres perspectivas, en función de la clase de distancia utilizada: distancia de Carlier Instantánea, distancia de Carlier Longitudinal y distancia de Carlier Mixta. Lo anterior se implementa con el objetivo de determinar si existen diferencias en cuanto a la estructura de grupos, según se enfatice al componente temporal de forma instantánea, evolutiva ó una combinación lineal de ambas.

En el presente trabajo se desarrolla con mayor precisión la metodología basada en la distancia de Carlier Mixta, ya que la misma toma en cuenta tanto el enfoque instantáneo así como el enfoque longitudinal en el tiempo. No obstante, los resultados obtenidos con los enfoques instantáneo y longitudinal se exponen en apéndice D.3.

Para realizar el presente análisis es necesario determinar el número de grupos a priori. Para ello, se analiza la estructura de grupos mediante un análisis de cluster jerárquico previo. En función de lo anterior, se decide trabajar para los tres enfoques con tres grupos¹⁴.

Asimismo, se predeterminaron los valores de ciertos parámetros para el funcionamiento del algoritmo. Es así como al parámetro que otorga el grado de superposición de los grupos (m) visto en sección 7.2 se le asigna un valor de 1.45 y la cantidad de iteraciones máximas que puede realizar el algoritmo son 1.000.

¹³Código en R para calcular distancias de Carlier fue otorgado por Andrés Castrillejo.

¹⁴Ver apéndice D.1

A continuación se presentan los resultados obtenidos mediante la utilización de la distancia de Carlier Mixta.

8.3.1. Características de los grupos

Para realizar la caracterización de los grupos se analizan las medidas de resumen de las variables caída de granizo y tamaño máximo de piedra, dentro de cada grupo conformado.

En el siguiente cuadro se observan las frecuencias de cada uno de los grupos:

Grupos	Frecuencia
1	7
2	7
3	5
Total	19

Cuadro 8.6: Frecuencias grupos

El siguiente cuadro ilustra el desempeño de cada grupo conformado en función de las variables de interés.

	Grupo 1		Grupo 2		Grupo 3	
	Mediana	Media	Mediana	Media	Mediana	Media
Caída de granizo	24	24	12	17	15	15
Tamaño máx. piedra	14.8	14.4	5.5	7.6	9.0	8.9

Cuadro 8.7: Caída de granizo y tamaño máximo de piedra según grupos

Como complemento del cuadro anterior se consideran los siguientes diagramas de cajas.

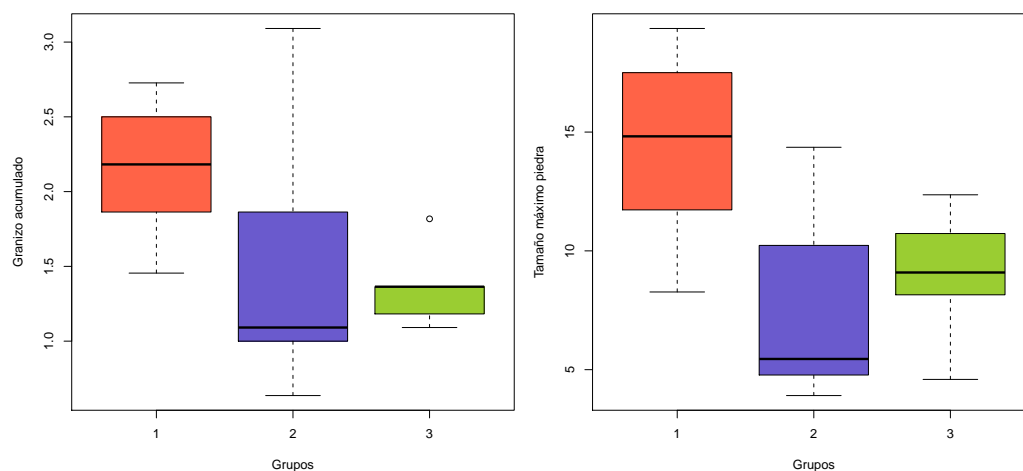


Figura 8.11: Diagramas de cajas grupos conformados

Los grupos conformados resultan caracterizados de la siguiente manera:

Grupo 1

Este grupo se caracteriza por poseer la mayor frecuencia de caída de granizo a lo largo de los once años, con un promedio de 24 granizadas. Asimismo, se caracteriza por poseer el mayor tamaño máximo de piedra, con un promedio de 14.4 mm de diámetro. Las estaciones que conforman este grupo son Artigas, Bella Unión, Colonia, Melo, Prado, Rivera y Tacuarembó.

Grupo 2

El grupo 2 es que el contiene mayor dispersión. El mismo está conformado por aquellas estaciones con frecuencia media de caída de granizo de 17. El tamaño máximo de piedra promedio de estas observaciones es de 7.6 mm de diámetro. Las estaciones que conforman este grupo son Carrasco, Melilla, Mercedes, Paso de los Toros, Punta del Este, Salto y Young.

Grupo 3

Este grupo se caracteriza por ser el más pequeño, con cinco observaciones, y el de menor variabilidad. Los valores de granizo oscilan en un entorno de 15 granizadas a lo largo de todo el período considerado, y el tamaño de piedra máximo promedio para el mismo es de 8.9 mm. Las estaciones contenidas en este grupo son Durazno, Paysandú, Rocha, Treinta y Tres y Trinidad.

En el siguiente gráfico se presenta la agrupación de las estaciones según los grupos conformados.

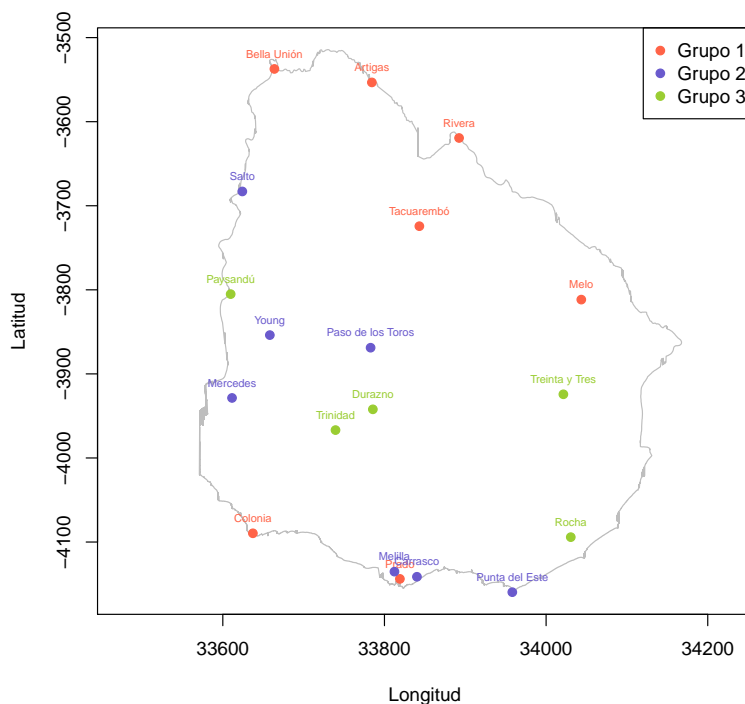


Figura 8.12: Conformación de grupos - Enfoque Mixto

Se observa que las estaciones localizadas en la región norte del territorio uruguayo son las que presentan mayor caída de granizo en los once años estudiados, así como el mayor tamaño de piedra promedio. Las estaciones meteorológicas ubicadas en el centro del país tienden a presentar bajos niveles de caída de granizo y tamaños de piedra promedio relativamente pequeños.

8.3.2. Análisis de coeficientes de pertenencia

Se analizan los coeficientes de pertenencia de las observaciones de los tres grupos conformados. De esta forma, se evalúan aquellas observaciones que tengan similares por lo menos dos de los grados de pertenencia más altos. A las estaciones que cumplan con lo mencionado anteriormente se las estudia con detalle y mediante un cierto criterio se decide si las mismas son clasificadas como estaciones difusas o no.

8.3. Análisis de Grupos Difusos

El siguiente cuadro presenta las medidas de resumen de esta información en cada grupo:

	Mínimo	1er cuartil	Mediana	Media	3er cuartil	Máximo
Grupo 1	41	44.5	48	47	48.5	54
Grupo 2	40	44.5	49	49.4	55	58
Grupo 3	41	52	52	56.2	66	70

Cuadro 8.8: *Coefficientes de pertenencia según grupos (en %)*

Se observa que las estaciones clasificadas al grupo 1, si bien fueron clasificadas al mismo, sus coeficientes de pertenencia a dicho grupo no superaron el 54 %. Mientras que las observaciones clasificadas en el grupo 3, contienen mayores grados de pertenencia a este grupo, con un máximo de 70 %.

Se escoge un criterio para clasificar a una observación como difusa. El mismo resuelve que aquella observación cuyo máximo grado de pertenencia sea menor a 50 % y cuya diferencia entre los grados de pertenencia restantes no supere 15 puntos porcentuales, sea considerada como observación difusa.

Las observaciones que cumplen con ambas condiciones mencionadas anteriormente son Artigas, Melilla, Melo, Paso de los Toros, Paysandú y Prado. Lo anterior significa que, estas estaciones meteorológicas, podrían pertenecer tanto al grupo que fueron clasificadas como a su grupo más cercano¹⁵.

¹⁵Ver cuadro en apéndice D.2.1

8.3. Análisis de Grupos Difusos

A continuación se presenta gráficamente el comportamiento de las observaciones consideradas como difusas respecto a los grupos.

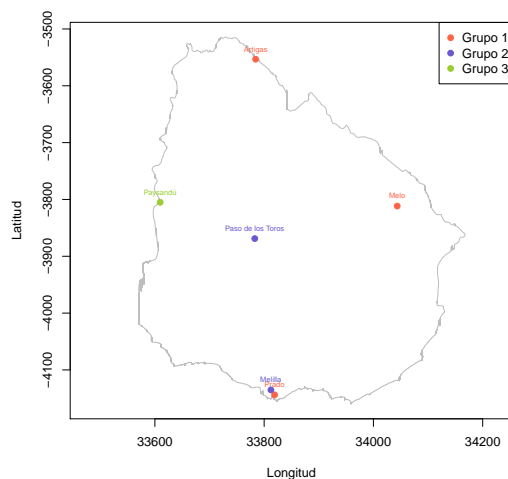


Figura 8.13: Grupos originales

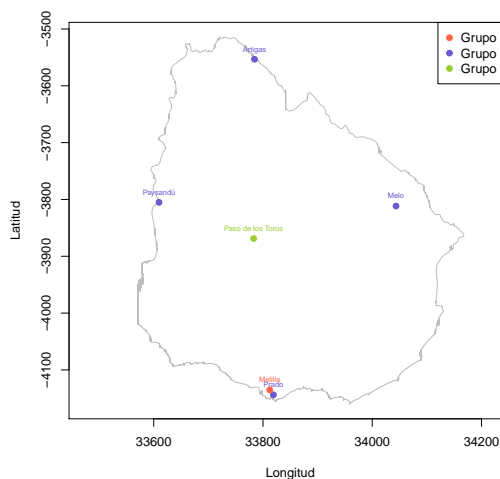


Figura 8.14: Grupos vecinos

En el gráfico de la izquierda se observa la pertenencia de las estaciones meteorológicas a los grupos, según la clasificación final del algoritmo. El gráfico de la derecha ilustra la agrupación de las estaciones difusas si fueran clasificadas al grupo vecino, es decir, el que comparte un grado de pertenencia similar al del grupo clasificado por el algoritmo.

En tal sentido, se aprecia que las estaciones que fueron clasificadas al grupo 1 podrían ser clasificadas al grupo 2, debido a la similitud entre los coeficientes de pertenencia. En cambio, las estaciones clasificadas al grupo 2, algunas podrían ser agrupadas en el grupo 1 y otras en el grupo 3. Tal es el caso de Paso de los Toros, perteneciente al grupo 2, la cual tiene como grupo vecino al grupo 3, mientras que Melilla podría ser clasificada al grupo 1. En cuanto a la estación ubicada en Paysandú, perteneciente al grupo 3, la misma puede ser agrupada en su grupo vecino, el grupo 2.

8.3.3. Conclusiones

Las principales conclusiones de aplicar la metodología de Análisis de Grupos Difusos son las siguientes:

1. La decisión de desarrollar con mayor precisión la metodología basada en la distancia de Carlier Mixta, radica en que la misma toma en cuenta tanto el enfoque instantáneo así como el enfoque longitudinal en el tiempo.
2. Las características de los tres grupos conformados son las siguientes:
 - **Grupo 1:** Este grupo se caracteriza por poseer la mayor frecuencia de caída de granizo a lo largo de los 11 años, con un promedio de 24 granizadas. Asimismo, se caracteriza por poseer el mayor tamaño máximo de piedra, con un promedio de 14.4 mm de diámetro. Las estaciones que conforman este grupo son Artigas, Bella Unión, Colonia, Melo, Prado, Rivera y Tacuarembó. Asimismo, las estaciones Artigas, Melo y Prado son consideradas como difusas, pues las mismas también poseen alto grado de pertenencia con el grupo 2.
 - **Grupo 2:** Este grupo es que el contiene mayor dispersión. El mismo está conformado por aquellas estaciones con frecuencia media de caída de granizo de 17. El tamaño máximo de piedra promedio de estas observaciones es de 7.6 mm de diámetro. Las estaciones que conforman este grupo son Carrasco, Melilla, Mercedes, Paso de los Toros, Punta del Este, Salto y Young. En este grupo se encuentran dos estaciones meteorológicas consideradas como difusas, Paso de los Toros y Melilla. La estación ubicada en Paso de los Toros podría ser clasificada al grupo 3, debido a la similitud de entre los grados de pertenencia y la estación localizada en Melilla podría ser agrupada en el grupo 1 por la misma razón.
 - **Grupo 3:** Este grupo se caracteriza por ser el más pequeño, con 5 observaciones, y el de menor variabilidad. Los valores de granizo oscilan en un entorno de 15 granizadas a lo largo de todo el período considerado, y el tamaño de piedra máximo promedio para el mismo es de 8.9 mm. Las estaciones contenidas en este grupo son Durazno, Paysandú, Rocha, Treinta y Tres y Trinidad. El grupo 3 contiene una observación difusa, Paysandú, la cual tiene como grupo vecino al grupo 2.
3. Se observa que las estaciones clasificadas al grupo 1, si bien fueron clasificadas al mismo, sus grados de pertenencia a dicho grupo no superaron el 54 %.

8.3. Análisis de Grupos Difusos

Mientras que las observaciones clasificadas en el grupo 3, contienen mayores probabilidades de pertenencia a este grupo, con un máximo de 70 %.

Conclusiones generales

9.1. Conclusiones

Luego de la investigación de metodologías para el análisis espacio-temporal del evento granizo en Uruguay, se encuentran dos técnicas que permiten complementariamente cumplir con el objetivo planteado.

La primer metodología que se analiza se basa en la interpolación espacial denominada *Interpolación por Kriging*. La segunda metodología que se presenta es una técnica multivariante denominada *Análisis de Grupos Difusos (Fuzzy Clustering Analysis)*.

La aplicación de ambas técnicas en forma complementaria resultan en la mejor aproximación al conocimiento del comportamiento espacio-temporal del evento climático granizo en Uruguay.

En la interpolación por Kriging se estima el valor del fenómeno meteorológico en estudio en un período dado, en zonas donde se carece de información, mientras que a través del análisis de grupos difusos se encuentran tipologías de estaciones en función de las variables relacionadas al fenómeno, conformando grupos homogéneos de estaciones meteorológicas. Se considera de gran interés el aporte que brindan los grados de pertenencia de las observaciones a los grupos, lo cual incorpora al análisis la superposición entre dichos grupos. A esta última técnica se le incorpora además un factor temporal, el cual considera la evolución en el tiempo de las observaciones en estudio. Es por ello que ambas metodologías son complementarias.

9.2. Limitaciones

Las limitaciones encontradas en este trabajo se basan en su mayoría en la información disponible. La información meteorológica en Uruguay contiene muchas limitantes a la hora de analizar cualquier fenómeno: existencia de datos faltantes durante

períodos extensos en un mismo punto de muestreo y grandes áreas del territorio sin información. Para una aplicación de la metodología que pueda ser de mayor utilidad a las distintas producciones sería necesario contar con una red de captura de datos más densa y geográficamente mejor distribuída.

En particular en Uruguay, no se cuenta con información sobre variables asociadas al fenómeno granizo, como ser actividad eléctrica, viento y altura de las nubes, entre otras.

9.3. Futuras investigaciones

Para futuras investigaciones, sería de gran interés incorporar un componente de penalización en la función objetivo del algoritmo de grupos difusos, de manera que se tome en cuenta el aspecto espacial (distancia geográfica entre las estaciones meteorológicas) para este tipo de datos utilizados.

Sería interesante analizar el comportamiento del fenómeno granizo en las diferentes estaciones del año, verano, otoño, invierno y primavera, mediante las técnicas estudiadas.

Por último, sería de gran interés implementar las metodologías vistas considerando diferentes períodos de tiempo, por ejemplo, tomar ventanas de tiempo cada 5 años.

Bibliografía

- [1] Alfaro, P. *Procedimiento Interpolación Espacial de Variables. Versión 1*. Dirección Nacional de Meteorología. Montevideo, 2010.
- [2] Blanco, J. *Introducción al análisis multivariado*. Instituto de Estadística. UdeLaR. Montevideo, 2006.
- [3] Coppi, R; D'Urso, P; Giordani, P. *A Fuzzy Clustering Model for Multivariate Spatial Time Series*. 2010.
- [4] Cressie, N. *Statistics for Spatial Data*. Ed. John Wiley & Sons, Inc. New York, 1993.
- [5] Cressie, N; Grondona, M.O. *A comparison of variogram estimation with covariogram estimation*. The art of statistical science. Ed. John Wiley & Sons, Inc. 1992.
- [6] Gattinoni, N; Mercuri, P; Mezher, R. *Distribución espacio-temporal del granizo en Argentina*. Argentina, 2008.
- [7] Gattinoni, N; Mercuri, P; Mezher, R. *Variabilidad estacional de la ocurrencia de granizo en el centro, este y noreste de Argentina*. Argentina, 2008.
- [8] Gómez Pérez, P. *Análisis espacial de datos para clasificar musáceas en zonas bananeras del Ecuador*. Ecuador, 2009.
- [9] Kitanidis, P. *Introduction to Geostatistics*. Ed. Cambridge University Press. 1999.
- [10] López Hernández, F; Palacios Sánchez, M. *Distintos modelos de dependencia espacial. Análisis de autocorrelación*. Universidad politécnica de Cartagena. Junio, 2000.
- [11] Lee, L. P. *The Transverse Mercator Projection of the spheroid*. Survey Review, Volumen 8, Número 58. Ed. Maney Publishing. Octubre 1945.

- [12] Nardecchia, G; Zabal, M. *Modelización de un proceso espacial para la producción de petróleo en una región de la cuenca neuquina*.
- [13] Reyes Cortés, M. *Estimación paramétrica y no paramétrica de la tendencia en datos con dependencia espacial. Un estudio de simulación*. Junio, 2010.
- [14] Torres Ruiz, E. *Agrometeorología*. 1995.

Otras Referencias

- [15] *Anuario Estadístico 2010*. Oficina de Estadísticas Agropecuarias, Ministerio de Ganadería, Agricultura y Pesca. Montevideo, 2010.
- [16] *Introducción a la Geoestadística. Teoría y Aplicación*. Departamento de Estadística. Universidad Nacional de Colombia.
- [17] *Mapas de Riesgo Agroclimáticos*. Agroseguro España y Ministerio de Ganadería, Agricultura y Pesca. Montevideo, 2008.
- [18] R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [19] *Golden Software Surfer 8*. 2007.

Referencias Paquetes estadísticos

- [20] Fortran code by H. Akima R port by Albrecht Gebhardt aspline function by Thomas Petzoldt. petzoldt@rcs.urz.tu-dresden.de. enhancements and corrections by Martin Maechler (2009). *akima: Interpolation of irregularly spaced data*. R package version 0.5-4. <http://CRAN.R-project.org/package=akima>.
- [21] Barry Rowlingson, Peter Diggle, adapted, packaged for R by Roger Bivand, *pcp* functions by Giovanni Petris and goodness of fit by Stephen Eglen (2009). *splanx: Spatial and Space-Time Point Pattern Analysis*. R package version 2.01-25. <http://CRAN.R-project.org/package=splanx>.
- [22] Jonathan M. Lees (2010). *GEOmap: Topographic and Geologic Mapping*. R package version 1.5-4. <http://CRAN.R-project.org/package=GEOmap>

BIBLIOGRAFÍA

- [23] Paulo J. Ribeiro Jr & Peter J. Diggle. *geoR: a package for geostatistical analysis*. R-NEWS, 1(2):15-18. June, 2001.
- [24] Adrian Trapletti and Kurt Hornik (2009). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-22.
- [25] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. (2005). *Cluster Analysis Basics and Extensions*; unpublished.
- [26] David B. Dahl with contributions from many others (2008). *xtable: Export tables to LaTeX or HTML*. R package version 1.5-4.
- [27] Achim Zeileis, Torsten Hothorn (2002). *Diagnostic Checking in Regression Relationships*. R News 2(3), 7-10. URL <http://CRAN.R-project.org/doc/Rnews/>.
- [28] John Fox. I am grateful to Douglas Bates, David Firth, Michael Friendly, Gregor Gorjanc, Spencer Graves, Richard Heiberger, Georges Monette, Henric Nilsson, Derek Ogle, Brian Ripley, Sanford Weisberg, and Achim Zeileis for various suggestions and contributions. (2009). *car: Companion to Applied Regression*. R package version 1.2-16. <http://CRAN.R-project.org/package=car>.
- [29] Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.
- [30] Ligges, U. and Mächler, M. (2003). *Scatterplot3d - an R Package for Visualizing Multivariate Data*. Journal of Statistical Software 8(11), 1-20.
- [31] Deepayan Sarkar (2009). *lattice: Lattice Graphics*. R package version 0.17-26. <http://CRAN.R-project.org/package=lattice>.

Parte V

Apéndice Metodológico

Apéndice A

Interpolación por Kriging Ordinario

A.1. Análisis de tendencia

Para analizar los modelos de regresión lineal planteados se realizan las siguientes pruebas de significación de parámetros y del modelo.

Prueba de significación del modelo

En la prueba de significación del modelo se testea el siguiente juego de hipótesis:

- $H_0) \beta_1 = \dots = \beta_k = 0$
- $H_1) \text{ Algún } \beta_k \neq 0, \text{ con } k=\text{número de parámetros a estimar}$

El estadístico utilizado en esta prueba es el siguiente: $F = \frac{SCExp/k}{SCErr/n-(k+1)} \sim F_{k,n-(k+1)}$

Prueba de significación de parámetros

En la prueba de significación de parámetros se testea el siguiente juego de hipótesis:

- $H_0) \beta_i = 0$
- $H_1) \beta_i \neq 0, \forall i = 1, \dots, k$

El estadístico de prueba utilizado es el siguiente: $t_i = \frac{\hat{\beta}_i}{\sqrt{Var(\hat{\beta}_i)}} \sim t_{n-(k+1)}$

A.2. Obtención de las ecuaciones de Kriging

El problema de optimización presentado en sección 6.7.1 es equivalente a resolver:

$$\begin{cases} \min \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(s - s_i) \\ \text{sujeto a } \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

El mismo se resuelve aplicando la técnica de Multiplicadores de Lagrange que se desarrolla a continuación.

Sea f una función de la siguiente forma:

$$f(\lambda_1, \dots, \lambda_n; \beta) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(s - s_i) - 2\beta \left(1 - \sum_{i=1}^n \lambda_i \right)$$

Dicha función se denomina función de Lagrange y β el multiplicador de Lagrange. Así minimizando f se obtienen los valores de λ_i que minimizan la varianza del error.

Luego de derivar f respecto a $\lambda_1, \dots, \lambda_n$ y respecto a β e igualando los resultados a cero, se obtienen las ecuaciones de Kriging.

- $\frac{\partial f}{\partial \lambda_i} = \sum_{j=1}^n \lambda_j C(s_i - s_j) - C(s - s_i) - \beta = 0$
- $\frac{\partial f}{\partial \beta} = \sum_{i=1}^n \lambda_i - 1 = 0$

Considerando la relación $\gamma(h) = C(0) - C(h)$, las ecuaciones de Kriging se escriben de la siguiente forma:

- $-\sum_{j=1}^n \lambda_j \gamma(s_i - s_j) + \gamma(s - s_i) - \beta = 0 \quad i = 1, \dots, n$
- $\sum_{i=1}^n \lambda_i = 1$

A.3. Variograma

A.3.1. Interpretación del variograma

El variograma puede presentar diferentes comportamientos según se le estudie a pequeñas distancias o a grandes distancias.

Comportamiento a pequeñas distancias

Permite estudiar cuán rápido puede variar la variable en estudio a pequeñas distancias. Básicamente existen cuatro formas que puede presentar el variograma a cortas distancias:

1. *Comportamiento Discontinuo o Efecto pepita*: Si $\gamma(h) \rightarrow c_0 \geq 0$ cuando $h \rightarrow 0$ entonces c_0 se denomina efecto *pepita* según Matheron. Esto demuestra que existe una discontinuidad en el origen. Esto no es posible para los procesos que pertenecen a C^2 (o sea si $Y(.) \in C^2$ entonces $E(Y(s+h) - Y(s))^2 \rightarrow 0$ si $\|h\| \rightarrow 0$). Por lo tanto, si a pequeña escala se presenta una discontinuidad, ésta solo es posible por un error de medida.
2. *Comportamiento Lineal*: Indica que para pequeñas distancias el variograma tiene un comportamiento lineal. Representa variables continuas pero no diferenciales. La variabilidad de la variable en estudio dependerá de la pendiente de la recta en el origen. A mayor pendiente, mayor variabilidad o viceversa.
3. *Comportamiento Cuadrático*: El comportamiento cuadrático a pequeñas distancias representa variables continuas e infinitamente diferenciables. Así la variable no varía tan rápidamente de un punto a otro.
4. *Comportamiento Híbrido*: Es cuando se verifican variaciones más "suaves" a cortas distancias y variaciones más "fuertes" a grandes distancias.

Comportamiento a grandes distancias

Se puede verificar que no todos los variogramas presentan un sill y rango finito. Esto se da por la presencia de una tendencia global (deriva o drifts), indicando que la variable en estudio no es estacionaria.

A.3.2. Estimación del variograma

Datos irregularmente espaciados

Los datos muestrales pueden estar distribuidos espacialmente de forma irregular. Por lo tanto, puede ocurrir que no existan valores de la variable en estudio a una distancia h o que no existan valores de la variable para una dirección θ .

Lo anterior hace que se consideren **clases de distancia** y **clases de dirección**.

Clases de distancias:

Para cada lag (h) se define una tolerancia Δh y se tienen en cuenta aquellos puntos que se encuentren entre $h - \Delta h$ y $h + \Delta h$.

¿Cómo se debe elegir Δh de forma apropiada?

El valor de Δh se debe escoger de manera que no queden superpuestas las clases de distancia y de que no hayan valores de la variable que queden por fuera de las clases. La manera óptima es tomar a Δh como el 50% del valor del lag h .

Clases de dirección:

Para cada dirección θ se define una tolerancia $\Delta\theta$ y se utilizan únicamente los puntos que se encuentran entre las direcciones $\theta - \Delta\theta$ y $\theta + \Delta\theta$.

¿Cómo determinar los valores de n , h , θ y $\Delta\theta$?

- **n**: Cuando se calcula un variograma sobre un dominio D , se escoge n de manera que:

$$nh = \frac{|D|}{2}$$

- **h**: Distancia promedio entre las estaciones meteorológicas. Esto se puede deducir a partir de dos variogramas: el *Variograma Cloud* y el *Variograma Omnidireccional*
- **θ** : Se escoge como la dirección de anisotropía de la variable. Se utiliza aquí el *Mapa de Variograma*

A.3.3. Variograma Omnidireccional

Es aquel que no depende de la dirección. Se obtiene al escoger la tolerancia angular $\Delta\theta$ de forma tal que las direcciones $\theta - \Delta\theta$ y $\theta + \Delta\theta$ sean opuestas y perpendiculares a la dirección θ .

Si por el contrario se realiza variogramas para ciertas direcciones, por ejemplo para detectar existencia de anisotropía, se realiza variogramas direccionales.

A.3.4. Variograma Cloud

Se obtiene graficando el valor de los pares versus la distancia. El variograma cloud permite:

- Detectar valores atípicos o cambios bruscos
- Escoger un valor inicial del lag
- Observar la dispersión alrededor del valor de γ

A.3.5. Mapa de Variograma

El mapa de variograma es una herramienta que permite determinar las direcciones de anisotropía de la variable en estudio.

A.3.6. Modelos de variograma

Los modelos de variograma isotrópicos más comunes son:

Modelo Efecto Pepita Puro

Este modelo representa a un fenómeno completamente aleatorio, en el cual no hay correlación espacial. No importa a la distancia que se encuentren los valores de la variable en estudio, siempre serán no correlacionados.

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| = 0 \\ s & \text{si } |h| \neq 0 \end{cases} \quad (\text{A.1})$$

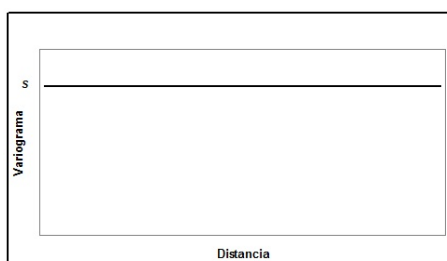


Figura A.1: *Modelo Efecto Pepita*

Modelo Esférico

Su comportamiento es casi lineal en el origen. La pendiente es igual a $\frac{1,5s}{a}$ siendo s el *sill* y a su *rango*. Representa fenómenos continuos pero no diferenciables. Es uno de los modelos más utilizados.

$$\gamma(h) = \begin{cases} s \left(\frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \frac{|h|^3}{a^3} \right) & \text{si } |h| \leq a \\ s & \text{si } |h| \geq a \end{cases} \quad (\text{A.2})$$



Figura A.2: Modelo esférico

Modelo Exponencial

Su comportamiento en el origen es lineal pero con pendiente $\frac{3s}{a}$, siendo s el *sill* que alcanza asintóticamente y su *rango* es igual a a . Representa fenómenos continuos pero no diferenciables.

$$\gamma(h) = s \left(1 - e^{-\frac{|h|}{a}} \right) \quad (\text{A.3})$$



Figura A.3: Modelo exponencial

Modelo Gaussiano

Posee un comportamiento cuadrático en el origen. Posee un *sill* s que alcanza asintóticamente y un *rango* igual a a . Representa fenómenos continuos e infinitamente diferenciables.

$$\gamma(h) = s \left(1 - e^{-\frac{|h|^2}{a^2}} \right) \quad (\text{A.4})$$

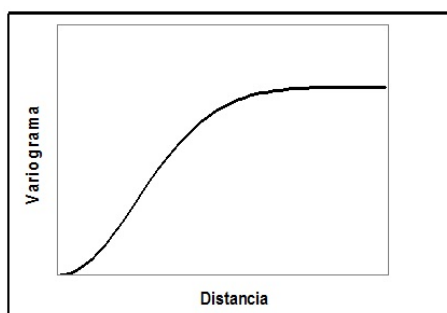


Figura A.4: *Modelo gaussiano*

Modelo Cúbico

Posee un comportamiento cuadrático en el origen. Posee un *rango* a y un *sill* s . Representa fenómenos continuos.

$$\gamma(h) = \begin{cases} s \left(7\frac{|h|^2}{a^2} - 8,75\frac{|h|^3}{a^3} - 0,75\frac{|h|^7}{a^7} \right) & \text{si } |h| \leq a \\ s & \text{si } |h| \geq a \end{cases} \quad (\text{A.5})$$

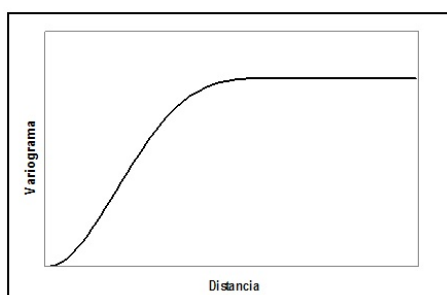


Figura A.5: *Modelo cúbico*

Modelo Seno Cardinal

Posee un comportamiento cuadrático en el origen con *rango* a y *sill* s que alcanza asintóticamente. Representa fenómenos continuos con periodicidades.

$$\gamma(h) = s \left(1 - \frac{\text{sen}\left(\frac{|h|}{a}\right)}{\frac{|h|}{a}} \right) \quad (\text{A.6})$$

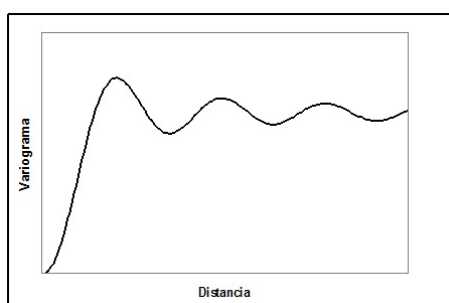


Figura A.6: Modelo seno cardinal

Modelo Potencia

El comportamiento en el origen depende del valor de p ($0 \leq p \leq 2$). Posee otro parámetro llamado *factor de escala* s . Representa fenómenos no estacionarios.

$$\gamma(h) = s |h|^p \quad (\text{A.7})$$

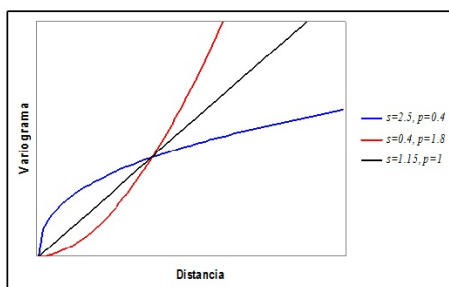


Figura A.7: Modelo potencia

Modelo Potencia-Exponencial

Su fórmula es la siguiente:

$$\gamma(h) = e^{-\frac{h}{a} \text{kappa}} \quad (\text{A.8})$$

Siendo a el rango, y $kappa$ un parámetro adicional de suavidad que oscila entre 0 y 2.

A.4. Covariograma

Otra herramienta para estudiar la dependencia espacial está dada por la función covarianza o covariograma.

Se denomina covariograma a:

$$Cov(Z(s), Z(s+h)) = C(h) \quad \forall s \in \mathfrak{R}^d$$

Las características de dicha función son las siguientes:

- Es una función acotada superiormente por la varianza, es decir:

$$|C(h)| \leq C(0)$$

- Es una función par:

$$C(-h) = C(h)$$

- $C(0) = Var(Z(s))$

Cuando un proceso es intrínsecamente estacionario, la relación existente entre variograma y covariograma es la siguiente:

$$\gamma(h) = \frac{1}{2}E[(Z(s+h) - Z(s))^2] = C(0) - C(h)$$

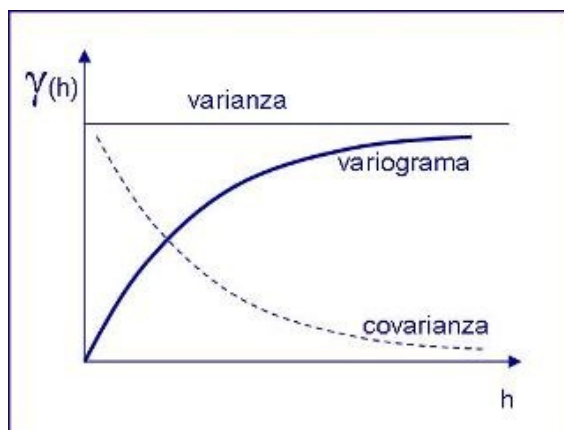


Figura A.8: Relación variograma-covariograma

A.5. Anisotropía

La anisotropía es una característica en la cual ciertas propiedades físicas varían según la dirección en que son examinadas. Algo anisótropo podrá presentar diferentes características según la dirección en que se lo estudie¹.

Para el caso particular del presente trabajo se dice que existe anisotropía si el variograma experimental al ser calculado para distintas direcciones presenta distintos comportamientos.

Existen diferentes tipos de anisotropía:

- Geométrica
- Zonal
- Híbrida

Anisotropía Geométrica

Es aquella en la que el variograma en distintas direcciones presenta el mismo sill (meseta) pero rangos distintos. Esto significa que existe mayor continuidad espacial en la dirección de mayor rango y que existe menor continuidad espacial en la dirección de menor rango.

Anisotropía Zonal

La anisotropía zonal ocurre cuando los variogramas en distintas direcciones presentan distintas mesetas (sill) para los mismos rangos.

Anisotropía Híbrida

Es aquella en la que el variograma en distintas direcciones presenta rangos diferentes y distintos *sill*.

¹La Isotropía en física es la característica de los cuerpos cuyas propiedades físicas no dependen de la dirección.

Análisis de Grupos Difusos

B.1. Distancia

Se denomina distancia a todo índice de disimilaridad que satisfaga las siguientes propiedades:

- $d_{ij} = 0 \Leftrightarrow i = j$ para todo $(i, j) \in E$
- $d_{ij} = d_{ji} \Leftrightarrow i = j$ para todo $(i, j) \in E$ (simetría)
- $d_{ij} \leq d_{ik} + d_{kj}$ para todo $(i, j, k) \in E$ (desigualdad triangular)

Un conjunto E en el cual se define una distancia determina un Espacio Métrico.

Por tanto valores elevados de este índice indica una menor similitud.

B.2. Algoritmos de clasificación

Se define el algoritmo de clasificación, para medir la distancia entre observaciones y grupos, y grupos entre sí, de modo de poder determinar cuáles observaciones pertenecen a un mismo grupo.

Existen distintos métodos jerárquicos de clasificación, entre ellos se encuentran: *Método del vecino más cercano*, *Método del vecino más lejano*, *Método del enlace medio*, *Método del centroide* y *Método de Ward (De la mínima varianza)*.

El algoritmo de unión utilizado en el presente trabajo es el *Método de Ward*. El mismo es un método jerárquico agregativo que busca optimizar, en cada etapa, la dispersión de las clases de la partición obtenida por agregación de dos objetos.

En este algoritmo la unión entre dos grupos se realiza de modo tal que se minimice la variación dentro de los grupos en la nueva partición. Si bien la variación en los

grupos de la clase que resulta de la unión de $\{KUL\}$ es más elevada que la suma de las variaciones de los grupos K y L, este criterio de agregación reúne los grupos que minimicen el incremento de la variación intra-grupos provocado por la unión de elementos más heterogéneos.

B.3. Reglas de detención

Se debe buscar algún mecanismo que permita decidir cuántos grupos considerar. Para ello se debe disponer de instrumentos que ayuden al investigador a tomar una decisión eligiendo las diversas particiones que se presenta de acuerdo a cada método.

Como método gráfico se puede destacar el *Dendrograma*, con él se puede tener una idea de cómo se van uniendo los grupos y las distancias a la que esto ocurre. Para interpretar el dendrograma debe trazarse una línea perpendicular a su desarrollo y contar los grupos que se han formado con esa partición.

Asimismo, se cuenta con ciertos índices creados para esta operación:

- R^2 : establece la relación entre la variación explicada y la variación total, donde la variación explicada la representa la estructura de grupos hallada en cada nivel.

$$R^2 = 1 - \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^J (x_{ij(k)} - \bar{x}_{kj})^2}{\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_j)^2}$$

Donde I es el número de individuos, J es el número de variables y K es el número de grupos.

En cada etapa de particiones encajadas se observa el valor del indicador y el incremento que se produce en el mismo al pasar de k grupos a $k + 1$ grupos. Si el aporte deja de ser sustancial entonces se seleccionan k grupos.

- **Pseudo F**: Se define el indicador de la siguiente manera:

$$F_{J(K-1), J(I-K)} = \frac{\frac{trB}{K-1}}{\frac{trW}{I-K}} = \frac{R^2/(K-1)}{(1-R^2)/(I-K)}$$

Donde:

trB es la suma de las variaciones entre los grupos (variación explicada)

trW es la suma de las variaciones dentro de los grupos (variación residual)

I es el número de observaciones

J es el número de variables

K es el número de grupos

Empíricamente se han determinado algunas reglas que contribuyen a su utilización:

- Si el indicador crece monótonamente al crecer el número de grupos k , entonces no se puede determinar una estructura clara.
 - Si el indicador disminuye monótonamente al crecer el número de grupos k , entonces no se puede determinar una estructura clara, pero sí se puede decir que existe una estructura jerárquica.
 - Si el indicador crece, llega a un máximo y luego decrece, entonces los datos presentan un número definido de grupos en ese máximo.
- **Pseudo t^2** : compara las trazas de las matrices de variaciones intragrupos G y L con la traza de la matriz de variaciones que surge de unir los grupos G y L .

$$pseudot^2 = \frac{trW_{GL} - (trW_G + trW_L)}{(trW_G + trW_L)/(n_G + n_L - 2)}$$

Este indicador analiza la significación de fusionar dos grupos. En cada paso considera los candidatos a unirse. Trata de determinar en cada paso si la disminución en la suma de cuadrados residuales como resultado de pasar de k a $k + 1$

Parte VI

Apéndice Resultados

Interpolación por Kriging Ordinario

C.1. Análisis de cuatro escenarios

Se estudian cuatro escenarios de interpolación por Kriging, los cuales difieren en la combinación de dos aspectos:

- El modelo utilizado para modelar la tendencia
- El variograma que mejor ajusta

Se analizan cada uno de estos escenarios que se detallan a continuación:

1. Se utiliza un modelo de tendencia con polinomio de primer orden y variograma Seno Cardinal
2. Se utiliza un modelo de tendencia con polinomio de primer orden y variograma Potencia-Exponencial
3. Se utiliza un modelo de tendencia con polinomio de segundo orden reducido y variograma Seno Cardinal
4. Se utiliza un modelo de tendencia con polinomio de segundo orden reducido y variograma Potencia-Exponencial

Los polinomios de primer y segundo orden utilizados son los siguientes:

- $Z_i = \beta_1 Longitud_i + \beta_2 Latitud_i + \epsilon_i \quad \forall i = 1, \dots, 19$
- $Z_i = \beta_1 (Longitud * Latitud)_i + \beta_2 Longitud_i^2 + \epsilon_i \quad \forall i = 1, \dots, 19$

C.1. Análisis de cuatro escenarios

Donde Z_i representa el promedio anual de granizo para la estación i .

Los resultados se exponen a continuación:

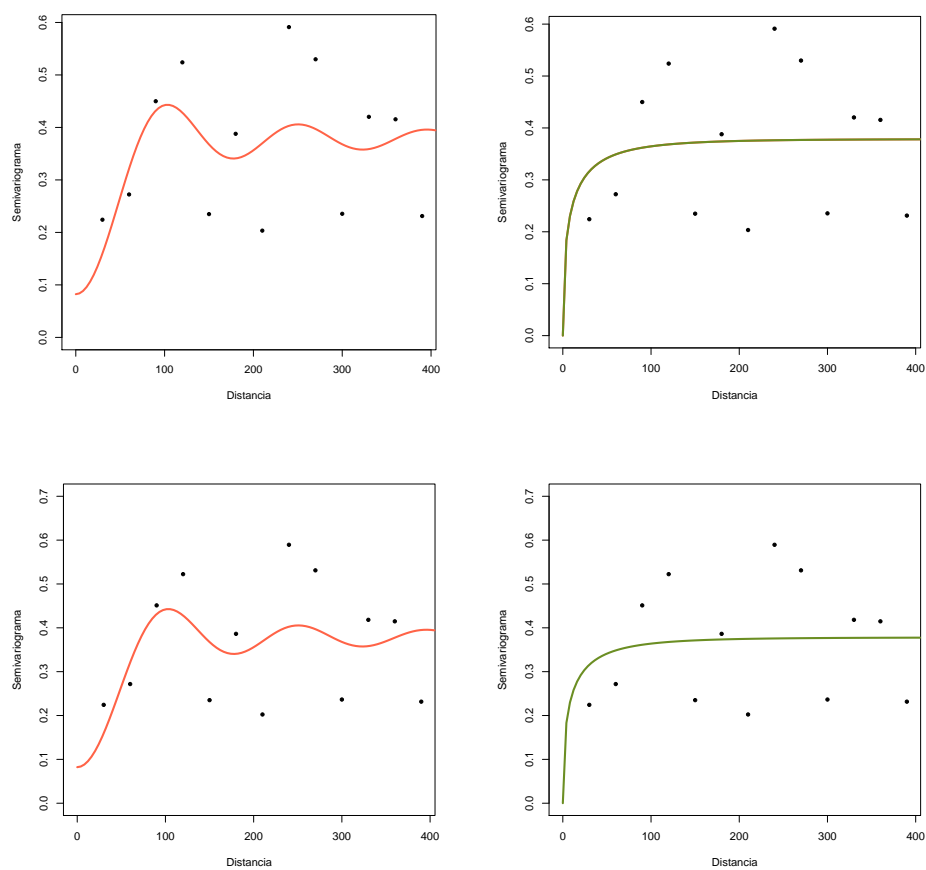


Figura C.1: Ajuste variogramas experimentales. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro

C.1. Análisis de cuatro escenarios

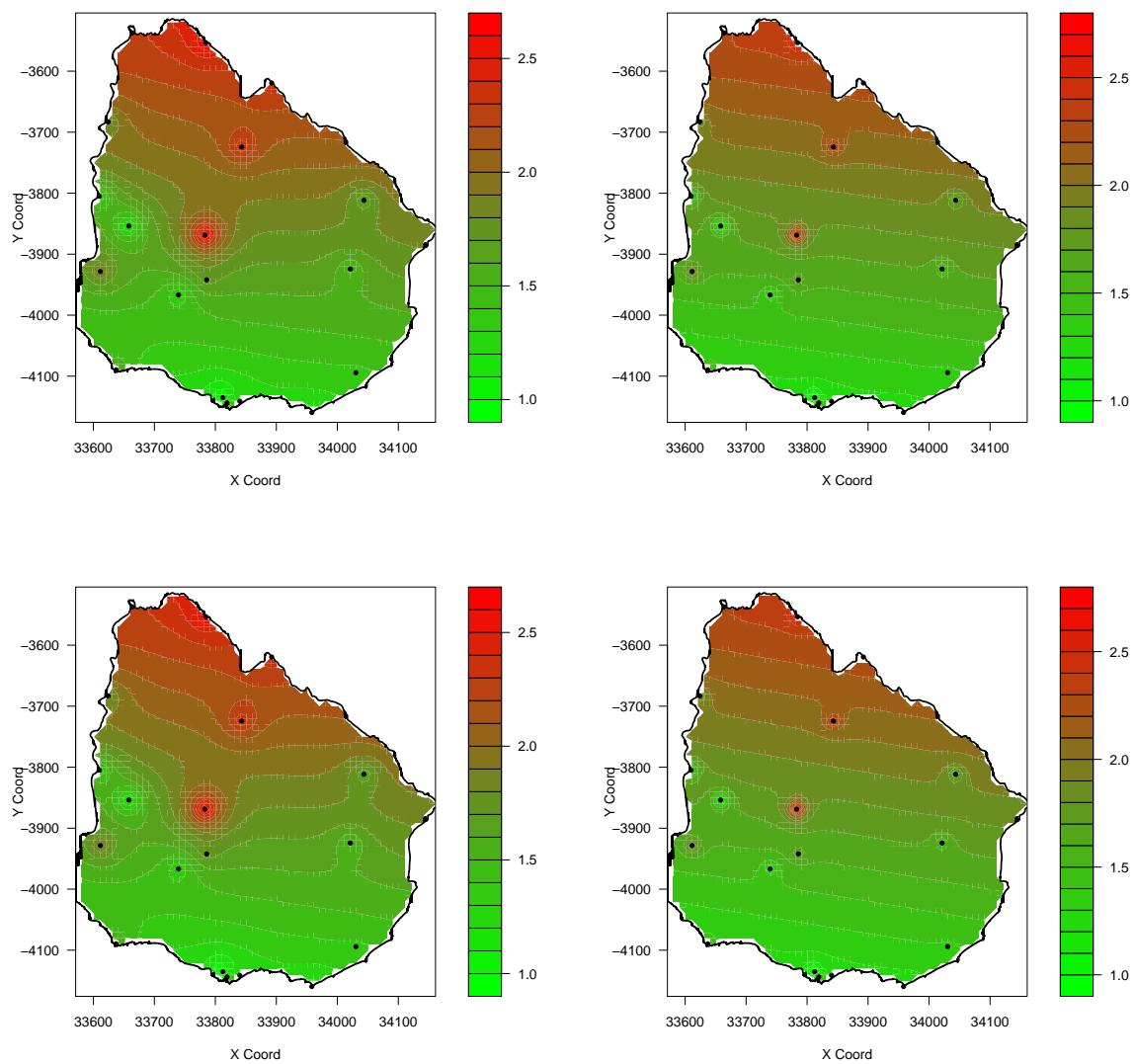


Figura C.2: Estimaciones Kriging. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro

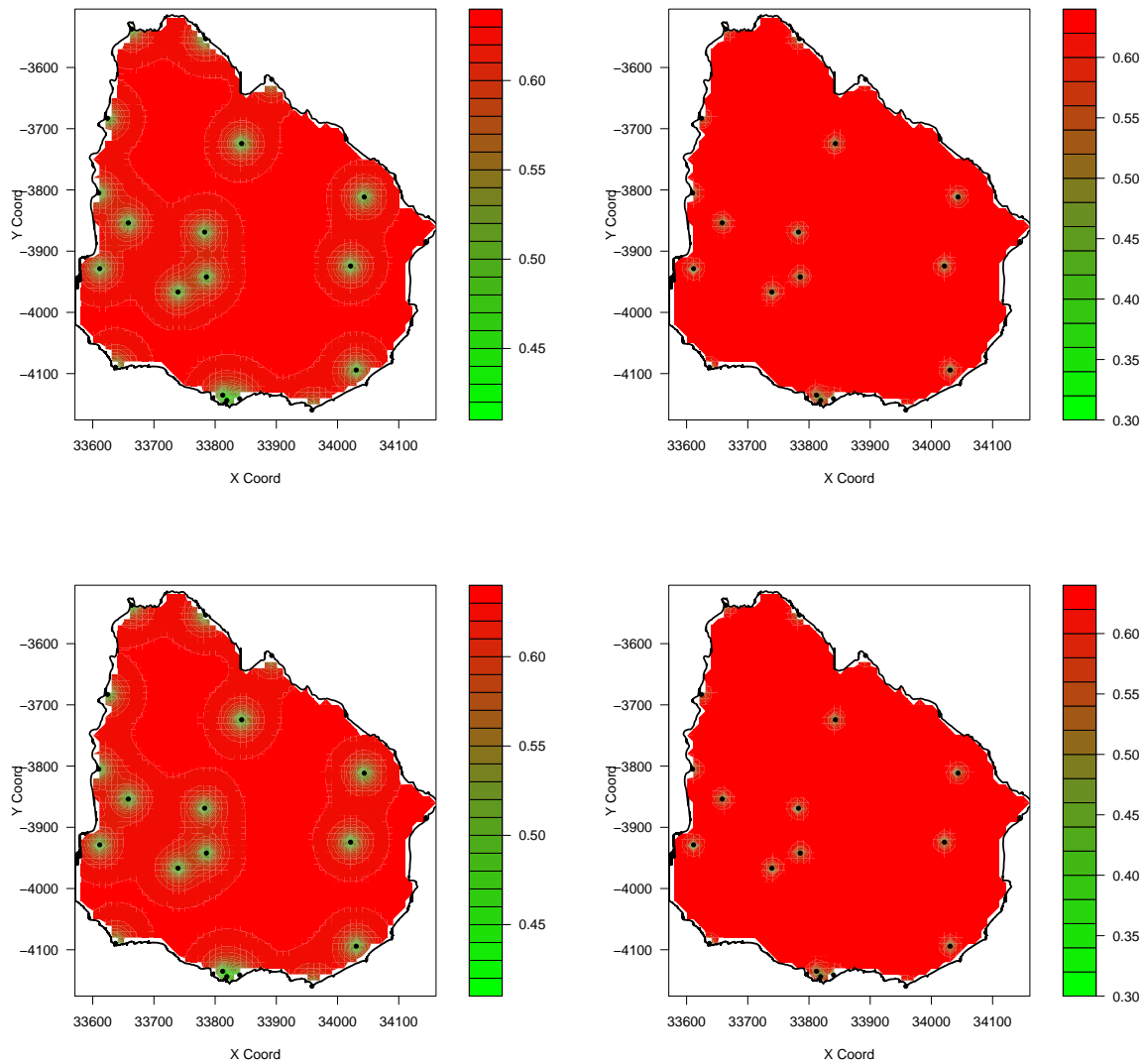


Figura C.3: Error estimación Kriging. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro

Si bien los resultados de estimación por Kriging son similares entre sí, en los gráficos de error de estimación se observa que con el modelo de tendencia utilizando un polinomio de primer orden se minimiza dicho error.

Una vez analizados con detalle, se decide trabajar con el escenario número uno, ya que por una parte el modelo utilizado para modelar la media del proceso es significativo al 5 % como así también sus parámetros. También posee el mayor R^2 ajustado.

Desde el punto de vista del variograma a utilizar, el *Seno Cardinal* es el que minimiza el error cuadrático medio, minimizando así también los errores de estimación cometidos por la metodología de Kriging Ordinario.

Se desarrolla con más profundidad el escenario elegido.

C.2. Análisis de la tendencia

Se testea si los residuos son homoscedásticos mediante los tests de Varianza no constante (Non-constant Variance), Breusch-Pagan y Goldfeld-Quandt. Los resultados obtenidos se detallan a continuación:

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 0.01237076 Df = 1 p = 0.9114388

studentized Breusch-Pagan test

data: modelo2

BP = 0.0094, df = 1, p-value = 0.923

Goldfeld-Quandt test

data: modelo2

GQ = 0.6619, df1 = 8, df2 = 7, p-value = 0.7131

Los p-valores asociados a las distintas pruebas de homoscedasticidad realizadas son mayores que el nivel de significación predeterminado 5 %, por lo tanto se acepta la hipótesis nula de que los residuos son homoscedásticos.

Para testear si los residuos del modelo son incorrelacionados se utiliza el test propuesto por Durbin-Watson. Los resultados son los siguientes:

Durbin-Watson test

data: modelo2

DW = 2.1937, p-value = 0.6612

alternative hypothesis: true autocorrelation is greater than 0

El p-valor de la prueba anterior supera el nivel de significación del 5% por lo tanto se concluye estadísticamente que los residuos son incorrelacionados.

Por último, se estudia si los residuos provienen de una distribución Normal a través de pruebas de bondad de ajuste. Los test aplicados son los de Kolmogorov-Smirnov, Shapiro-Wilk y Jarque-Bera. Los resultados de los mismos se detallan a continuación:

One-sample Kolmogorov-Smirnov test

```
data: modelo2$residuals
D = 0.2013, p-value = 0.3741
alternative hypothesis: two-sided
```

Shapiro-Wilk normality test

```
data: modelo2$residuals
W = 0.9579, p-value = 0.5316
```

Jarque Bera Test

```
data: modelo2$residuals
X-squared = 0.9587, df = 2, p-value = 0.6192
```

Luego de la aplicación de las pruebas de normalidad, según sus p-valores se puede concluir que los residuos provienen de una distribución normal.

C.3. Estimación del Variograma

A continuación se presenta el variograma experimental y los distintos modelos de variogramas que ajustan al mismo.

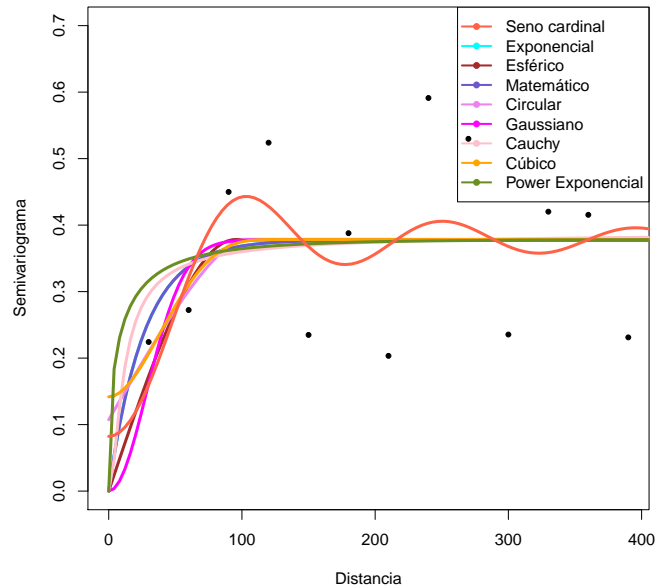


Figura C.4: Variograma experimental ajustado con distintos modelos

C.4. Estimación por Kriging

Como complemento de los gráficos exhibidos en la sección 8.2, se presentan gráficos alternativos, los cuales grafican la estimación de Kriging Ordinario y los errores de estimación. Asimismo, se ilustran también los mismos desde la perspectiva tri-dimensional.

C.4. Estimación por Kriging

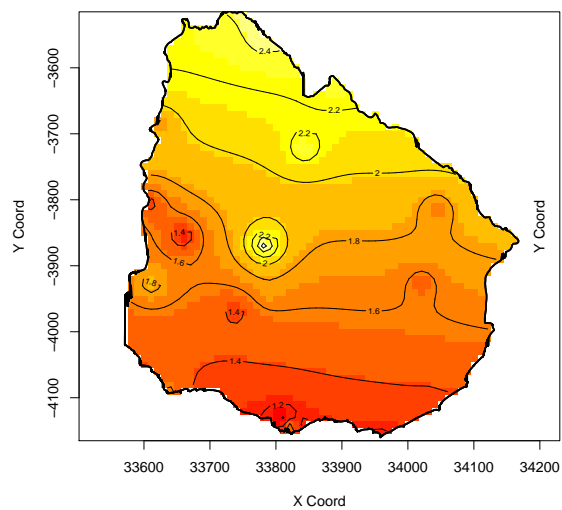


Figura C.5: Estimación Kriging

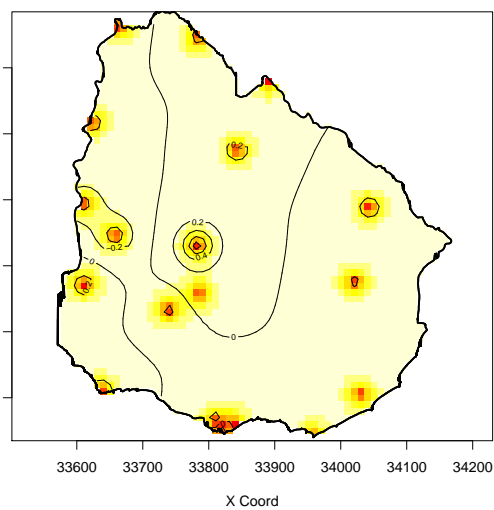


Figura C.6: Error estándar Kriging

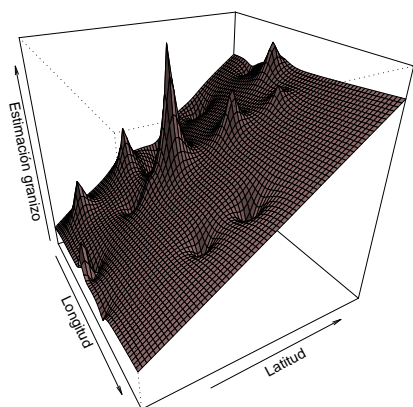


Figura C.7: Estimación Kriging

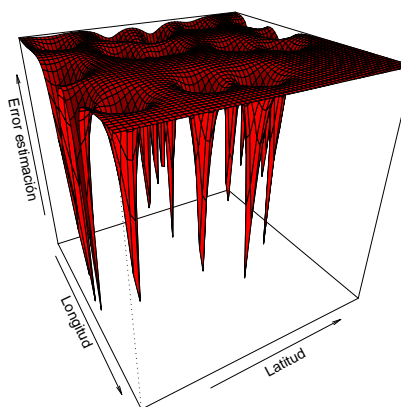


Figura C.8: Error estándar Kriging

Análisis de Grupos

D.1. Análisis de Grupos Jerárquico

El algoritmo de unión utilizado para la conformación de los grupos es el algoritmo de *Ward*¹, dado que el mismo presenta una clara estructura de grupos respecto al resto de los algoritmos.

En base a los indicadores R^2 , *pseudo F* y *pseudo t*² y al dendrograma, se decide retener tres grupos de observaciones.

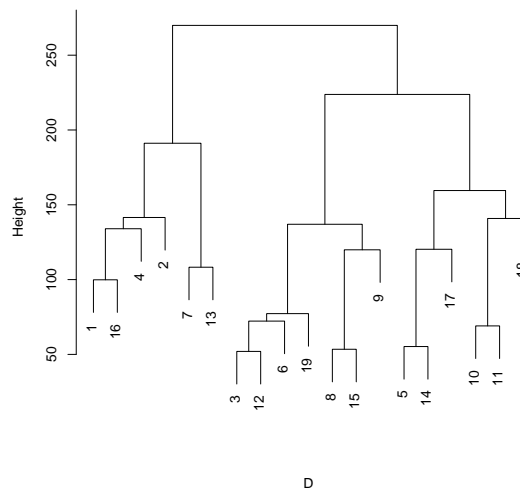


Figura D.1: Dendrograma Método Ward

El valor del R^2 para esta partición es de 40 %.

¹Ver apéndice B.2

D.2. Análisis de Grupos Difusos

A continuación se presentan los gráficos resultantes de aplicar la técnica de grupos difusos con la distancia de Carlier Mixta. En el primer gráfico se presentan los grupos junto a sus centroides y en el siguiente gráfico se presenta la información denominada *Silhouette*, la cual proporciona los valores de los coeficientes de pertenencia por grupo y brinda una idea del grado de superposición de los mismos.

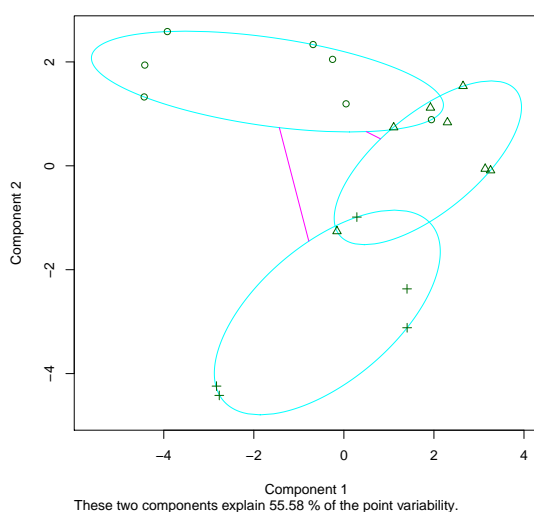


Figura D.2: Grupos y centroides - Análisis Mixto

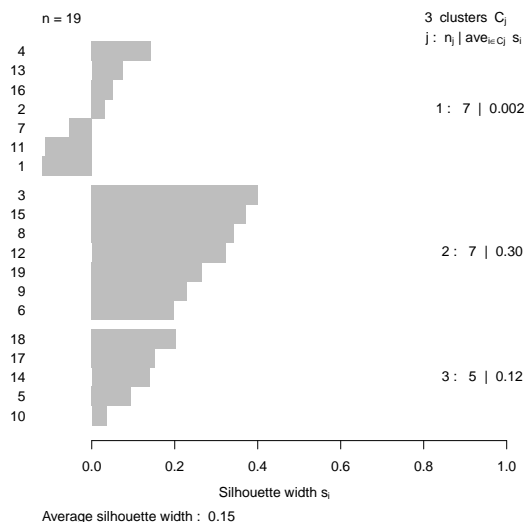


Figura D.3: Información Silhouette - Análisis Mixto

D.2.1. Coeficientes de pertenencia

En el siguiente cuadro se observan los coeficientes de pertenencia de las estaciones a cada uno de los grupos (u_1 , u_2 y u_3), así como los grupos más cercanos (grupos vecinos)².

Estaciones	u1	u2	u3	Grupo	Grupo vecino
Artigas	45.4	33.9	20.6	1	2
Bella Unión	48.3	32.7	19.0	1	2
Carrasco	18.1	57.4	24.5	2	3
Colonia	48.7	26.2	25.1	1	2
Durazno	15.3	18.5	66.2	3	2
Melilla	33.6	40.7	25.6	2	1
Melo	44.3	29.9	25.8	1	2
Mercedes	26.7	52.9	20.4	2	1
Paso de los Toros	26.5	40.3	33.2	2	3
Paysandú	30.8	28.6	40.5	3	2
Prado	40.7	29.6	29.7	1	2
Punta del Este	21.6	49.1	29.4	2	3
Rivera	47.9	30.6	21.5	1	2
Rocha	12.7	17.4	69.9	3	2
Salto	23.9	57.8	18.3	2	3
Tacuarembó	54.5	26.7	18.9	1	2
Treinta y Tres	22.6	25.1	52.3	3	2
Trinidad	22.5	25.2	52.3	3	2
Young	32.0	48.0	20.0	2	1

Cuadro D.1: Estaciones meteorológicas y sus coeficientes de pertenencia

D.3. Análisis Instantáneo y Longitudinal

Se realiza la metodología de agrupación de estaciones, esta vez utilizando las distancias de Carlier Instantánea y Longitudinal con el objetivo de determinar si existe movilidad de las estaciones según el tipo de distancia que se utilice.

El mecanismo implementado para realizar los análisis es el mismo que el implementado para el enfoque mixto. Asimismo, la caracterización de los tres grupos en cada uno de los dos enfoques resulta ser la misma que la del enfoque mixto. Por lo tanto, el objetivo en este apartado consiste en estudiar particularmente la movilidad de ciertas observaciones respecto a las distintas distancias aplicadas.

²Se considera como grupo vecino a aquel que presente el segundo grado de pertenencia más alto luego del máximo

D.3. Análisis Instantáneo y Longitudinal

Se presentan a continuación los gráficos correspondientes a la conformación de los grupos en cada uno de los enfoques.

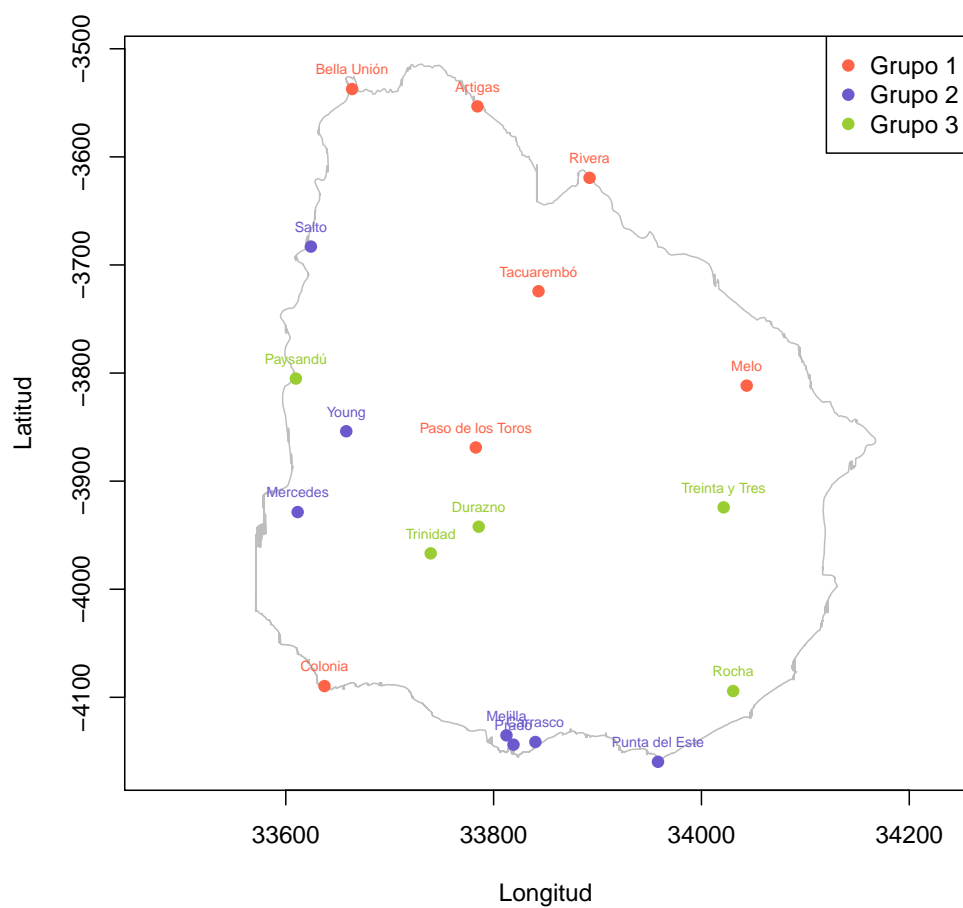


Figura D.4: Conformación de grupos - Análisis Instantáneo

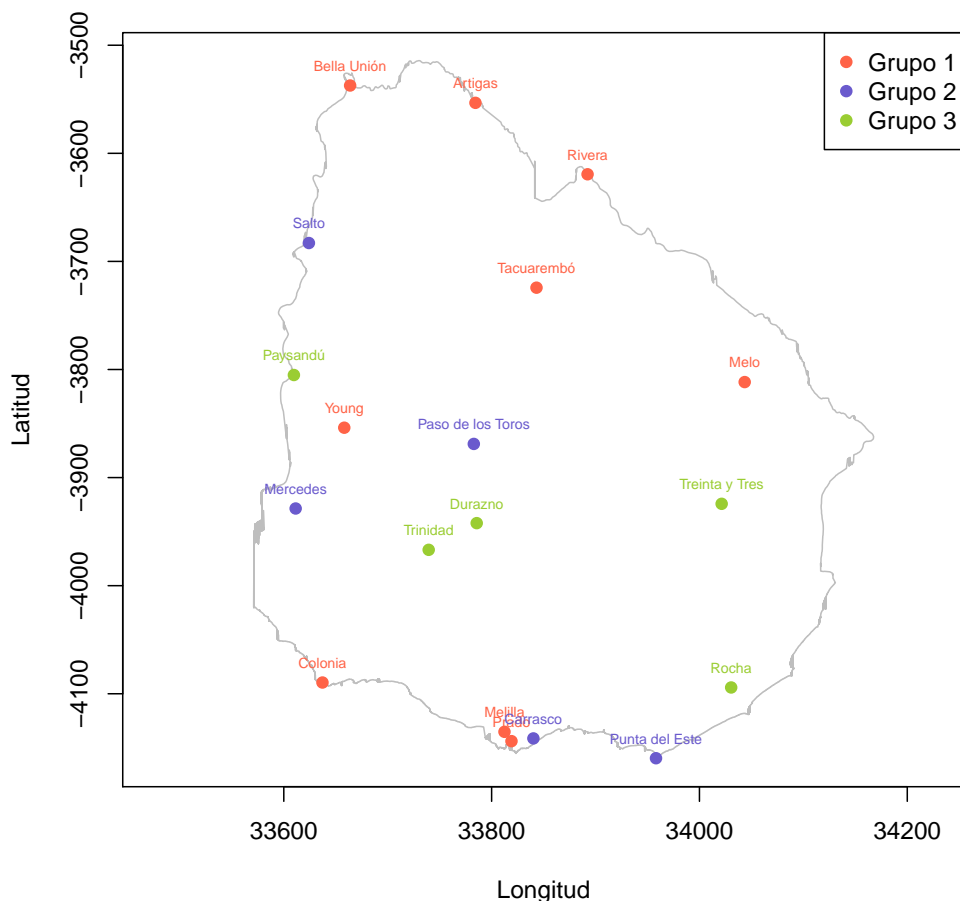


Figura D.5: Conformación de grupos - Análisis Longitudinal

Las estaciones que han presentado movilidad respecto de los grupos, según los distintos análisis, han sido Young, Paso de los Toros, Melilla y Prado.

Si se estudia detenidamente, bajo la distancia instantánea, Young, Melilla y Prado pertenecen al grupo 2, sin embargo, bajo la distancia longitudinal se movilizan todas al grupo 1. Lo anterior estaría indicando que el efecto del estudio longitudinal del tiempo hace que las estaciones mencionadas se agrupen en el grupo de mayor caída acumulada de granizo. Mientras que el efecto instantáneo (o transversal) del tiempo provoca agrupar a las mismas en el grupo cuya caída acumulada de granizo es media.

En cuanto a la estación ubicada en Paso de los Toros, bajo el enfoque instantáneo, la misma se encuentra en el grupo de media frecuencia acumulada de granizo, mientras que si se analiza mediante una distancia longitudinal la misma se traslada al grupo 1, el de mayor frecuencia acumulada de granizo.

Parte VII
Apéndice Código R

Apéndice E

Código en R

E.1. Análisis descriptivo

```
rm(list=ls())
library(xlsReadWrite)
library(xtable)

# Lee contorno de Uruguay
uru=read.xls("CONTORNO URUGUAY.xls")
dim(uru)
head(uru)

# Grafica Uruguay en coordenadas geográficas
plot(uru$LONG, uru$LATITUD, type="l",pch=20, cex=0.3)
dev.off()

# Convierte uruguay a coordenadas cartesianas
library(akima)
library(splancs)
library(GEOmap)

# Convierte coordenadas geográficas de Uruguay
# a coordenadas cartesianas X-Y mediante la
# conversión de Mercator Esférica

# Hace proyección de Mercator Esférica
proj = setPROJ(type = 1, LATS = uru$LATITUD, LONS = uru$LONG)

# Convierte a coordenadas cartesianas (km)
uru1= GLOB.XY(uru$LATITUD, uru$LONG, proj) #crea coord como "list"
uru2=data.frame(uru1)
head(uru2)
names(uru2)=c("longitud","latitud")
head(uru2)

# Grafica Uruguay con coordenadas cartesianas
plot(uru2$longitud, uru2$latitud, type="l",pch=20, cex=0.3, xlim=c(33475,34225))
dev.off()

# Lee datos
datos=read.xls("BASE KRIGING-FUZZY.xls")
dim(datos)
head(datos)
```

E.1. Análisis descriptivo

```
summary(datos[,c("tot","tmax")])

# Crea variable que es el promedio de granizada anual en 11 años
datos$prom=round(datos$tot/11,3)

#####
# Convierte coordenadas geográficas en datos a
# coordenadas cartesianas X-Y, mediante conversión
# de Mercator Esférica
#####

# Utiliza la proyección hecha para coords de Uruguay
# para proyectar las coordenadas de los datos
coord= GLOB.XY(datos$latitud, datos$longitud, proj) #crea coord como "list"
coord=data.frame(coord)
head(coord)
names(coord)=c("longitud","latitud")

# Agrega a "coord" la estación, el granizo y el tmax
datos1=data.frame(coord,est=datos[, "est"],tot=datos[, "tot"],tmax=datos[, "tmax"],prom=datos[, "prom"])
head(datos1)

# Medidas de resumen
s=summary(datos1[,c("tot","tmax","prom")])
xtable(s)

# Coeficientes de correlación
corr1=cor(datos1[,c("longitud","latitud","tot","tmax","prom")])
corr1
xtable(corr1)

#####

# Ubicación geográfica de las estaciones meteorológicas
plot(uru2$longitud,uru2$latitud, type="l", pch=20,cex=0.3,lwd=2,
col="grey",xlab="longitud",ylab="latitud",xlim=c(33475,34225))
points(datos1$longitud,datos1$latitud,pch=20,xlab="longitud",
ylab="latitud",cex=0.8)
text(datos1$longitud,datos1$latitud,datos1$est,cex=0.6,pos=3)

dev.off()
#####

# URUGUAY CON CUARTILES

# Gráfico granizo por cuantiles de caída acumulada de granizo en 11 años
q=quantile(datos1$tot,seq(0,1,0.25))
q1=datos1[datos1$tot<=q[2],]
q2=datos1[datos1$tot>q[2] & datos1$tot<=q[3],]
q3=datos1[datos1$tot>q[3] & datos1$tot<=q[4],]
q4=datos1[datos1$tot>q[4],]

plot(uru2$longitud,uru2$latitud, type="l", pch=20,cex=0.3,lwd=2,
col="grey",xlab="longitud",ylab="latitud",xlim=c(33475,34225))
points(q1$longitud,q1$latitud,pch=20,xlab="longitud",ylab="latitud",col="turquoise2")
text(q1$longitud,q1$latitud,q1$est,cex=0.8,pos=3,col="turquoise2")
points(q2$longitud,q2$latitud,pch=20,xlab="longitud",ylab="latitud",col="yellowgreen")
text(q2$longitud,q2$latitud,q2$est,cex=0.8,pos=3,col="yellowgreen")
points(q3$longitud,q3$latitud,pch=20,xlab="longitud",ylab="latitud",col="slateblue")
```

E.2. Interpolación por Kriging Ordinario

```
text(q3$longitud,q3$latitud,q3$est,cex=0.8,pos=3,col="slateblue")
points(q4$longitud,q4$latitud,pch=20,xlab="longitud",ylab="latitud",col="red")
text(q4$longitud,q4$latitud,q4$est,cex=0.8,pos=3,col="red")
legend("topright",legend=c("1er cuartil","2do cuartil","3er cuartil",
"4to cuartil"),pch=20,col=c("turquoise2","yellowgreen","slateblue","red"))

dev.off()

# Gráfico granizo por cuartiles del promedio anual de granizo
q=quantile(datos1$prom,seq(0,1,0.25))
q1=datos1[datos1$prom<=q[2],]
q2=datos1[datos1$prom>q[2] & datos1$prom<=q[3],]
q3=datos1[datos1$prom>q[3] & datos1$prom<=q[4],]
q4=datos1[datos1$prom>q[4],]

plot(uru2$longitud,uru2$latitud,type="l",pch=20,cex=0.3,lwd=2,
col="grey",xlab="longitud",ylab="latitud",xlim=c(33475,34225))
points(q1$longitud,q1$latitud,pch=20,xlab="longitud",ylab="latitud",col="turquoise2")
text(q1$longitud,q1$latitud,q1$est,cex=0.8,pos=3,col="turquoise2")
points(q2$longitud,q2$latitud,pch=20,xlab="longitud",ylab="latitud",col="yellowgreen")
text(q2$longitud,q2$latitud,q2$est,cex=0.8,pos=3,col="yellowgreen")
points(q3$longitud,q3$latitud,pch=20,xlab="longitud",ylab="latitud",col="slateblue")
text(q3$longitud,q3$latitud,q3$est,cex=0.8,pos=3,col="slateblue")
points(q4$longitud,q4$latitud,pch=20,xlab="longitud",ylab="latitud",col="red")
text(q4$longitud,q4$latitud,q4$est,cex=0.8,pos=3,col="red")
legend("topright",legend=c("1er cuartil","2do cuartil","3er cuartil",
"4to cuartil"),pch=20,col=c("turquoise2","yellowgreen","slateblue","red"))

dev.off()
```

E.2. Interpolación por Kriging Ordinario

```
rm(list=ls())
library(xlsReadWrite)
library(xtable)

# Lee datos
datos=read.xls('BASE KRIGING.xls')
head(datos)
dim(datos)
summary(datos)

# Calcula promedio anual de granizadas en el período de 11 años
datos$TOTAL=round(datos$TOTAL/11,3)

head(datos)
summary(datos)

# Lee contorno de Uruguay
uru=read.xls("CONTORNO URUGUAY.xls")
dim(uru)
head(uru)
tail(uru)

# Grafica Uruguay con coordenadas geográficas (expresadas en decimales)
```

E.2. Interpolación por Kriging Ordinario

```
plot(uru$LONG, uru$LATITUD, type="l",pch=20, cex=0.3)

dev.off()

# Convierte coordenadas de Uruguay a coordenadas cartesianas
library(akima)
citation("akima")
library(splancs)
citation("splancs")
library(GEOmap)
citation("GEOmap")

# Convierte coordenadas geográficas de Uruguay
# a coordenadas cartesianas X-Y mediante la
# conversión Esférica de Mercator

# Hace proyección de Mercator Esférica
proj = setPROJ(type = 1, LATS = uru$LATITUD, LONS = uru$LONG)

# Convierte a coordenadas cartesianas (km)
uru1= GLOB.XY(uru$LATITUD, uru$LONG, proj) #crea coord como "list"
uru2=data.frame(uru1)
names(uru2)=c("LONGITUD", "LATITUD")
head(uru2)

# Grafica Uruguay en coordenadas cartesianas
plot(uru2$LONGITUD, uru2$LATITUD, type="l",pch=20, cex=0.3, xlim=c(33475,34225))

dev.off()

#####
# Convierte coordenadas geográficas en datos a
# coordenadas cartesianas X-Y, mediante conversión
# Esférica de Mercator
#####

# Utiliza la proyección hecha para coords de Uruguay
# para proyectar las coordenadas de los datos
coord= GLOB.XY(datos$LATITUD, datos$LONGITUD, proj) #crea coord como "list"
coord=data.frame(coord)
names(coord)=c("LONGITUD", "LATITUD")

datos1=data.frame(coord,TOTAL=datos[, "TOTAL"],EST=datos[, "est"])
head(datos1)

#####

# Ubicación geográfica de las estaciones
plot(uru2$LONGITUD,uru2$LATITUD, type="l", pch=20,cex=0.3,lwd=2,
col="grey", xlab="Longitud",ylab="Latitud",xlim=c(33475,34225))
points(datos1$LONGITUD,datos1$LATITUD,pch=20,xlab="LONGITUD",
ylab="LATITUD",cex=0.8)
text(datos1$LONGITUD,datos1$LATITUD,datos1$EST,cex=0.6,pos=3)

dev.off()

#####
##### ANÁLISIS DE TENDENCIA #####
```

E.2. Interpolación por Kriging Ordinario

```
#####

library(nnet)
library(lmtest)
library(car)
library(tseries)
citation("tseries")

# Matriz de correlación
corr=cor(datos1[,c("LONGITUD", "LATITUD", "TOTAL")])
corr
xtable(corr)

#####
## Escenario 1: Modelo con polinomio de primer orden
#####

# Modelo 2:
modelo2=lm(TOTAL~LONGITUD+LATITUD-1,data=datos1)
summary(modelo2)

#####
##### CUMPLIMIENTO #####
##### DEL #####
##### SUPUESTOS #####
##### MODELO #####
#####

### NORMALIDAD
# Testea gráficamente normalidad de residuos
hist(modelo2$residuals, probability=T)
lines(density(modelo2$residuals))
boxplot(modelo2$residuals)
qq.plot(modelo2$residuals)
plot(density(modelo2$residuals))
dev.off()

# Testea normalidad residuos
# Kolmogorov-Smirnov
ks.test(modelo2$residuals, "pnorm")
# Shapiro-Wilk
shapiro.test(modelo2$residuals)
# Jarque Bera
jarque.bera.test(modelo2$residuals)

### HOMOSCEDASTICIDAD
# Score test
ncv.test(modelo2)
# Breusch-Pagan
bptest(modelo2)
# Goldfeld-Quandt
gqtest(modelo2)

### INCORRELACIÓN
# Test de Durbin-Watson
dwtest(modelo2)

### Evalúa la superficie de tendencia estimada en la grilla ###
```

E.2. Interpolación por Kriging Ordinario

```
# Crea grilla
rlong1=range(uru2$LONGITUD)
rlat1=range(uru2$LATITUD)
grilla<-expand.grid(seq(rlong1[1],rlong1[2],20), seq(rlat1[1],rlat1[2],20))
names(grilla)
names(grilla)=c('LONGITUD','LATITUD')
head(grilla)

# Se modifica tamaño de grilla

grilla3<-expand.grid(seq(rlong1[1],rlong1[2],10), seq(rlat1[1],rlat1[2],10))
names(grilla3)=c('LONGITUD','LATITUD')
head(grilla3)

# Evalúa tendencia estimada en la grilla 3

pred3=predict(modelo2,newdata=grilla3)
head(pred3)
length(pred3)

# Grafica en 3D la superficie de tendencia estimada
# en la grilla 3

library(scatterplot3d)
library(lattice)

scatterplot3d(grilla3$LONGITUD, grilla3$LATITUD, pred3, xlab="Longitud",
ylab="Latitud",zlab="Promedio anual de granizo",pch=20,cex.symbols=0.3,color="slateblue",angle=30)

cloud(pred3~grilla3$LONGITUD*grilla3$LATITUD,xlab="Longitud",
ylab="Latitud",zlab="Granizo",col=2,screen=list(z=60,x=-60,y=3),
pch=20,cex=0.3)

dev.off()

#####
##### ESTIMACIÓN DEL VARIOGRAMA #####
##### CON RESIDUOS #####
#####

# Transforma los datos a tipo "geodata"
library(geoR)
citation("geoR")

dat.geo<-as.geodata(datos1)
plot.geodata(dat.geo)

# Eliminación de tendencia
dat.sin=modelo2$residuals
mean(dat.sin)
hist(dat.sin)
plot(dat.sin)
abline(h=0)
dev.off()

# Utiliza como datos (Z) a los residuos (e)
dats.geo<-dat.geo
dats.geo$data<-dat.sin
```

E.2. Interpolación por Kriging Ordinario

```
plot.geodata(dats.geo)
dev.off()

#####
##### VARIOGRAMA EXPERIMENTAL #####
#####

## REALIZA VARIOGRAMA CON RESIDUOS ##

# Estimador variograma clásico
varios<-variog(dats.geo)
plot(varios,pch=20)

## Superposición del variograma empírico y de un modelo de variograma

# Se escoge variograma experimental
variog=variog(dats.geo,uvec=seq(0,400,30))
plot(variog,pch=20)
dev.off()

# Genera valores iniciales de sill y rango para ajustar los modelos de variogramas
ini.vals<-expand.grid(seq(0,1,by=0.1),seq(0,400,by=20) )

# Diferentes modelos de ajuste del variograma
wave=variofit(variog, ini = ini.vals,cov.model="wave")
summary(wave)
expon=variofit(variog, ini = ini.vals,cov.model="exponential")
summary(expon)
spheric=variofit(variog, ini = ini.vals,cov.model="spherical")
summary(spheric)
matern=variofit(variog, ini = ini.vals)
summary(matern)
circ=variofit(variog, ini = ini.vals,cov.model="circular")
summary(circ)
gauss=variofit(variog, ini = ini.vals,cov.model="gaussian")
summary(gauss)
cub=variofit(variog, ini = ini.vals,cov.model="cubic")
summary(cub)
pow.exp=variofit(variog, ini = ini.vals,cov.model="powered.exponential")
summary(pow.exp)
cauch=variofit(variog, ini = ini.vals,cov.model="cauchy")
summary(cauch)

plot(variog, pch=20, ylim=c(0,0.7),xlab="Distancia",ylab="Semivariograma")
lines(expon,col='cyan',lwd=3)
lines(spheric,col='brown',lwd=3)
lines(matern,col='slateblue',lwd=3)
lines(circ,col='violet',lwd=3)
lines(gauss,col='magenta',lwd=3)
lines(cauch,col='pink',lwd=3)
lines(cub,col='orange',lwd=3)
lines(pow.exp,col='olivedrab',lwd=3)
lines(wave,col='tomato',lwd=3)
legend("topright",legend=c("Seno cardinal","Exponencial","Esférico","Matemático",
"Circular","Gaussiano","Cauchy","Cúbico","Power Exponencial"),pch=20,lwd=2,
col=c("tomato","cyan","brown","slateblue","violet","magenta","pink","orange","olivedrab"))

# Se selecciona el siguiente modelo de variograma
wave1=variofit(variog, ini = ini.vals,cov.model="wave")
```


E.2. Interpolación por Kriging Ordinario

```
wave2=variofit(variog, ini = ini.vals,cov.model="wave",wei="equal")
wave3=variofit(variog, ini = ini.vals,cov.model="wave",wei="equal",minimisation.function="nlm")

plot(variog, pch=20, xlab="Distancia",ylab="Semivariograma")
lines(wave1,col='olivedrab',lwd=3,lty=2)
lines(wave2,col='tomato',lwd=3)
lines(wave3,col='slateblue',lwd=3,lty=1)

dev.off()

##### VALIDACIÓN DEL MODELO DE VARIOGRAMA #####
##### MEDIANTE CROSS-VALIDATION #####

crossv_wave=xvalid(dats.geo, model=wave)
Ewave=sqrt(sum(crossv_wave$error^2)/19) #ECM

crossv2=xvalid(dats.geo, model=wave2)
Ewave2=sqrt(sum(crossv2$error^2)/19) #ECM

crossv3=xvalid(dats.geo, model=wave3)
Ewave3=sqrt(sum(crossv3$error^2)/19) #ECM

crossv_sph=xvalid(dats.geo, model=spheric)
EspH=sqrt(sum(crossv_sph$error^2)/19) #ECM

crossv_cau=xvalid(dats.geo, model=cauch)
Ecau=sqrt(sum(crossv_cau$error^2)/19) #ECM

crossv_pow=xvalid(dats.geo, model=pow.exp)
Epow=sqrt(sum(crossv_pow$error^2)/19) #ECM

crossv_mat=xvalid(dats.geo, model=matern)
Emat=sqrt(sum(crossv_mat$error^2)/19) #ECM

crossv_cub=xvalid(dats.geo, model=cub)
Ecub=sqrt(sum(crossv_cub$error^2)/19) #ECM

crossv_cir=xvalid(dats.geo, model=circ)
Ecir=sqrt(sum(crossv_cir$error^2)/19) #ECM

crossv_exp=xvalid(dats.geo, model=expon)
Eexp=sqrt(sum(crossv_exp$error^2)/19) #ECM

crossv_gauss=xvalid(dats.geo, model=gauss)
Egauss=sqrt(sum(crossv_gauss$error^2)/19) #ECM

E=c(Ewave,Ewave2,Ewave3,Esph, Ecau, Epow, Emat, Ecub, Ecir, Eexp, Egauss)
E
min(E) # Corresponde al variograma wave2

# Grafica valores predichos versus observados
crossv_plot=data.frame(pred=crossv2$predicted,est=datos1$EST,data=crossv2$data)
plot(crossv_plot$pred, crossv_plot$data, xlim=range(crossv_plot$data),
ylim=range(crossv_plot$data), pch=20,xlab="Residuos predichos",
ylab="Residuos observados",col="tomato")
text(crossv_plot$pred,crossv_plot$data,crossv_plot$est,cex=0.6,pos=3)
abline(0,1)
```

E.2. Interpolación por Kriging Ordinario

```
# Hace crossvalidation con la varianza de Kriging
crossv2=data.frame(crossv2$krige.var,crossv2$error,datos1$EST)
head(crossv2)
crossv2[crossv2$crossv2.krige.var<0.2,]

##### CONCLUSIÓN: EL MENOR ECM SE LOGRA UTILIZANDO EL
##### VARIOGRAMA WAVE2

#####
##### KRIGING ORDINARIO #####
##### CON RESIDUOS #####
#####

# Utiliza misma grilla que la construída en tendencia

# Grafica Uruguay con grilla de interpolación
plot(grilla3,cex=0.3,pch=20,xlim=c(33475,34225),col="gray", xlab="Longitud",ylab="Latitud")
points(dat.geo$coords,pch=20)
text(datos1$LONGITUD,datos1$LATITUD,datos1$est,cex=0.6,pos=3)
points(uru2$LONGITUD,uru2$LATITUD, pch=20,cex=0.3, xlab="Longitud",ylab="Latitud",type="l")

dev.off()

# Hace kriging ordinario con residuos para las diferentes grillas

kc1<-krige.conv(dats.geo,locations=grilla,krige=krige.control(
cov.pars=wave2$cov.pars,nugget=wave2$nugget))

kc2<-krige.conv(dats.geo,locations=grilla2,krige=krige.control(
cov.pars=wave2$cov.pars,nugget=wave2$nugget))

kc3<-krige.conv(dats.geo,locations=grilla3,krige=krige.control(
cov.pars=wave2$cov.pars,nugget=wave2$nugget))

kc4<-krige.conv(dats.geo,locations=grilla4,krige=krige.control(
cov.pars=wave2$cov.pars,nugget=wave2$nugget))

#####

# Resumen de varianzas de Kriging según diferentes grillas

summary(kc1$krige.var)
summary(kc2$krige.var)
summary(kc3$krige.var)
summary(kc4$krige.var)

plot(kc1$krige.var)
plot(kc2$krige.var)
plot(kc3$krige.var)
plot(kc4$krige.var)

#####
#### SEGÚN LAS MEDIDAS DE RESUMEN ANTERIORES ####
#### SE SELECCIONA LA GRILLA3 #####

##### GRÁFICOS DE KRIGING #####
```

E.3. Análisis de Grupos

```
# Uruguay con isolíneas
contour(kc3,val=kc3$predict+pred3,borders=uru2,col="olivedrab")
points(dat.geo$coords, pch=20, col="tomato")

# Valores estimados con función "image"
image.kriging(kc3,loc=grilla3,val=kc3$predict+pred3,main='estimacion kriging',borders=uru2)
contour(kc3,val=kc3$predict+pred3,borders=uru2,add=TRUE)

# Residuos estimados con función "image"
image.kriging(kc3,loc=grilla3,val=sqrt(kc3$krige.var),main='error estandar',borders=uru2)
contour(kc3,borders=uru2,add=TRUE)

# Gráficos en 3D
persp.kriging(kc3,loc=grilla3,val=kc3$predict+pred3,xlab="Longitud",
ylab="Latitud",zlab="Estimación granizo",phi=30,theta=60,col='rosybrown',shade=0.25)
persp.kriging(kc3,loc=grilla3,val=sqrt(kc3$krige.var),xlab="Longitud",
ylab="Latitud",zlab="Error estimación",phi=30,theta=60,col=2,shade=0.25)

#####
# GRÁFICO URUGUAY CON FUNCIÓN CONTOUR
#####

# Estimación Z
contour(kc3,val=kc3$predict+pred3,coords.data=dat.geo$coords,filled=TRUE
borders=uru2,color.palette=colorRampPalette(c('green', 'olivedrab','red'),
space = "rgb"))

# Error
contour(kc3,val=sqrt(kc3$krige.var,filled=TRUE,borders=uru2,
color.palette=colorRampPalette(c('red','olivedrab','lemonchiffon'),
space = "rgb"))
```

E.3. Análisis de Grupos

E.3.1. Análisis de Grupos Jerárquico

```
#####
## Crea matriz de disimilaridad con distancias de Carlier
#####

# Lee array de datos
mi.array=read.table("Array de datos.txt", sep='\t', dec='.')

# Carga función de Carlier
source('carlier.R')

# Calcula distancia de Carlier mixta,
# dando el mismo peso a cada distancia (0.5)
D=carlier(mi.array, 0.5)

# Calcula distancia instantánea
Dinstant=carlier(mi.array,1)
```

E.3. Análisis de Grupos

```
# Calcula distancia longitudinal
Dlongit=carlier(mi.array,0)

# Guarda distancias calculadas
write.table(D,'Matriz disimilaridad Carlier MIXTA.txt')
write.table(Dinstant,'Matriz disimilaridad Carlier INSTANTANEA.txt')
write.table(Dlongit,'Matriz disimilaridad Carlier LONGITUDINAL.txt')

#####

### Hace análisis de cluster con matriz de disimilaridad
### de Carlier y variables granizo acumulado y tamaño
### máximo de piedra promedio

rm(list=ls())
library(xlsReadWrite)

library(cluster)
citation("cluster")
source('standard.r')
source('indicadores.r')

# Observa distribución de granizo y tmax previamente
plot(datos$tot, datos$tmax, pch=20,xlab="Granizo acumulado",
ylab="Tamaño máximo de piedra", main="Diagrama de dispersión")
text(datos$tot, datos$tmax, datos$est, cex=0.6, pos=3)

# Lee datos
datos=read.xls("BASE KRIGING-FUZZY.xls")
dim(datos)
head(datos)

# Lee matriz de disimilaridad de Carlier mixta
D=read.table("Matriz disimilaridad Carlier MIXTA.txt")
dim(D)
head(D)

# Hace análisis de cluster con matriz D MIXTA
gruposw<-agnes(D, diss= TRUE, stand = FALSE, method = "ward")
gruposss<-agnes(D, diss= TRUE, stand = FALSE, method = "single")
gruposc<-agnes(D, diss= TRUE, stand = FALSE, method = "complete")
gruposa<-agnes(D, diss= TRUE, stand = FALSE, method = "average")

indw<-indicadores(gruposw[4], D)
inds<-indicadores(gruposss[4], D)
indc<-indicadores(gruposc[4], D)
inda<-indicadores(gruposa[4], D)

# Dendrograma
plot(gruposw,which=2, main="", sub="")
#plot(gruposss,which=2)
#plot(gruposc,which=2)
#plot(gruposa,which=2)

k=3
cl=as.factor(cutree(gruposw[4],k))
table(cl)
```

E.3. Análisis de Grupos

```
si=silhouette(cutree(gruposw, k = 2),dist=D)
plot(si)
dev.off()

datos=cbind(datos,c1)

# Caracterización de grupos
par(mfrow=c(1,2))
boxplot(datos$tot~datos$c1)
boxplot(datos$tmax~datos$c1)

# Caracterización según granizo
by(datos$tot1, datos$c1, summary)

# Caracterización según tamaño máximo
by(datos$tmaxt, datos$c1, summary)

# Caracterización según departamentos
by(datos$dep, datos$c1, summary)
addmargins(table(datos$dep, c1))

# Caracterización según años
by(datos$año, datos$c1, summary)
```

E.3.2. Análisis de Grupos Difusos

```
rm(list=ls())
library(xlsReadWrite)
library(cluster)
library(xtable)

# Lee datos
datos=read.xls("BASE KRIGING-FUZZY.xls")
dim(datos)
head(datos)
summary(datos[,c("tot","tmax")])

# Transforma a promedio anual la caída de granizo acumulada
datos$tot=round(datos$tot/11,3)
head(datos)
summary(datos[,c("tot","tmax")])

# Lee base con contorno de Uruguay
uru=read.xls("CONTORNO URUGUAY.xls")
dim(uru)
head(uru)

# Convierte uruguay a coordenadas cartesianas
library(akima)
library(splancs)
library(GEOmap)

# Convierte coordenadas geográficas de Uruguay
# a coordenadas cartesianas X-Y mediante la
# conversión de Mercator Esférica
```

E.3. Análisis de Grupos

```
# Hace proyección de Mercator Esférica
proj = setPROJ(type = 1, LATS = uru$LATITUD, LONS = uru$LONG)

# Convierte a coordenadas cartesianas (km)
uru1= GLOB.XY(uru$LATITUD, uru$LONG, proj) #crea coord como "list"
uru2=data.frame(uru1)
head(uru2)
names(uru2)=c("LONGITUD", "LATITUD")
head(uru2)

# Observa máximos y mínimos de latitud y longitud
rlong=range(uru2$LONGITUD)
rlong
rlat=range(uru2$LATITUD)
rlat
plot(uru2$LONGITUD, uru2$LATITUD, pch=20, cex=0.3,type="l", xlim=c(33450,34250))

dev.off()

#####
# Convierte coordenadas geográficas en datos a
# coordenadas cartesianas X-Y, mediante conversión
# de Mercator Esférica
#####

# Utiliza la proyección hecha para coords de Uruguay
# para proyectar las coordenadas de los datos
coord= GLOB.XY(datos$latitud, datos$longitud, proj) #crea coord como "list"
coord=data.frame(coord)
head(coord)
names(coord)=c("longitud", "latitud")

# Agrega a "coord" la estación, el granizo promedio y el tmax
datos1=data.frame(coord,est=datos[, "est"],tot=datos[, "tot"],tmax=datos[, "tmax"])
head(datos1)

#####
##### FUZZY CON DISTANCIA MIXTA #####
#####

# Lee matriz de disimilaridad de Carlier mixta
D=read.table("Matriz disimilaridad Carlier MIXTA.txt")
dim(D)
head(D)

# Hace cluster fuzzy con función "fanny"
fuzzy=fanny(D, 3, diss = TRUE, memb.exp = 1.45,
            maxit = 1000, tol = 1e-15, trace.lev = 0)

summary(fuzzy)

par(mfrow=c(1,2))
# Grafica grupos
plot(fuzzy, ask = FALSE, which.plots = 1,
     nmax.lab = 40, max.strlen = 5, data = D, dist = NULL,
     cor = TRUE, stand = FALSE, lines = 2,main="")

# Grafica Silhouette
```

E.3. Análisis de Grupos

```
plot(fuzzy, ask = FALSE, which.plots = 2,
nmax.lab = 40, max.strlen = 5, data = D, dist = NULL,
cor = TRUE, stand = FALSE, lines = 2, main="")

# Agrega coeficientes de pertenencia a base
memb=data.frame(round(fuzzy$membership*100,1))
names(memb)=c("u1", "u2", "u3")

datos2=data.frame(datos1,memb)
head(datos2)

# Busca los máximos de todos los coefs de pertenencias
datos2$maxu=0
for(i in 1:nrow(datos2)){
datos2$maxu[i]=max(datos2[i,c("u1", "u2", "u3")])
}
head(datos2)

# Trabajamos con la silhouette info
si=fuzzy$silinfo
sil=data.frame(si$widths)
summary(sil)

#####
###write.table(sil,"silu.txt")
#####

# Lee la información dentro de "sil.txt" para lograr obtener los
# ids que están desordenados
datossi=read.table("silu.txt", header=T)
head(datossi)

# Agrega id a la base de datos
datos2$id=seq(1,nrow(datos2),1)

# Une "datos" con "sil" según id
datos3=merge(datos2,datossi[,c("id","cluster","neighbor","sil_width")],by="id")
datos3
summary(datos3)

by(datos3$sil_width, datos3$cluster, mean)

# Cantidad de observaciones por grupos
tabla1=table(datos3$cluster)
tabla1

#####
##### CARACTERIZACIÓN #####
#####

# Distribución de granizo en los clusters
boxplot(datos3$tot~datos3$cluster)
# Distribución del tamaño de piedra por clusters
boxplot(datos3$tmax~datos3$cluster)

par(mfrow=c(1,2))
boxplot(datos3$tot~datos3$cluster, xlab="Grupos",ylab="Granizo acumulado")
```

E.3. Análisis de Grupos

```
boxplot(datos3$txmax~datos3$cluster,xlab="Grupos",ylab="Tamaño máximo piedra")

dev.off()

by(datos3$tot, datos3$cluster, summary)
by(datos3$txmax, datos3$cluster, summary)

#####
##### ANÁLISIS DESCRIPTIVO #####
#####

# Analiza coeficientes de pertenencia por clusters

op=par(mfrow=c(1,3))
boxplot(datos3$u1[datos3$cluster==1],ylim=c(0,100),main="Coef. pertenencia del grupo 1")
boxplot(datos3$u2[datos3$cluster==2],ylim=c(0,100),main="Coef. pertenencia del grupo 2")
boxplot(datos3$u3[datos3$cluster==3],ylim=c(0,100),main="Coef. pertenencia del grupo 3")
par(op)
dev.off()

summary(datos3$u1[datos3$cluster==1])
summary(datos3$u2[datos3$cluster==2])
summary(datos3$u3[datos3$cluster==3])

op=par(mfrow=c(1,3))
boxplot(datos3$u1~datos3$cluster,ylim=c(0,100),main="Coef. pertenencia al grupo 1")
boxplot(datos3$u2~datos3$cluster,ylim=c(0,100),main="Coef. pertenencia al grupo 2")
boxplot(datos3$u3~datos3$cluster,ylim=c(0,100),main="Coef. pertenencia al grupo 3")
par(op)
dev.off()

# Cruza departamentos según grupos
table(datos3$est, datos3$cluster)

# Busca los segundos máximos de todos los coefs de pertenencias
datos3$dif1=datos3$maxu-datos3$u1
datos3$dif2=datos3$maxu-datos3$u2
datos3$dif3=datos3$maxu-datos3$u3

datos3$min2=0
for(i in 1:nrow(datos3)){
  datos3$min2[i]=min(datos3[i,c("dif1","dif2","dif3")][datos3[i,c("dif1","dif2","dif3")]!=0])
}
head(datos3)

# Define difusas y las selecciona
difusas=datos3[datos3$maxu<50 & datos3$min2<=15,]
difusas

# Crea tabla para colocar en anexo del informe
anexo=datos3[,c("est","u1","u2","u3","cluster","neighbor")]
anexo

# Convierte a latex
xtable(anexo, digits=0,caption="Estaciones meteorológicas y sus coeficientes de pertenencia")

### Gráficamente ###
```


E.3. Análisis de Grupos

```
# Grafica a las estaciones en función de los clusters
plot(uru2$LONG, uru2$LATITUD, pch=20, cex=0.3, col="grey",
xlab="Longitud",ylab="Latitud", type="l",xlim=c(33475,34225))
points(datos3$longitud[datos3$cluster==1],datos3$latitud[datos3$cluster==1],
pch=19,col="tomato")
text(datos3$longitud[datos3$cluster==1],datos3$latitud[datos3$cluster==1],
labels=levels(datos3$est)[datos3$cluster==1],cex=0.6,pos=3, col="tomato")
points(datos3$longitud[datos3$cluster==2],datos3$latitud[datos3$cluster==2],
col="slateblue", pch=19)
text(datos3$longitud[datos3$cluster==2],datos3$latitud[datos3$cluster==2],
labels=levels(datos3$est)[datos3$cluster==2],cex=0.6,pos=3, col="slateblue")
points(datos3$longitud[datos3$cluster==3],datos3$latitud[datos3$cluster==3],
col="yellowgreen", pch=19)
text(datos3$longitud[datos3$cluster==3],datos3$latitud[datos3$cluster==3],
labels=levels(datos3$est)[datos3$cluster==3],cex=0.6,pos=3, col="yellowgreen")
legend("topright", legend = c("Grupo 1","Grupo 2","Grupo 3"),
col=c("tomato","slateblue","yellowgreen"),pch=19)

#####
# Cluster y vecinos más cercanos

plot(uru2$LONG, uru2$LATITUD, pch=20, cex=0.3, col="grey", xlab="Longitud",ylab="Latitud",
type="l",xlim=c(33475,34225), main="Grupos originales")
points(difusas$longitud[difusas$cluster==1],difusas$latitud[difusas$cluster==1],
pch=19,col="tomato")
text(difusas$longitud[difusas$cluster==1],difusas$latitud[difusas$cluster==1],
labels=difusas$est[difusas$cluster==1],cex=0.6,pos=3, col="tomato")
points(difusas$longitud[difusas$cluster==2],difusas$latitud[difusas$cluster==2],
col="slateblue", pch=19)
text(difusas$longitud[difusas$cluster==2],difusas$latitud[difusas$cluster==2],
labels=difusas$est[difusas$cluster==2],cex=0.6,pos=3, col="slateblue")
points(difusas$longitud[difusas$cluster==3],difusas$latitud[difusas$cluster==3],
col="yellowgreen", pch=19)
text(difusas$longitud[difusas$cluster==3],difusas$latitud[difusas$cluster==3],
labels=difusas$est[difusas$cluster==3],cex=0.6,pos=3, col="yellowgreen")
legend("topright", legend = c("Grupo 1","Grupo 2","Grupo 3"),
col=c("tomato","slateblue","yellowgreen"), pch=19)

plot(uru2$LONG, uru2$LATITUD, pch=20,cex=0.3,col="grey",xlab="Longitud",ylab="Latitud",
type="l",xlim=c(33475,34225),main="Grupos vecinos")
points(difusas$longitud[difusas$cluster==1 & difusas$neighbor==2],
difusas$latitud[difusas$cluster==1 & difusas$neighbor==2],pch=19,col="slateblue")
text(difusas$longitud[difusas$cluster==1 & difusas$neighbor==2],
difusas$latitud[difusas$cluster==1 & difusas$neighbor==2],
labels=difusas$est[difusas$cluster==1 & difusas$neighbor==2],
cex=0.6,pos=3, col="slateblue")
points(difusas$longitud[difusas$cluster==2 & difusas$neighbor==1],
difusas$latitud[difusas$cluster==2 & difusas$neighbor==1], pch=19, col="tomato")
text(difusas$longitud[difusas$cluster==2 & difusas$neighbor==1],
difusas$latitud[difusas$cluster==2 & difusas$neighbor==1],
labels=difusas$est[difusas$cluster==2 & difusas$neighbor==1],cex=0.6,pos=3, col="tomato")
points(difusas$longitud[difusas$cluster==2 & difusas$neighbor==3],
difusas$latitud[difusas$cluster==2 & difusas$neighbor==3], col="yellowgreen", pch=19)
text(difusas$longitud[difusas$cluster==2 & difusas$neighbor==3],
difusas$latitud[difusas$cluster==2 & difusas$neighbor==3],
labels=difusas$est[difusas$cluster==2 & difusas$neighbor==3],
```

E.3. Análisis de Grupos

```
cex=0.6,pos=3, col="yellowgreen")
points(difusas$longitud[difusas$cluster==3 & difusas$neighbor==2],
difusas$latitud[difusas$cluster==3 & difusas$neighbor==2],col="slateblue", pch=19)
text(difusas$longitud[difusas$cluster==3 & difusas$neighbor==2],
difusas$latitud[difusas$cluster==3 & difusas$neighbor==2],
labels=difusas$est[difusas$cluster==3 & difusas$neighbor==2],
cex=0.6,pos=3, col="slateblue")
legend("topright", legend = c("Grupo 1","Grupo 2","Grupo 3"),
col=c("tomato","slateblue","yellowgreen"), pch=19)
```

Índice de figuras

5.1. Mapa físico de Uruguay	16
6.1. Sistema de coordenadas geográficas	22
6.2. Ejemplo de Variograma	33
8.1. Ubicación geográfica estaciones meteorológicas	46
8.2. Promedio anual de granizada según cuartiles	48
8.3. Representación de la superficie de tendencia	52
8.4. Variograma experimental	53
8.5. Variograma experimental ajustado con modelo seno cardinal	54
8.6. Grilla de interpolación	56
8.7. Curvas de nivel estimación por Kriging Ordinario	57
8.8. Estimación por Kriging Ordinario	58
8.9. Desviación estándar estimación por Kriging Ordinario	59
8.10. Residuos predichos versus residuos observados	61
8.11. Diagramas de cajas grupos conformados	65
8.12. Conformación de grupos - Enfoque Mixto	66
8.13. Grupos originales	68
8.14. Grupos vecinos	68
A.1. Modelo Efecto Pepita	81
A.2. Modelo esférico	82
A.3. Modelo exponencial	82
A.4. Modelo gaussiano	83
A.5. Modelo cúbico	83
A.6. Modelo seno cardinal	84
A.7. Modelo potencia	84
A.8. Relación variograma-covariograma	85
C.1. Ajuste variogramas experimentales. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro	92

ÍNDICE DE FIGURAS

C.2. Estimaciones Kriging. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro	93
C.3. Error estimación Kriging. Arriba: Escenarios uno y dos. Abajo: Escenarios tres y cuatro	94
C.4. Variograma experimental ajustado con distintos modelos	97
C.5. Estimación Kriging	98
C.6. Error estándar Kriging	98
C.7. Estimación Kriging	98
C.8. Error estándar Kriging	98
D.1. Dendrograma Método Ward	99
D.2. Grupos y centroides - Análisis Mixto	100
D.3. Información Silhouette - Análisis Mixto	100
D.4. Conformación de grupos - Análisis Instantáneo	102
D.5. Conformación de grupos - Análisis Longitudinal	103

Índice de cuadros

8.1. Estaciones meteorológicas en estudio	45
8.2. Matriz de correlaciones lineales	47
8.3. Medidas de resumen variables de interés	47
8.4. Cuartiles del promedio anual de caída de granizo	48
8.5. Estimaciones, errores estándar y estadístico de prueba modelo regresión	51
8.6. Frecuencias grupos	64
8.7. Caída de granizo y tamaño máximo de piedra según grupos	64
8.8. Coeficientes de pertenencia según grupos (en %)	67
D.1. Estaciones meteorológicas y sus coeficientes de pertenencia	101