Geoestadística en datos circulares

Alejandro Niño Chaparro

Universidad Nacional de Colombia
Facultad, Departamento de Estadística
Bogotá, Colombia

Geostatistics in circular data

Alejandro Niño Chaparro

National University of Colombia
Science Faculty, Statistics Department
Bogotá, Colombia

Geoestadística en datos circulares

Alejandro Niño Chaparro

Tesis presentada como requisito parcial para optar al título de
Magister en Estadística

Director:
PhD Ramón Giraldo Henao

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento Estadística
Bogotá, Colombia
2023

Dedicado a mis padres, esas bellas personas que hicieron posible vivir esta nueva experiencia.

Agradecimientos

Agradezco a mi familia, especialmente a mis padres y mi tío Juane por haber permitido tener esta oportunidad. También manifiesto mi gratitud con el profesor Ramón Giraldo por su orientación y dedicación, a Camilo Cotamo, Sebastián Barón, Jonatan Barón y Cristian Avendaño por su amistad.

Resumen

Se propone una nueva metodología en el contexto de geostadística no estacionaria que permite hacer predicción de datos circulares empleando kriging circular residual cuando la tendencia espacial es modelada a través de redes neuronales. Usando datos simulados y reales (tomados del proyecto NASA power) se hace comparación de la técnica propuesta con pulimento de medianas. Los resultados indican que la estrategia considerada mejora las predicciones.
Palabras clave: Datos circulares; geoestadística no estacionaria; kriging circular; pulimento de medianas; redes neuronales.

Abstract

We propose a new methodology in the context of nonstationary geostatistics that allows the prediction of circular data using residual circular kriging when the spatial trend is modeled through neural networks. Using simulated and real data (taken from the NASA power project), the proposed technique is compared with those obtained through median polish. The results indicate that the strategy proposed improves the predictions.
Keywords: Circular kriging, nonstationary geostatistics, neural networks, median polish, directional data.

Contenido

Agradecimientos VII

Resumen IX

1 Introducción 2

2 Marco Teórico 5

2.1 Geoestadística con datos de variables definidas en los reales . . 5

2.1.1 Geoestadística bajo estacionariedad 6

2.1.2 Simulación en geoestadística 7

2.1.3 Aplicaciones 8

2.2 Funciones para caracterizar la dependencia espacial con datos reales 8

2.2.1 Modelos teóricos de covariograma 8

2.2.2 Estimación del modelo de covariograma 9

2.3 Kriging ordinario con variables definidas en los reales 11

2.4 Geoestadística no estacionaria 12

2.5 Geoestadística no estacionaria basada en redes neuronales . . 13

2.6 Datos circulares 15

2.6.1 Distribución von Mises 16

2.6.2 Medidas descriptivas de datos angulares 18

2.7 Geoestadística con datos circulares 20

2.7.1 Modelos teóricos de covariograma circular 20

2.7.2 Función cosenograma 21

2.7.3 Simulación de campos aleatorios circulares 23

2.7.4 Kriging ordinario con datos circulares 23
2.7.5 Kriging residual de datos circulares usando pulimento de medianas 26

3 Geoestadística circular no estacionaria basada en redes neuronales 32
3.1 Kriging residual de datos circulares basado en redes neuronales 32
3.1.1 Estimación de la tendencia espacial empleando redes neuronales 32
3.1.2 Kriging residual circular basado en redes neuronales . . 34

4 Estudio de simulación 38
4.1 Resultados 39

5 Análisis de datos reales: Predicción espacial de dirección del viento en una región de Colombia 44
5.1 Análisis de validación cruzada 45
5.2 Predicción en sitios no muestreados y caracterización de los vientos 49

6 Conclusiones y trabajo futuro 51

Bibliografía 53

1 Introducción

La geostadística comenzó su desarrollo en los años 50 con el objetivo de modelar procesos estocásticos con continuidad espacial [35]. Este campo de la estadística espacial tiene aplicaciones en distintos escenarios aplicados [50], entre otros en hidrología [7], problemas ambientales [3] y minería [14]. El uso de los métodos geoestadísticos ha crecido considerablemente y se ha implementado con datos reales univariados y multivariados [48]. Un análisis geoestadístico se realiza en dos etapas. Inicialmente se lleva a cabo el estudio de la estructura de autocorrelación espacial empleando la función de covariograma y posteriormente se hace predicción espacial usando técnicas kriging [50].

La modelación de datos circulares se refiere a la representación y análisis de datos que tienen una estructura cíclica, como datos de tiempo, direcciones, coordenadas geográficas, horas del día o ángulos. Estos se pueden encontrar en una variedad de campos, como astronomía, geografía, meteorología, ingeniería o biología, entre otros. En este contexto suelen citarse trabajos pioneros de von Mises [16] quien propuso la distribución normal circular y Mardia [28] que estableció algunas técnicas inferenciales apropiadas para su estudio. La modelación de datos circulares implica considerar la naturaleza cíclica de los datos y utilizar técnicas y herramientas estadísticas adecuadas para su análisis y visualización.

El análisis de datos circulares ha cobrado también relevancia dentro de la modelación espacial [38]. En este ámbito hay aplicaciones en geología (orientación de las fallas), geofísica (dirección del campo magnético), meteorología

(dirección del viento), biología (dirección de la migración de especies), oceanografía (corrientes oceánicas), entre otras. Se han hecho adaptaciones que permiten aplicar técnicas geoestadísticas en este contexto. Tanto la estimación de la estructura de dependencia espacial como la predicción kriging han sido consideradas bajo este marco de trabajo. Morphet [33] presenta metodologías de visualización, predicción kriging y simulación de campos aleatorios circulares estacionarios. La estructura de correlación y consecuentemente la estimación de los pesos del kriging se obtienen a partir del ajuste de modelos teóricos de covariograma a los valores estimados de la estructura de correlación espacial circular hallados con la función de cosenograma. Surjotedjo [47] propone el kriging de datos circulares precedido por pulimento de medianas como método para estimar la tendencia espacial. En [5] se proponen modelos espacio-temporales. El relativamente reciente uso de la ciencia de datos y particularmente de técnicas de aprendizaje automático hace factible emplear otros procedimientos para eliminar la tendencia espacial antes de la predicción kriging. En este trabajo, se explora el uso de redes neuronales en combinación con el kriging de datos circulares como método de predicción espacial de procesos espaciales circulares no estacionarios.

El documento está organizado en seis secciones: Una breve introducción en la sección uno permite describir el alcance del trabajo y los temas tratados en cada capítulo. En el capítulo dos se revisan conceptos básicos de geoestadística con datos de variables reales incluyendo la definición de la función de covariograma y su estimación, la predicción por kriging ordinario para campos estacionarios y kriging residual cuando se asume no estacionariedad y que la tendencia es ajustada por medio de redes neuronales. Posteriormente se aborda el tema de datos circulares. Se incluye su estudio descriptivo, la distribución von Mises [28]), el análisis geoestadístico con observaciones de procesos aleatorios circulares estacionarios y no estacionarios (cuando la tendencia espacial de los datos circulares es modelada con pulimento de medianas). La parte central del trabajo corresponde al capítulo cuatro donde se propone el uso de redes neuronales para datos circulares en combinación con kriging ordinario circular como método de predicción espacial de datos

circulares no estacionarios. En el capítulo cinco se implementa una evaluación de la propuesta con datos simulados. En la sección seis se hace un estudio de datos reales correspondientes a direcciones del viento registradas en la costa occidental de Colombia. El documento finaliza con conclusiones y opciones de trabajo futuro en el contexto de geoestadística circular.

2 Marco Teórico

El uso de geoestadística en el análisis de datos circulares es distinto al tradicionalmente empleado con realizaciones de campos aleatorios reales. En este escenario, hay particularidades que deben ser abordadas de manera diferencial. En este capítulo, inicialmente se hace una revisión de las metodologías propuestas para llevar a cabo predicción espacial de datos reales. Se muestra la teoría básica del kriging en el estudio de campos aleatorios reales bajo estacionariedad y para el caso de no estacionariedad la aproximación empleando kriging residual cuando la tendencia se estima a partir de redes neuronales. Después se presenta un resumen de las técnicas de descripción de datos angulares y un estado del arte en la modelación geoestadística de datos circulares estacionarios y no estacionarios (empleando respectivamente kriging circular y kriging circular residual con pulimento de medianas).

2.1. Geoestadística con datos de variables definidas en los reales

La geostadística es un campo de la estadística espacial que permite modelar realizaciones de campos aleatorios georreferenciados cuando hay continuidad espacial [18]. Sus principales objetivos son la estimación (de la estructura de correlación espacial), predicción (de valores de la variable en sitios sin información) y simulación de estos fenómenos. A través del modelado de la correlación espacial, la geostadística permite realizar, a través de las técnicas kriging [39], predicción insesgada y de mínima varianza (MPLI) de la variable de interés en sitios sin información. Utilizando la información de sitios vecinos y la estimación de la estructura de correlación espacial del proceso estocástico

subyacente, se realiza la predicción espacial y se estima la incertidumbre en la predicción.
Es importante tener en cuenta que la geostadística es una disciplina dentro del análisis de datos espaciales que se enfoca en ubicaciones que provienen de un conjunto continuo y son seleccionadas a discreción del investigador. Esto implica que los datos utilizados no necesariamente se colectan bajo un muestreo regular, sino que se toman en puntos específicos según las necesidades y los objetivos del estudio. En geoestadística se realiza estimación de la estructura de autocorrelación espacial, predicción en sitios sin información y simulación del fenómeno de estudio. Esto último permite generar múltiples realizaciones del proceso que facilitan, entre otras cosas, la toma de decisiones, la gestión de recursos naturales, la planificación urbana y la evaluación de riesgos [9].

2.1.1. Geoestadística bajo estacionariedad

Un campo aleatorio se define como una colección de variables aleatorias $Y(s)$, donde s denota una posición en un espacio bidimensional S. Este tiene dos componentes: una no aleatoria $\mu(s)$ que corresponde a la tendencia sistemática en s y otra aleatoria $\epsilon(s)$ que representa las fluctuaciones locales. Sea S un espacio bidimensional con $S \subset R^2$ y $s \in S$ una posición dentro del espacio S. Se define $Y(s)$ una variable aleatoria en s como:

$$Y(s) = \mu(s) + \epsilon(s)$$

El campo aleatorio se concibe como el conjunto $CA = \{Y(s), s \in S\}$. Sea $s_1, \ldots, s_n \in S$ y $(Y(s_1), \ldots, Y(s_n))$ un vector aleatorio con media $E[Y(s)] = \mu(s) = \mu$ y función de covarianza $\text{Cov}(Y(s_i), Y(s_j))$, denotado como $\text{C}(h)$, $h = ||s_i - s_j||$. Bajo estas condiciones se dice que $Y(s)$ es estacionario. Se asume que $\epsilon(s)$ es una variable aleatoria con $E[\epsilon(s)] = 0$ y covarianza $C(h)$ definida como

$$C(h) = E\left(\epsilon(s)\epsilon(s+h)\right),$$

donde $\epsilon(s)$ y $\epsilon(s+h)$ denotan las variables aleatorias en las posiciones s y $(s+h)$ respectivamente. El supuesto de Gaussianidad y estacionariedad implica que el vector $\mathbf{Y(s)} = (Y(s_1), \ldots, Y(s_n))^T \sim N_p(\boldsymbol{\mu}, \Sigma)$, con $\Sigma_{n\times n}$ la matriz de varianzas y covarianzas con términos $C(Y(s_i), Y(s_j)) = \sigma_{ij}$, $i, j = 1 \ldots n$. Lo anterior implica que la combinación lineal de las variables del proceso es Gaussiana y que la distribución conjunta de cualquier subconjunto de variables aleatorias del campo es también Gaussiana. Una propiedad clave en geoestadística es la ergodicidad, que supone que las propiedades estadísticas del proceso se pueden estimar a partir de una única realización.

2.1.2. Simulación en geoestadística

La simulación de procesos estocásticos georreferenciados juega un papel importante en el análisis de datos espaciales porque permite visualizar las características (tendencia y variabilidad) del proceso subyacente. Un enfoque ampliamente utilizado para este propósito es el de la transformación espectral, que se basa en la descomposición en términos de funciones base como las sinusoidales o de Fourier. Este método permite obtener realizaciones del campo aleatorio en el espacio mediante la generación de coeficientes aleatorios en el dominio de la frecuencia. El proceso comienza con la selección de una función de autocorrelación que describe la estructura espacial del campo aleatorio deseado. A través de esta función, se determinan los coeficientes espectrales que posteriormente son utilizados para generar los componentes de frecuencia del campo. Al aplicar la transformación inversa de Fourier (o su equivalente en funciones de base elegidas), se obtiene una realización del campo aleatorio en el dominio espacial. El uso de técnicas de Monte Carlo es otra estrategia ampliamente empleada para la simulación de campos aleatorios. En este caso, se generan valores aleatorios de una distribución específica asociada al fenómeno bajo estudio. Estos permiten visualizar la tendencia y la variabilidad y evaluar procedimientos inferenciales.

2.1.3. Aplicaciones

Realizaciones de campos aleatorios en el contexto geoestadístico surgen en diversas disciplinas científicas y áreas de investigación. Su modelación permite caracterizar las estructuras de dependencia inherentes, predecir la(s) variables(s) de interés en sitios sin información y obtener simulaciones. Por ejemplo en meteorología se pueden modelar y predecir variables como temperatura, precipitación y velocidad del viento. Esto ayuda en la comprensión de los patrones de variabilidad y hace posible pronosticar patrones climáticos en diferentes escalas espaciales. En geología, se emplean para evaluar porosidad y permeabilidad, que son críticas para la exploración y producción de recursos naturales como petróleo y gas. En ecología, se aplican para modelar la distribución de especies, analizar la dinámica de las poblaciones y estudiar las interacciones al interior de los ecosistemas.

2.2. Funciones para caracterizar la dependencia espacial con datos reales

La dependencia espacial en geoestadística se establece a través de los modelos de covariograma o equivalentemente de la función de semivariograma. En esta sección se presentan las expresiones de los modelos teóricos de covariograma y posteriormente se explica como hacer la estimación de la dependencia espacial a partir de estos mismos.

2.2.1. Modelos teóricos de covariograma

La función de covariograma permite establecer la dependencia en términos del rezago espacial h (distancia entre sitios). Si $Y(s) \sim N(\boldsymbol{\mu}, \Sigma)$, con Σ una matriz con elementos $\sigma_{ij} = C(h), h = ||s_i - s_j||$, entonces las estimaciones de $\boldsymbol{\mu}$ y Σ permite determinar de manera completa el proceso [50]. Bajo estacionariedad

$$C(h) = E[Y(s)Y(s+h)] - E[Y(s)]^2$$

Es claro que $C(0) = \sigma^2$ y que $C(h) \to 0$ cuando $h \to \infty$. Existen varios modelos teóricos de covariograma que son comúnmente utilizados en el análisis geoestadístico. Entre otros están exponencial, esférico y Gaussiano [**?**]. Sus expresiones están dadas por [29]

- Esférico:

$$C(h) = \begin{cases} \tau^2 + \sigma^2 & h = 0 \\ \sigma^2 \left(1 - \left(\frac{3}{2}\left(\frac{h}{\phi}\right) - \frac{1}{2}\left(\frac{h}{\phi}\right)^3\right)\right) & 0 < h < a \\ 0 & h \geq a \end{cases} \tag{2-1}$$

- Exponencial:

$$C(h) = \begin{cases} \tau^2 + \sigma^2 & h = 0 \\ \sigma^2 \left(\exp\left(\frac{3h}{\phi}\right)\right) & 0 < h \end{cases} \tag{2-2}$$

- Gaussiano:

$$C(h) = \begin{cases} \tau^2 + \sigma^2 & h = 0 \\ \sigma^2 \left(1 - \exp\left(-\frac{h^2}{\phi^2}\right)\right) & 0 < h \end{cases} \tag{2-3}$$

con τ^2, σ^2 y ϕ correspondientes, respectivamente, al efecto pepita (discontinuidad en el origen), la varianza y la distancia hasta la cual hay correlación espacial. En la Figura **2-1** se presentan a manera de ilustración gráficos de las funciones de variograma y covariograma correspondientes a un modelos esférico con $\tau^2 = 2$, $\sigma^2 = 8$ y $\phi = 7$.

2.2.2. Estimación del modelo de covariograma

Dada una realización $(Y(s_1), \ldots, Y(s_n))$, $s \in \mathbb{R}^2$, de un proceso estocástico espacial estacionario $Y(s) = \mu + \epsilon(s)$, la estimación de la función de covariograma se obtiene en dos etapas:

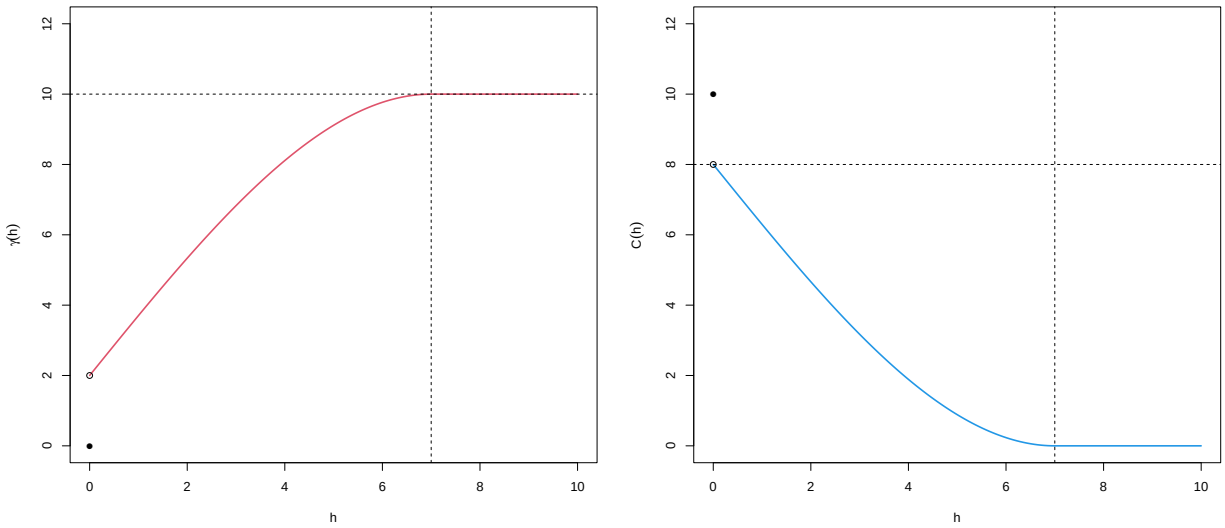

Figura 2-1: Variograma (izquierda) y covariograma (derecha) correspondientes a un modelo esférico con $\tau^2 = 2$, $\sigma^2 = 8$ y $\phi = 7$. El valor 2 en ambos representa el efecto pepita.

1. Inicialmente se calcula la función de covariograma experimental (muestral, empírica) definida como

$$\widetilde{C}(h) = \frac{\sum^{n(h)}(Y(s+h)Y(s))}{n(h)} - \tilde{\mu}^2, \tag{2-4}$$

donde $n(h)$ representa el número de parejas de sitios separados por una distancia h y $\tilde{\mu}$ la estimación de la media del proceso. En la práctica se requiere usar una tolerancia $h - \delta < h < h + \delta$ para determinar las parejas $(Y(s), Y(s+h))$ y el número de parejas $n(h)$ en cada rezago h. $Y(s)$ es el valor de la variable en un sitio s, $Y(s+h)$ es otro valor muestral separado del anterior por una distancia h. La función de covariograma experimental $\widetilde{C}(h)$ se calcula para varias distancias h. Se toman intervalos de distancia $\{[0,h], (h,2h], (2h,3h], \ldots\}$ y el valor del covariograma para un valor h corresponde a una distancia promedio entre los valores de las parejas de sitios dentro de cada intervalo y no a una distancia h específica. Obviamente el número de parejas de puntos $n(h)$ dentro de los intervalos no es constante.

2. Se ajusta alguno de los modelos teóricos de covariograma dados en la Sección 2.2.1 (esférico, exponencial, Gaussiano) a los gráficos de dispersión dados por $(h, \widetilde{C}(h))$. Para hacer lo anterior generalmente se emplea mínimos cuadrados ponderados [43], donde el peso de cada punto del dispersograma está asociado a $n(h)$ (a mayor número de observaciones más peso tiene el punto correspondiente en la estimación del modelo).

2.3. Kriging ordinario con variables definidas en los reales

Kriging es un método para predecir variables espacialmente dependientes cuando el fenómeno de interés varía de forma continua en la región estudiada [48]. Kriging funciona expresando el valor de una variable aleatoria en un punto dado como una combinación lineal de los valores de la variable aleatoria en otros puntos cercanos [35]. Los coeficientes de la combinación lineal se determinan de modo que se minimice la suma de los cuadrados de los errores de predicción [9]. Kriging se ha utilizado con éxito en una amplia variedad de contextos aplicados, incluyendo minería, geología, meteorología, bioestadística e ingeniería.

El objetivo del kriging ordinario es predecir el valor de una variable $Y(s_0)$ en el punto s_0 sin observación, a partir de un conjunto de mediciones $Y(s_i), i = 1, \ldots, n$ correlacionadas espacialmente. El predictor kriging ordinario en el sitio (s_0) dada la realización $Y(s_1), \ldots, Y(s_n)$ se define como la combinación lineal de n variables aleatorias

$$\widehat{Y}(s_0) = \sum_{i=1}^{n} \lambda_i Y(s_i), \tag{2-5}$$

donde $\lambda_i, i = 1, \ldots, n$ representa el peso de $Y(s_i)$ en la predicción. El predictor en (2-5) es insesgado si $\sum_{i=0}^{n} \lambda_i = 1$. Los $\lambda_i, i = 1, \ldots, n$ se escogen de tal forma que se tenga la mejor predicción lineal insesgada (MPLI), es decir se calculan los pesos que minimizan la varianza de predicción sujetos a

la restricción de insesgadez. Formalmente se tiene el siguiente problema de optimización:

$$\text{mín}\, V[\widehat{Y}(s_0) - Y(s_0)] \quad \text{s.a.} \quad \sum_{i=0}^{n} \lambda_i = 1 \tag{2-6}$$

Se puede comprobar [7] que los λ's se obtienen resolviendo el sistema matricial

$$\begin{pmatrix} C_{11} & \dots & C_{1n} & 1 \\ C_{21} & \dots & C_{2n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \nu \end{pmatrix} \begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{pmatrix}, \tag{2-7}$$

donde $C_{ij}, i, j = 1, \dots, n$ corresponde a la función de covarianza $C(h), h = ||s_i - s_j||$ que es estimada ajustando (por mínimos cuadrados o máxima verosimilitud) algún modelo de covarianza teórico (Sección 2.2.1) al covariograma experimental (Sección 2.2.2). El parámetro ν corresponde al multiplicador de Lagrange asociado a la condición de insesgadez.

2.4. Geoestadística no estacionaria

Un concepto fundamental en estadística espacial es el de estacionariedad, que indica que las propiedades (media y varianza) del proceso son constantes y que la estructura de autocorrelación solo depende del vector de separación y no de la dirección. En la vida real, este supuesto puede ser simplista, particularmente porque la media del fenómeno no es constante. En este contexto, se habla de geoestadística no estacionaria. Tres aspectos clave en la no estacionariedad son anisotropía (la correlación cambia a lo largo del espacio), heterogeneidad espacial (la variabilidad depende de la posición) y tendencia (la media del proceso no es constante). Las soluciones a estos problemas se fundamentan en el uso de modelos de covarianza anistrópicos, transformaciones y estimaciones de la tendencia con métodos de regresión, respectivamente.

La estadística espacial no estacionaria tiene una amplia variedad de aplicaciones en diversos campos, incluyendo ciencias ambientales [34], epidemiología [6] y minería [25]. En áreas de estudio como contaminación atmosférica, calidad del agua, evaluación de contenidos minerales y combustibles fósiles, entre otros, es común detectar ausencia de estacionariedad. Esto se debe principalmente a que la media del proceso sigue tendencias, lineales o no lineales, en la región de interés. Aunque la estadística espacial no estacionaria conlleva a una mayor precisión en numerosos escenarios prácticos, también presenta desafíos considerables. Los modelos no estacionarios tienden a ser intrincados y demandan mayores recursos computacionales tanto para su implementación como para su ajuste, en comparación con los modelos estacionarios. La elección de un modelo adecuado para capturar la no estacionariedad puede resultar un reto, ya que se busca lograr un equilibrio entre la flexibilidad y la simplicidad del modelo. En ocasiones, la disponibilidad de datos georreferenciados puede ser limitada, lo que complica la realización de un análisis completo de la no estacionariedad. En tales situaciones, puede ser necesario recurrir a técnicas suplementarias de interpolación o estimación para abordar estas limitaciones y obtener resultados robustos.

2.5. Geoestadística no estacionaria basada en redes neuronales

La combinación de redes neuronales y geoestadística ha demostrado ser útil en diversas áreas [1]. Las redes neuronales son modelos computacionales que se inspiran en la estructura y funcionamiento del cerebro humano. Están compuestas por neuronas artificiales interconectadas organizadas en capas, lo que les permite aprender patrones y relaciones complejas en los datos mediante un proceso de entrenamiento. Esta capacidad las hace idóneas para modelar y analizar fenómenos espaciales. La aplicación de redes neuronales en geoestadística ofrece varias ventajas, entre ellas su capacidad para capturar relaciones no lineales y modelar procesos complejos.

Las redes neuronales hacen factible manejar datos de alta dimensión. Entre otras áreas se han empleado para el tratamiento de imágenes de satélite y patrones puntuales espaciales. Estas pueden considerar de manera explícita la dependencia espacial [8]. La adaptación de las redes neuronales al contexto espacial y particularmente en geoestadística incluye además de la predicción la detección de anomalías.

El uso de redes neuronales en estadística espacial presenta desafíos. La interpretabilidad del modelo, la heterogeneidad de los datos y la complejidad computacional son consideraciones importantes que deben ser abordadas para garantizar resultados confiables. El desarrollo de técnicas para mejorar la interpretabilidad del modelo y la creación de arquitecturas especializadas para el análisis de datos espaciales son campos de interés en la investigación en redes neuronales[24]. Numerosos estudios han respaldado la eficacia de las redes neuronales en estadística espacial. Por ejemplo Zhang et al. [52] empleó redes neuronales convolucionales para llevar a cabo la clasificación de cobertura terrestre utilizando imágenes de teledetección. En el ámbito de geoestadística no estacionaria, el método de kriging residual basado en redes neuronales se fundamenta en la estimación de la tendencia espacial a través del uso de redes neuronales artificiales seguido por la aplicación del kriging ordinario sobre los residuos correspondientes ([12], [44], [53]). Las redes se componen de varias capas (una de entrada, una o varias ocultas y una de salida). Cada capa consta de un conjunto de nodos que tienen entradas de las capas anteriores y salidas a las capas siguientes. Es importante considerar que cada capa utiliza una función de activación para transformar las entradas y generar las salidas

Seleccionar el número adecuado de neuronas en la capa oculta es esencial, puesto que que tener muy pocas privaría a la red de los recursos necesarios para resolver el problema, mientras que tener demasiadas aumentaría el tiempo de entrenamiento y podría llevar al sobre-entrenamiento (la red aprendería aspectos insignificantes del conjunto de entrenamiento que no son relevantes para datos más generales). Es esencial investigar cómo se comportan las predicciones y los residuos al variar el número de capas ocultas y neuronas, a fin

de determinar la configuración óptima para el modelo.
Dentro del contexto de las redes neuronales artificiales, se llevan a cabo pruebas con el propósito de evaluar su capacidad para generalizar y responder adecuadamente en la predicción. Durante la fase de operación, la red neuronal artificial recibe coordenadas en una cuadrícula regular y genera predicciones para las salidas correspondientes. Después de completar el proceso de aprendizaje, se realiza un análisis de los residuos utilizando técnicas de variografía. Los residuos son estacionarios, lo que permite estimar la función de de covariograma ajustando un modelo teórico de covariogrma (Sección 2.2.1) al covariograma experimental calculado con los residuales. Para generar mapas de predicción se emplea kriging ordinario (sección 2.3) de los residuales en combinación con la tendencia estimada con la red neuronal. Finalmente, se realiza una evaluación del método haciendo validación cruzada (dejando una o varios observaciones fuera de la muestra de entrenamiento y haciendo predicción de ellas).

2.6. Datos circulares

Los datos circulares emergen en situaciones en las que las mediciones siguen un ciclo natural o periódico, como las direcciones del viento o las horas del día. En contraposición a la habitual escala real, los datos circulares se expresan en grados o radianes a lo largo de un círculo completo (360° o 2π radianes). Los datos circulares son aquellos que pueden representarse como puntos sobre un círculo unitario. Estos se utilizan comúnmente en matemáticas, física, geografía, navegación y muchas otras disciplinas para describir la orientación y la dirección de objetos o puntos en el espacio. A manera de ilustración, en la Figura 2.1 se muestra una representación de 100 datos simulados de dirección del viento. Dado que un círculo es una curva cerrada, resulta evidente que la teoría estadística convencional, diseñada para el análisis de datos en la línea real, no es adecuada para los datos recolectados en un círculo unitario. En esta situación, se hace necesario definir medidas descriptivas, funciones de densidad y en general técnicas estadísticas que consideren la naturaleza

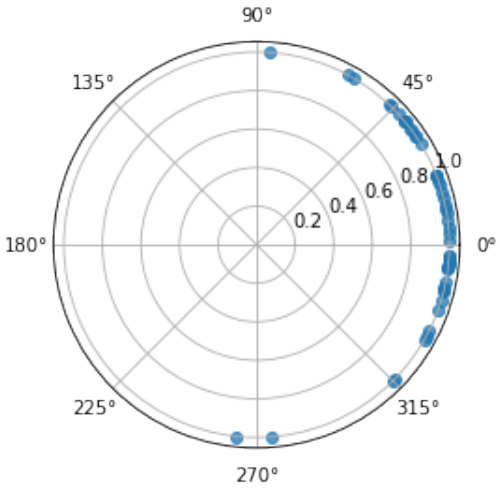

Figura 2-2: Representación gráfica de 100 observaciones de dirección del viento generadas aleatoriamente

periódica propia del círculo unitario. En el contexto probabilístico se han desarrollado funciones de distribución que tienen en cuenta las estructuras algebraicas del círculo unitario, como entre otras la von Mises, introducida en 1905 por Langevin [26]. A continuación se presenta de manera resumida una revisón de esta misma.

2.6.1. Distribución von Mises

Esta sección está basada en [28]. La distribución von Mises juega un papel fundamental en inferencia estadística con datos circulares. Su importancia en este contexto es equivalente a la que tiene la densidad normal en el caso de datos reales. Una variable aleatoria circular θ sigue una distribución von Mises, denotado como $\theta \sim$ von Mises (θ_0, κ), si su función de densidad de probabilidad está dada por

$$f(\theta; \theta_0, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \theta_0)},\ 0 < \theta \leq 2\pi,\ ,0 \leq \theta_0 < 2\pi,\ \kappa > 0, \qquad (2\text{-}8)$$

donde $I_0(\kappa)$ es la función de Bessel de primer tipo y orden cero definida como [51]

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{1}{2}\kappa\right)^{2r} \tag{2-9}$$

El parámetro κ está asociado a la concentración (variabilidad). A mayor κ mayor agrupación alrededor de la dirección media. En la Figura 2-2 se presentan densidades von Mises con distintos valores de κ. Puede observarse que a mayor κ menor varianza.

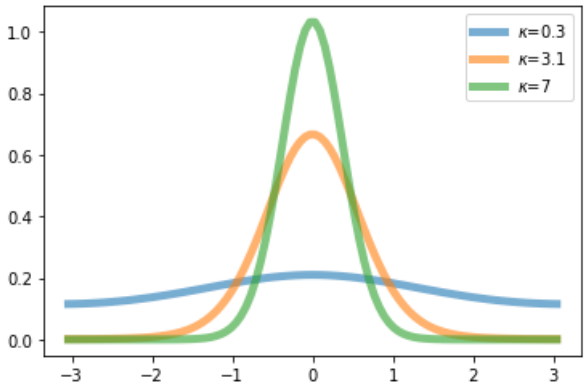

Figura 2-3: Función de densidad von Mises para tres valores del parámetro de concentración κ.

Si $\theta \sim$ von Mises (θ_0, κ) se tiene que

- $E(\theta) = \theta_0$
- $V(\theta) = [1 - E(\cos(\theta))]$,

con

$$E[\cos(\theta)] = \frac{\int_0^{2\pi} \frac{\cos(\theta)}{2\pi} e^{\kappa \cos(\theta - \theta_0)}}{I_0(\kappa)}$$
$$= A(\kappa)$$

Los conceptos anteriores juegan un rol esencial en el estudio de datos circulares puesto que permiten cuantificar la localización y dispersión de los mismos.

2.6.2. Medidas descriptivas de datos angulares

Las medidas descriptivas para datos circulares difieren de las de datos reales debido al comportamiento modular inherente de las observaciones. Cuando se dispone de una muestra aleatoria angular $\theta_1, \ldots, \theta_n$ de una función de distribución $F_\Theta(\theta)$, cada ángulo θ_i con $i = 1, ..., n$ tiene asociado un vector $z_i = (\cos(\theta_i), \text{seno}(\theta_i))$ en el círculo unitario. La media muestral $\bar{\theta}$ está relacionada a la dirección promedio de los vectores z_i. Esta se obtiene a partir del centro de gravedad (C, S) definido a partir de los promedios como:

$$C = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i)$$
$$S = \frac{1}{n}\sum_{i=1}^{n}\text{seno}(\theta_i)$$

La media de datos circulares se define como [28]

$$\bar{\theta} = \begin{cases} \arctan(\frac{S}{C}) & C > 0 \\ \arctan(\frac{S}{C}) + \pi & S \geq 0, C < 0 \\ \arctan(\frac{S}{C}) - \pi & S < 0, C < 0 \\ \frac{\pi}{2} & S > 0, C = 0 \\ -\frac{\pi}{2} & S < 0, C = 0 \\ \text{indefinido} & S = 0, C = 0 \end{cases} \tag{2-10}$$

Sea $\bar{R} = \sqrt{C^2 + S^2}$. C y S se obtienen a partir de $\bar{R}$ y $\bar{\theta}$ como $C = \bar{R}\cos(\bar{\theta})$ y $S = \bar{R}\text{seno}(\bar{\theta})$.
La longitud media resultante $\bar{R}$ es la medida más común de concentración de datos circulares [37]. $\bar{R}$ es igual a 1 cuando todos los datos están agrupados

en un solo punto del círculo unitario, lo que indica una alta concentración alrededor de $\bar{\theta}$. Si los datos están dispersos de manera uniforme en el círculo, $\bar{R}$ tendrá un valor cercano a cero. Lo contrario no es cierto, es decir un valor de 0 en $\bar{R}$ no necesariamente significa que los datos están uniformemente distribuidos en el círculo. Dada una muestra aleatoria $\theta_1, \ldots, \theta_n$ de tamaño n se define la dispersión alrededor de α y la varianza (tomando $\alpha = \bar{\theta}$) como

$$D = \frac{1}{n}\sum_{i=1}^{n}(1 - \cos(\theta_i - \alpha))$$

$$V = 1 - \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i - \bar{\theta}) = 1 - \bar{R}$$

A continuación se presenta un ejemplo par ilustrar lo anterior: Suponga que se tienen 3 muestras tamaño 100 de variables aleatorias circulares θ_1, θ_2 y θ_3 con valores de concentración $\bar{R}_1$=0.16, $\bar{R}_2$=0.78 y $\bar{R}_3$ =0.994, respectivamente (Figura **2-4**). Puede observarse que a mayor $\bar{R}$ mayor es la concentración de puntos sobre la media.

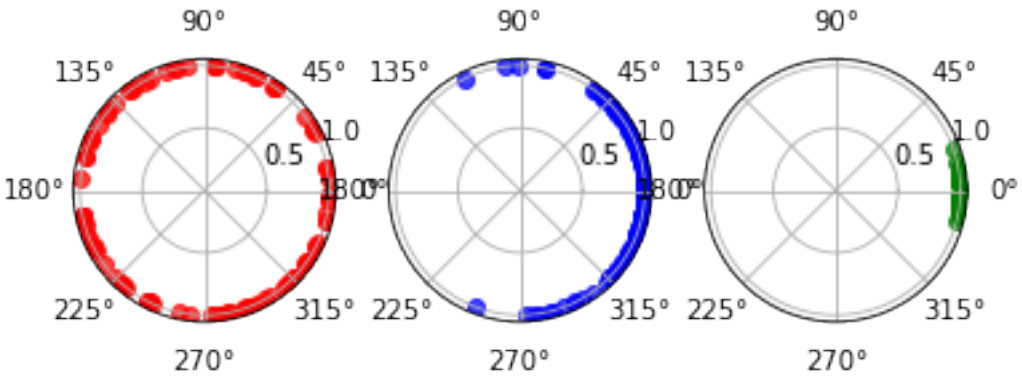

Figura 2-4: Tres muestras tamaño 100 de variables aleatorias θ_1, θ_2 y θ_3 distribuidas con distribución von Mises y $\bar{R}_1$=0.16 (izquierda), $\bar{R}_2$=0.78 (centro) y $\bar{R}_3$ =0.994 (derecha).

2.7. Geoestadística con datos circulares

Las técnicas geoestadísticas generalmente se aplican a datos continuos medidos lo largo de una escala lineal. Estas han sido adaptadas al contexto angular (dirección del viento, dirección de las olas, rumbos de la brújula, horas de día). A continuación se revisan las particularidades de la geoestadística en el contexto de datos circulares. Se presentan los conceptos de covariaograma y kriging bajo estacionariedad y también se revisa una aproximación no estacionaria basada en pulimento de medianas.

2.7.1. Modelos teóricos de covariograma circular

Un campo aleatorio circular se define como el conjunto $\{\Theta(s), s \in R^2\}$. Bajo estacionariedad la realización $\theta(\mathbf{s}) = (\theta(s_1), \ldots, \theta(s_n))^T$ tiene función media $E(\theta(s_j)) = \theta$, $s_j \in D \subset R^2$ y función de covariograma $C(h) = \text{Cov}(\theta(s_i), \theta(s_j))$, $h = ||s_i - s_j||$. Específicamente

$$\begin{aligned} C(h) &= \text{Cov}(\theta(s), \theta(s+h)) \\ &= E[\theta(s)\theta(s+h)] - E[\theta(s)]^2 \end{aligned}$$

Algunos de modelos teóricos de covariograma en el ámbito de datos circulares son los siguientes [33]:
Esférico

$$C(h) = \begin{cases} 1 & h = 0 \\ 1 - \tau^2 - (1 - \tau^2 - \rho^2)\exp\left(\frac{3}{2}\left(\frac{h}{\phi}\right) - \frac{1}{2}\left(\frac{h}{\phi}\right)^3\right) & 0 < h < \phi \\ \rho^2 & h > \phi \end{cases}$$

Exponencial

$$C(h) = \begin{cases} 1 & h = 0 \\ \rho^2 + (1 - \tau^2 - \rho^2)\exp\left(-3\left(\frac{h}{\phi}\right)\right) & h > 0 \end{cases}$$

Gaussiano

$$C(h) = \begin{cases} 1 & h = 0 \\ \rho^2 + (1 - \tau^2 - \rho^2)\exp\left(-3\left(\frac{h}{\phi}\right)^2\right) & h > 0 \end{cases}$$

donde ρ^2 es la meseta, τ^2 es la discontinuidad en el origen (efecto pepita) y ϕ es el rango (máxima distancia de correlación). De manera análoga a la función de covarianza en el caso de datos reales se tiene que $C(h) \to \rho^2$ cuando $h \to \infty$. En algunas publicaciones se emplea $\zeta(h)$ en vez de $C(h)$. Por concordancia con las expresiones usadas en la sección 2.2.1 para definir el covariograma de datos reales, denotamos a la función de covariograma circular como $C(h)$. A manera de ilustración en la Figura **2-5** se presenta una función de cosenograma experimental (puntos azules) y el ajuste de un modelo teórico de covariograma con $\tau^2 = 0$, ρ^2 =0.63 y ϕ =1.12.

2.7.2. Función cosenograma

A la función de covariograma experimental en el contexto de datos circulares se le denomina función de cosenograma y es usualmente denotada como $\zeta(d)$ [33]. En el trabajo se emplea $\widehat{C}(h)$ por homogeneidad con la notación empleada para el caso de datos reales. La estimación de la función de covariograma dada la muestra $\theta_i, i = 1, \ldots, n$ del campo circular se obtiene ajustando alguno de los modelos teóricos de covariograma dados en la sección 2.7.1 (esférico, exponencial, Gaussiano) al cosenograma definido como [33]:

$$\widetilde{C}(h) = \frac{1}{n(h)} \sum_{||x_i - x_j| - h| < \delta} \cos(\theta_i - \theta_j), \tag{2-11}$$

donde $n(h)$ es el número de pares de observaciones circulares separadas por una distancia h. Como en el caso de datos reales en la práctica se considera una tolerancia $h - \delta < h < h + \delta$ para definir $n(h)$. A manera de ilustración, en la Figura **2-5** se muestra el ajuste de un modelo esférico de covariograma

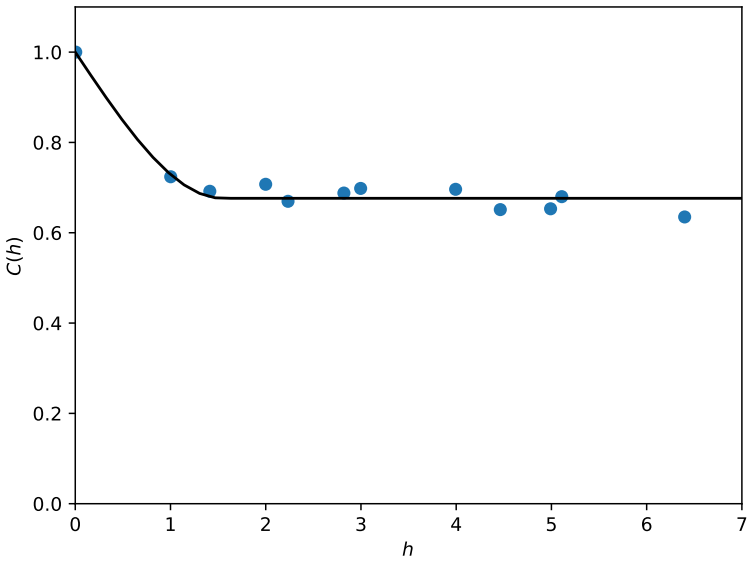

Figura 2-5: Ejemplo del cosenograma experimental (puntos) correspondiente a datos simulados a partir de una distribución von Mises con parámetro κ=3.1 (ver sección 2.7.3) y ajuste por mínimos cuadrados ordinarios de un modelo esférico de covariograma circular ($\tau^2 = 0$, ρ^2=0.63, ϕ=1.12).

circular (curva negra) al cosenograma obtenido con un conjunto de datos circulares (puntos azules) de un proceso circular simulado.

2.7.3. Simulación de campos aleatorios circulares

Para simular un campo aleatorio circular se puede seguir el procedimiento propuesto en [33] que está basado en la generación inicial de valores de un campo aleatorio Gaussiano. Las etapas son:

1. Genere una simulación de un campo aleatorio Gaussiano $\left\{Z(s), s \in R^2\right\}$ asumiendo $\mu = 0, \sigma^2 = 1$ y un modelo de covarianza dado (esférico, Gaussiano, exponencial, etc) obteniendo así el vector $(z(s_1), \ldots, z(s_n))$. En [10] se dan los pasos para llevar a cabo este proceso.

2. Para cada observación $z(s_i)$ generada en el punto anterior obtenga $\theta(s_i) = F_{\Theta}^{-1}(F(z_i))$, donde F_{Θ} representa una función de distribución circular (von Mises, uniforme, triangular, cardioide, etc) y $F(z_i)$ es la función de probabilidad acumulada en cada valor de la simulación del campo aleatorio Gaussiano. Es relevante resaltar que al generar este campo se debe considerar que el valor cero depende de la dirección que se elija como referencia. En el caso de $\theta \sim$ von Mises(θ_0, κ)

$$F_{\Theta}(\theta) = \int_{-\pi}^{\theta} \frac{\exp\left(\kappa \cos(\phi)\right)}{2\pi \sum_{j=0}^{\infty} \left(\frac{\kappa}{2}\right)^2 \left(\frac{1}{j!}\right)^2} d\phi \tag{2-12}$$

   Otras distribuciones circulares se pueden consultar en la tabla 5.1 de [33]

2.7.4. Kriging ordinario con datos circulares

El kriging circular ordinario se obtiene a partir de una combinación lineal de observaciones de datos angulares, asumiendo que se tiene una realización de un proceso estocástico con media y varianza común, donde la función

de covarianza es obtenida por el ajuste de un modelo teórico de covariograma circular a una estimación de la función de cosenograma definida como $\mathbb{E}[\cos(\theta(s_i) - \theta(s_j))]$ con $h = ||s_i - s_j||$. Sea $\theta(s_0)$ la variable en la posición s_0, que tiene una representación vectorial definida como

$$u(s_0) = \begin{bmatrix} \cos(\theta(s_0)) \\ \text{seno}(\theta(s_0)) \end{bmatrix} \tag{2-13}$$

En kriging circular se hacen predicciones sobre cada una de las componentes de $u(s_0)$. De esta manera se puede caracterizar el ángulo en un contexto conocido. La predicción de $u(s_0)$ en la posición s_0 sin medición se halla a partir de una combinación lineal de las mediciones $u(s_i)$, $i = 1, ..., n$, donde los pesos se determinan teniendo en cuenta la correlación espacial. Los vectores $u(s_i)$ están correlacionados espacialmente con $u(s_0)$, lo que implica que los vectores $u(s_i)$ más cercanos a $u(s_0)$ tienen mayor ponderación que los distantes. Específicamente la predicción se define como

$$\widehat{u}(s_0) = \sum_{i=1}^{n} u(s_i) w_i, \ w_i \in \mathbb{R} \tag{2-14}$$

$$\begin{bmatrix} \widehat{\cos(\theta(s_0))} \\ \widehat{\text{seno}(\theta(s_0))} \end{bmatrix} = \begin{bmatrix} \cos(\theta(s_1)) \\ \text{seno}(\theta(s_1)) \end{bmatrix} w_1 + \ldots + \begin{bmatrix} \cos(\theta(s_n)) \\ \text{seno}(\theta(s_n)) \end{bmatrix} w_n$$

La representación anterior es análoga a la definida en el contexto de kriging ordinario de datos reales. Los vectores $u(s_i)$ se asumen unitarios, es decir se tiene

$$\begin{aligned} (\widehat{u}(s_0))^T \widehat{u}(s_0) &= 1 \\ W^T U U^T W &= 1 \\ W^T K W &= 1, \end{aligned}$$

donde

$$K = \begin{bmatrix} 1 & (u(s_1))^T u(s_2) & \ldots & (u(s_1))^T u(s_n) \\ (u(s_2))^T u(s_1) & 1 & \ldots & (u(s_2))^T u(s_n) \\ \vdots & \vdots & \ddots & \vdots \\ (u(s_n))^T u(s_1) & (u(s_n))^T u(s_2) & \cdots & 1 \end{bmatrix} \tag{2-15}$$

Los pesos w_i se obtienen minimizando la varianza del error definida como

$$E[1 - \cos(\hat{\theta}(s_0) - \theta(s_0))] \tag{2-16}$$

Se tiene que

$$\cos[\widehat{\theta}(s_0) - \theta(s_0)] = \cos(\widehat{\theta}(s_0))\cos(\theta(s_0)) + \text{seno}(\widehat{\theta}(s_0))\text{seno}(\theta(s_0))$$

$$= \frac{\sum_{i=1}^{n} w_i \left[\cos(\theta(s_i))\cos(\theta(s_0)) + \text{seno}(\theta(s_i))\text{seno}(\theta(s_0))\right]}{\sqrt{W^T K W}}$$

$$= \frac{\sum_{i=1}^{n} w_i \cos[\theta(s_i) - \theta(s_0)]}{\sqrt{W^T K W}}$$

$$= \frac{W^T C}{W^T K W}$$

El problema de minimización en (2-16) es equivalente al de maximización de

$$E\left[\frac{W^T C}{\sqrt{W^T K W}}\right], \text{ con } C^T = [\cos(\theta(s_1) - \theta(s_0)), \ldots, \cos(\theta(s_n) - \theta(s_0))]$$

Usando la expansión de Taylor de primer orden se tiene

$$E\left[\frac{W^T C}{\sqrt{W^T K W}}\right] = \frac{W^T \widetilde{C}}{\sqrt{W^T \widetilde{K} W}}$$

En la práctica $\widetilde{K}$ y $\widetilde{C}$ se calculan ajustando un modelo de covariograma circular teórico al cosenograma experimental. Usando el método de los multiplicadores de Lagrange se obtiene

$$W = \frac{\widetilde{K}^{-1}\widetilde{C}}{\sqrt{\widetilde{C}^T \widetilde{K}^{-1} \widetilde{C}}} \tag{2-17}$$

Note que para hallar W no es necesario calcular el denominador ya que este desaparece cuando se obtiene la función arcotangente de las componentes estimadas. Una vez hallada la predicción $\widehat{u}(s_0)$ se obtiene la predicción angular

como

$$\widehat{\theta}(s_i) = \arctan\left(\frac{\widehat{\text{seno}(\theta(s_i))}}{\widehat{\cos(\theta(s_i))}}\right) \tag{2-18}$$

La varianza de predicción del kriging circular y su correspondiente estimación están dadas por [33]

$$\sigma^2_{\text{ck}} = 2 - 2\sqrt{C^T K^{-1} C} \tag{2-19}$$

y

$$\widehat{\sigma}^2_{\text{ck}} = 2 - 2\sqrt{\widetilde{C}^T \widetilde{K}^{-1} \widetilde{C}} \tag{2-20}$$

2.7.5. Kriging residual de datos circulares usando pulimento de medianas

Esta sección es basada en [47]. El kriging residual es un método de predicción empleado en geostadística cuando el campo aleatorio de interés no es estacionario [9]. Consiste en ajustar la tendencia empleando técnicas de regresión y en estimar la estructura de correlación espacial a partir de los residuales del modelo estimado. La predicción del proceso en sitios sin información se obtiene como la suma de la tendencia y la predicción del kriging residual [**?**]. Una alternativa para llevar a cabo kriging residual es estimar la tendencia empleando el pulimento de medianas [9]. Esta es una estrategia que busca mitigar los efectos de atípicos o extremos en el conjunto de residuos. Estos valores inusuales pueden distorsionar las predicciones del modelo y afectar la precisión del proceso de interpolación. Se aplica el pulimento de medianas para identificar y filtrar aquellos residuos que se desvían significativamente de la mediana de los residuos. Al suavizar los valores extremos se logra una estabilidad en el proceso de predicción y se reduce la influencia de datos anómalos en la interpolación.

La combinación de kriging residual y pulimento de medianas da como resultado una metodología robusta y confiable de interpolación espacial. Esta

técnica mejora la precisión de las predicciones y proporciona una forma de trabajar con valores atípicos que podrían afectar los resultados. Empleando pulimento de medianas en conjunción con kriging residual se puede hacer predicción espacial de un campo aleatorio no estacionario.
Sea $\theta(s_{kl}), s_{kl} \in S \subset R^2$ una realización de un campo aleatorio circular asociado a un enmallado rectangular con componentes horizontales y verticales kl, $k = 1, \ldots, p$ y $l = 1, \ldots q$. Por simplicidad en la notación a continuación se usa θ_{kl} en vez de $\theta(s_{kl})$ El algoritmo del pulimento de medianas se puede realizar en cinco etapas:

1. Defina Θ como una matriz de tamaño $(p+1) \times (q+1)$

$$\Theta_{kl} = \begin{cases} \theta_{kl} & k = 1, \ldots, p;\ l = 1, \ldots, q \\ 0 & k = (p+1) \text{ o } l = (p+1) \end{cases}$$

2. Θ_{kl} es descompuesta en dos matrices dadas por

$$H_{kl}^{(0)} = \begin{cases} \cos(\theta_{kl}) & k = 1, \ldots, p;\ l = 1, \ldots, q \\ 0 & \text{en otro caso} \end{cases}$$

y

$$V_{kl}^{(0)} = \begin{cases} \text{seno}(\theta_{kl}) & \text{con} \quad k = 1, \ldots, p;\ l = 1, \ldots, q \\ 0 & \text{en otro caso} \end{cases}$$

3. Realice hasta que haya convergencia un procedimiento iterativo para obtener las matrices $H^{(i)}$ y $V^{(i)}$, $i = 1, 3, 5, \ldots$, con

$$\begin{aligned} H_{kl}^{(i)} &= H_{kl}^{(i-1)} - \text{Mediana}\{H_{kl}^{(i-1)}, l = 1, ..., q\} \text{ si } k = 1, ..., p+1; l = 1, ..., q \\ H_{k,q+1}^{(i)} &= H_{k,q+1}^{(i-1)} + \text{Mediana}\{H_{kl}^{(i-1)}, l = 1, ..., q\} \text{ si } k = 1, ..., p+1, l = q+1 \\ V_{kl}^{(i)} &= V_{kl}^{(i-1)} - \text{Mediana}\{V_{kl}^{(i-1)}, l = 1, ..., q\} \text{ si } k = 1, ..., p+1; l = 1, ..., q \\ V_{k,q+1}^{(i)} &= V_{k,q+1}^{(i-1)} + \text{Mediana}\{V_{kl}^{(i-1)}, l = 1, ..., q\} \text{ si } k = 1, ..., p+1 \end{aligned}$$

4. Lleve a cabo un procedimiento iterativo para obtener las matrices $H^{(i)}$ y $V^{(i)}$, $i = 2, 4, 6, \ldots$, con

$$
\begin{aligned}
H_{kl}^{(i)} &= H_{kl}^{(i-1)} - \text{Mediana}\{H_{kl}^{(i-1)}, k = 1, ..., p\} \text{si} k = 1, ..., p; l = 1, ..., q+1 \\
H_{p+1,l}^{(i)} &= H_{p+1,l}^{(i-1)} + \text{Mediana}\{H_{kl}^{(i-1)}, k = 1, ..., p\} \text{ si } l = 1, \ldots, q+1 \\
V_{kl}^{(i)} &= V_{kl}^{(i-1)} - \text{Mediana}\{V_{kl}^{(i-1)}, k = 1, ..., p\} \text{ si } k = 1, ..., p; l = 1, ..., q+1 \\
V_{p+1,l}^{(i)} &= V_{p+1,l}^{(i-1)} + \text{Mediana}\{V_{kl}^{(i-1)}, k = 1, ..., p\} \text{ si } l = 1, ..., q+1
\end{aligned}
$$

5. Repita los pasos hasta obtener $V^{(\infty)}$ y $H^{(\infty)}$ que hacen referencia a la iteración en la que converge el algoritmo. Las matrices de esta etapa se denotan como

$$
\begin{aligned}
\widetilde{a}_H &= H_{p+1,q+1}^{(\infty)} \\
\widetilde{a}_V &= V_{p+1,q+1}^{(\infty)} \\
\widetilde{r}_{H,k} &= H_{k,q+1}^{(\infty)}, k = 1, ..., p \\
\widetilde{r}_{V,k} &= V_{k,q+1}^{(\infty)}, k = 1, ..., p \\
\widetilde{c}_{H,l} &= H_{p+1,l}^{(\infty)}, l = 1, ..., q \\
\widetilde{c}_{V,l} &= V_{p+1,l}^{(\infty)}, l = 1, ..., q
\end{aligned}
$$

6. Retransforme los efectos estimados (total, de fila, de columna y residual) a datos circulares usando la siguiente regla

$$
\theta_{kl}^{\infty} = \begin{cases}
\arctan\left(\frac{H_{k,l}^{(\infty)}}{V_{k,l}^{(\infty)}}\right) & H_{k,l}^{(\infty)} > 0 \\
\arctan\left(\frac{H_{k,l}^{(\infty)}}{V_{k,l}^{(\infty)}}\right) + \pi & V_{k,l}^{(\infty)} \geq 0, H_{k,l}^{(\infty)} < 0 \\
\arctan\left(\frac{H_{k,l}^{(\infty)}}{V_{k,l}^{(\infty)}}\right) - \pi & V_{k,l}^{(\infty)} < 0, H_{k,l}^{(\infty)} < 0 \\
\frac{\pi}{2} & V_{k,l}^{(\infty)} \geq 0, H_{k,l}^{(\infty)} = 0 \\
-\frac{\pi}{2} & V_{k,l}^{(\infty)} < 0, H_{k,l}^{(\infty)} = 0 \\
\text{indefinido} & V_{k,l}^{(\infty)} = 0, H_{k,l}^{(\infty)} = 0
\end{cases} \tag{2-21}
$$

7. Haga kriging ordinario para datos circulares (sección 2.7.4) usando los residuales.

   a) Obtenga el cosenograma de los residuales

   b) Seleccione un modelo válido de covariograma circular y ajústelo al cosenograma del item anterior

   c) Aplique kriging circular a los residuales empleando el modelo de covariograma ajustado para hacer predicciones de los residuales angulares en sitios sin observación s_0

8. Estime de la tendencia espacial $\left(\tilde{\theta}(s_0)\right)$ en un sitio $s_0 = (x_0, y_0)^T$ no observado como

$$\tilde{\theta}(s_0) = \tilde{a} + \tilde{r}_k + \left(\frac{y_0 - y_k}{y_{k+1} - y_k}\right)(\tilde{r}_{k+1} - \tilde{r}_k) + \tilde{c}_l + \left(\frac{x_0 - x_l}{x_{l+1} - x_l}\right)(\tilde{c}_{l+1} - \tilde{c}_l),$$

   con $x_l < x_0 < x_{l+1}$ y $y_k < y_0 < y_{k+1}$, donde $\hat{a}$ es la estimación del efecto total, $\hat{r}_k$ es el efecto de fila k, $\hat{c}_l$ es el efecto de columna l

9. La predicción kriging por pulimento de medianas en el sitio s_0 es igual a la suma de la tendencia $\tilde{\theta}(s_0)$ más la predicción del residual circular $(\hat{u}^e(s_0))$ obtenida por kriging ordinario como

$$\widehat{u}^e(s_0) = \sum_{i=1}^{n} u^e(s_i) w_i,\ w_i \in \mathbb{R} \tag{2-22}$$

$$\begin{bmatrix} \widehat{\cos(\theta^e(s_0))} \\ \widehat{\text{seno}(\theta^e(s_0))} \end{bmatrix} = \begin{bmatrix} \cos(\theta^e(s_1)) \\ \text{seno}(\theta^e(s_1)) \end{bmatrix} w_1 + \ldots + \begin{bmatrix} \cos(\theta^e(s_n)) \\ \text{seno}(\theta^e(s_n)) \end{bmatrix} w_n$$

   Los vectores $u^e(s_i)$ se asumen unitarios, luego

$$\begin{aligned} (\widehat{u}^e(s_0))^T \widehat{u}^e(s_0) &= 1 \\ W^T U^e (U^e)^T W &= 1 \\ W^T K W &= 1, \end{aligned}$$

donde

$$K = \begin{bmatrix} 1 & (u^e(s_1))^T u^e(s_2) & \dots & (u^e(s_1))^T u^e(s_n) \\ (u^e(s_2))^T u^e(s_1) & 1 & \dots & (u^e(s_2))^T u^e(s_n) \\ \vdots & \vdots & \ddots & \vdots \\ (u^e(s_n))^T u^e(s_1) & (u^e(s_n))^T u^e(s_2) & \cdots & 1 \end{bmatrix} \tag{2-23}$$

Los pesos w_i en este caso se obtienen minimizando la varianza del error definida como

$$\mathbb{E}[1 - \cos(\hat{\theta}^e(s_0) - \theta^e(s_0))] \tag{2-24}$$

Se tiene que

$$\cos[\widehat{\theta^e}(s_0) - \theta(s_0)] = \cos(\widehat{\theta^e}(s_0))\cos(\theta^e(s_0)) + \text{seno}(\widehat{\theta^e}(s_0))\text{seno}(\theta^e(s_0))$$

$$= \frac{\sum_{i=1}^{n} w_i \left[\cos(\theta^e(s_i))\cos(\theta^e(s_0)) + \text{seno}(\theta^e(s_i))\text{seno}(\theta^e(s_0))\right]}{\sqrt{W^T K W}}$$

$$= \frac{\sum_{i=1}^{n} w_i \cos[\theta^e(s_i) - \theta^e(s_0)]}{\sqrt{W^T K W}}$$

$$= \frac{W^T C}{W^T K W}$$

El problema de minimización es equivalente al de maximización de

$$E\left[\frac{W^T C}{\sqrt{W^T K W}}\right], \text{ con } C^T = [\cos(\theta^e(s_1) - \theta^e(s_0)), \dots, \cos(\theta^e(s_n) - \theta^e(s_0))]$$

Usando la expansión de Taylor de primer orden se tiene

$$E\left[\frac{W^T C}{\sqrt{W^T K W}}\right] = \frac{W^T \widetilde{C}}{\sqrt{W^T \widetilde{K} W}}$$

$\widetilde{K}$ y $\widetilde{C}$ se obtienen ajustando un modelo de covariograma circular teórico al cosenograma experimental residual. Usando nuevamente el método de los multiplicadores de Lagrange se obtiene

$$W = \frac{\widetilde{K}^{-1}\widetilde{C}}{\sqrt{\widetilde{C}^T\widetilde{K}^{-1}\widetilde{C}}} \tag{2-25}$$

Una vez hallada la predicción $\widehat{u}^e(s_0)$ se obtiene la predicción angular como

$$\widehat{\theta}^e(s_i) = \arctan\left(\frac{\widehat{\text{seno}(\theta^e(s_i))}}{\widehat{\cos(\theta^e(s_i))}}\right) \tag{2-26}$$

La varianza de predicción del kriging circular y su correspondiente estimación están dadas por [33]

$$\sigma^2_{\text{ck}} = 2 - 2\sqrt{C^T K^{-1} C} \tag{2-27}$$

y

$$\widehat{\sigma}^2_{\text{ck}} = 2 - 2\sqrt{\widetilde{C}^T\widetilde{K}^{-1}\widetilde{C}} \tag{2-28}$$

3 Geoestadística circular no estacionaria basada en redes neuronales

3.1. Kriging residual de datos circulares basado en redes neuronales

Kriging residual basado en redes neuronales es un método ampliamente usado en el análisis de datos geoestadísticos cuando no hay estacionariedad. Por ejemplo [12] y [44] muestran aplicaciones con datos climáticos, [49] emplea esta técnica en estudios geológicos y [36] lo utiliza en un análisis de suelos. En esta sección se extiende esta metodología al contexto de datos circulares. Específicamente se propone una técnica que permite hacer predicción espacial de datos angulares combinando redes neuronales y kriging residual angular. En 3.1.1 se describe el procedimiento para estimar la tendencia espacial de los datos angulares. Posteriormente en 3.1.2 se establece como hacer predicción residual angular. Las dos etapas anteriores son combinadas en para obtener la predicción en términos de la variable angular original.

3.1.1. Estimación de la tendencia espacial empleando redes neuronales

Suponga que tiene una estructura de datos circulares georreferenciados como la representada en la Tabla **3-1**. En el contexto de datos angulares se requiere hacer estimaciones separadas de la tendencia espacial para dos conjuntos de datos reales en el intervalo $[-1, 1]$. Se procede inicialmente obteniendo a partir de la muestra aleatoria angular $\theta(s_i)$, $i = 1, ..., n$, los valores de seno $(\theta(s_i))$

Sitio	x	y	θ	seno$(\theta(s_i))$	cos$(\theta(s_i))$
s_1	x_1	y_1	$\theta(s_1)$	seno $(\theta(s_1))$	cos$(\theta(s_1))$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s_n	x_n	y_n	$\theta(s_n)$	seno $(\theta(s_n))$	cos$(\theta(s_n))$

Tabla 3-1: Estructura de un conjunto de datos circulares georreferenciados con coordenadas x y y (latitud y longitud) y $\theta(s_i)$ una variable angular en el sitio $s_i = (x_i, y_i), i = 1, \ldots, n$. Los valores seno$(\theta(s_i))$ y cos$(\theta(s_i)$, pueden ser usados para caracterizar el comportamiento de $\theta(s)$ en la región de interés.

y cos $(\theta(s_i))$ (Tabla **3-1**). Se asume que cada una de las componentes tiene una tendencia espacial que es estimada de manera independiente empleando redes neuronales (Tabla **3-2**). En cada caso hay una capa de entrada, una de salida y una o varias capas ocultas. En estas se considera una función de activación tangente hiperbólico puesto que su utilización limita los valores de respuesta al intervalo [-1, 1]. La primera capa de la red toma como parámetros de entrada las coordenadas de ubicación geográfica x (longitud) y y (latitud). La capa de salida retorna de manera independiente, según el caso, los valores estimados $\widetilde{\text{seno}(\theta(s_i))}$ y $\widetilde{\cos(\theta(s_i))}$. La función de error utilizada en la red se establece mediante la comparación de la estimación realizada durante el proceso de aprendizaje $\widetilde{\theta}(s_i)$ y el ángulo registrado en la muestra $\theta(s_i)$. Específicamente se busca minimizar la función de error

$$\mathrm{E} = \sum_{i=1}^{n} \eta\left(\widetilde{\theta}(s_i)\right)$$

Según el caso, estimación de seno$(\theta(s_i))$ o cos$(\theta(s_i))$, la función $\eta\left(\widetilde{\theta}(s_i)\right)$ está dada por:

$$\begin{aligned}\eta\left(\widetilde{\theta}(s_i)\right) &= \frac{1}{2}\left(1-\cos\left(\text{arcoseno}(\widetilde{\text{seno}(\theta(s_i))}-\theta(s_i)\right)\right)\\ &= \frac{1}{2}\left(1-\cos\left(\widetilde{\theta}(s_i)-\theta(s_i)\right)\right)\end{aligned} \tag{3-1}$$

o

$$\begin{aligned}\eta\left(\widetilde{\theta}(s_i)\right) &= \frac{1}{2}\left(1-\cos\left(\text{arcocos}(\widetilde{\cos(\theta(s_i))}-\theta(s_i)\right)\right)\\ &= \frac{1}{2}\left(1-\cos\left(\widetilde{\theta}(s_i)-\theta(s_i)\right)\right)\end{aligned} \tag{3-2}$$

Las expresiones (3-1) y (3-2) en el contexto circular son equivalentes a la diferencia de cuadrados en el análisis de variables reales. La estimación angular final de la tendencia espacial (después de optimizadas las funciones de error 3-1 y 3-2) es definida como

$$\widetilde{\theta}(s_i) = \arctan\left(\frac{\widetilde{\text{seno}(\theta(s_i))}}{\widetilde{\cos(\theta(s_i))}}\right) \tag{3-3}$$

En la tabla **3-2** se resume el procedimiento de estimación de la variable angular $\theta(s_i), i = 1,\ldots,n$ empleando las estimaciones de las funciones seno y coseno en cada sitio de observación Una vez entrenada las redes neuronales de ambas funciones se puede hacer estimación de la tendencia angular en sitios sin observación s_0 tomando como entrada sus coordenadas de ubicación espacial (x_0, y_0). Específicamente

$$\widetilde{\theta}(s_0) = \arctan\left(\frac{\widetilde{\text{seno}(\theta(s_0))}}{\widetilde{\cos(\theta(s_0))}}\right) \tag{3-4}$$

3.1.2. Kriging residual circular basado en redes neuronales

Dada la estimación de la tendencia espacial $\widetilde{\theta}(s_i), i = 1,\ldots,n$ (Ecuación 3-3), se emplea inicialmente kriging residual para hacer predicción del error

Sitio	x	y	$\theta(s)$	$\widetilde{\text{seno}\,(\theta(s))}$	$\widetilde{\cos\,(\theta(s))}$	$\widetilde{\theta}(s)$=arctang $\left(\frac{\widetilde{\text{seno}\,(\theta(s))}}{\widetilde{\cos\,(\theta(s))}}\right)$
s_1	x_1	y_1	$\theta(s_1)$	$\widetilde{\text{seno}\,(\theta(s_1))}$	$\widetilde{\cos\,(\theta(s_1))}$	$\widetilde{\theta}(s_1) = \text{arctang}\left(\frac{\widetilde{\text{seno}\,(\theta(s_1)}}{\widetilde{\cos\,(\theta(s_1)}}\right)$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$
s_n	x_n	y_n	$\theta(s_n)$	$\widetilde{\text{seno}\,(\theta(s_n))}$	$\widetilde{\cos\,(\theta(s_n))}$	$\widetilde{\theta}(s_n) = \text{arctang}\left(\frac{\widetilde{\text{seno}\,(\theta(s_n))}}{\widetilde{\cos\,(\theta(s_n))}}\right)$

Tabla 3-2: Estructura de datos espaciales con respuesta angular $(x_i, y_i, \theta(s_i))$, con (x_i, y_i) representan las coordenadas espaciales (longitud, latitud) en s_i, $\theta(s_i)$ la *i-ésima* observación de la variable angular $\theta(s)$, $\widetilde{\theta}(s_i)$ la estimación de $\theta(s_i)$ basada en $\widetilde{\text{seno}\,(\theta(s_i))}$ y $\widetilde{\cos\,(\theta(s_i))}$ que son obtenidas las estimaciones de estas funciones halladas por medio de redes neuronales para datos circulares.

angular en sitios no observados. Posteriormente las predicciones son sumadas a la tendencia estimada por redes neuronales para obtener la predicción en términos de la variable angular original. La metodología comprende varias etapas:

1. Se calcula para cada uno de los sitios de observación el residual circular $\theta^{\text{e}}(s_i) = \left(\widetilde{\theta}(s_i) - \theta(s_i)\right)$, con $\theta(s_i), s_i, i = 1, \ldots, n$ la muestra del proceso estocástico espacial angular registrado.

2. Se ajusta un modelo de covariograma teórico (esférico, exponencial, Gaussiano) (Sección 2.7.1) al cosenograma calculado con $\theta^e(s_i), i = 1, \ldots, n$

$$\widetilde{C}(h) = \frac{1}{n(h)} \sum_{|||s_i - s_j|| - h| < \epsilon} \cos\left(\theta^e(s_i) - \theta^e(s_j)\right), \tag{3-5}$$

   En la práctica es difícil encontrar muchas parejas de sitios (s_i, s_j) cuyas distancias sean exactamente iguales a h. Por ello, en aras de llevar a

cabo el ajuste de un modelo de covariograma circular al cosenograma residual (ver Sección 3-5) se fija la tolerancia $|||s_i - s_j|| - h| < \epsilon$, es decir se buscan todas aquellas parejas de sitios con distancias aproximadas (por debajo o por encima) a h.

3. Se define $\mathbf{u}_i^e = \begin{bmatrix} \text{seno}(\theta^e(s_i)) \\ \cos(\theta^e(s_i)) \end{bmatrix}, i = 1, \ldots, n$, el i-ésimo vector angular residual. Se hace predicción de los residuales angulares en sitios no observados por medio de kriging ordinario circular empleando como entrada los residuales angulares.

$$\begin{aligned} \widehat{\mathbf{u}}_0^e = \begin{bmatrix} \widehat{\text{seno}(\theta^e(s_0)} \\ \widehat{\cos(\hat{\theta}_0^e)} \end{bmatrix} &= w_1^e \begin{bmatrix} \text{seno}(\theta^e(s_1)) \\ \cos(\theta^e(s_1)) \end{bmatrix} + \ldots + w_n^e \begin{bmatrix} \text{seno}(\theta^e(s_n)) \\ \cos(\theta^e(s_n)) \end{bmatrix} \\ &= [\mathbf{u}_1^e, \ldots, \mathbf{u}_n^e] \begin{bmatrix} w_1^e \\ \vdots \\ w_n^e \end{bmatrix} \\ &= (\mathbf{U}^e)^T \mathbf{W}^e. \end{aligned}$$

Los pesos son estimados como (ver sección 2.7.4)

$$\mathbf{W}^e = \frac{(\mathbf{K}^e)^{-1}\mathbf{C}^e}{\sqrt{(\mathbf{C}^e)^T \mathbf{K}^{e-1}\mathbf{C}^e}}, \text{ donde}$$

$$\mathbf{K}^e = \begin{bmatrix} 1 & (\mathbf{u}_1^e)^T\mathbf{u}_2^e & \cdots & (\mathbf{u}_1^e)^T\mathbf{u}_n^e \\ (\mathbf{u}_2^e)^T\mathbf{u}_1^e & 1 & \cdots & (\mathbf{u}_2^e)^T\mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{u}_n^e)^T\mathbf{u}_1^e & (\mathbf{u}_n^e)^T\mathbf{u}_2^e & \cdots & 1 \end{bmatrix},$$

$$\text{y } (\mathbf{C}^e)^T = [\cos(\theta^e(s_1) - \theta^e(s_0)), \ldots, \cos(\theta^e(s_n) - \theta^e(s_0))]$$

4. Dado $\widehat{\mathbf{u}}_0^e$ la predicción residual angular se calcula como

$$\widehat{\theta}^e(s_0) = \arctan\left(\frac{\widehat{\text{seno}(\theta^e(s_0)}}{\widehat{\cos(\theta^e(s_0)}}\right) \tag{3-6}$$

5. La varianza de predicción del error angular se calcula mediante

$$\widehat{\sigma}^2_{\text{ck}} = 2 - 2\sqrt{(\widetilde{\mathbf{C}}^e)^T(\widetilde{\mathbf{K}}^e)^{-1}\widetilde{\mathbf{C}}^e}, \tag{3-7}$$

   con $(\widetilde{\mathbf{C}}^e)$ la estimación de $(\mathbf{C}^e)$

6. La predicción angular en un sitio de interés $s_0 = (x_0, y_0)$ en la región de estudio se obtiene sumando la tendencia estimada por redes neuronales (ecuación 3-4) y la predicción residual circular (ecuación 3-6). Específicamente

$$\widehat{\theta}(s_0) = \widetilde{\theta}(s_0) + \widehat{\theta}^e(s_0) \tag{3-8}$$

7. La varianza de predicción de la variable angular es igual a la definida en (3-7)

8. Los resultados de las predicciones y de las varianzas de predicción pueden representarse de diversas maneras. La más común es empleando gráficos de vectores que indican las direcciones predichas y sus respectivas magnitudes.

4 Estudio de simulación

El estudio de datos simulados es una alternativa comúnmente empleada en estadística para comprobar las bondades de nuevas metodologías de análisis. Las simulaciones facilitan la comparación de estrategias novedosas de modelación con otras previamente establecidas. Los resultados de estudios de simulación ofrecen la posibilidad de identificar los pros y contras de técnicas innovadoras. En geoestadística es muy común su utilización para cuantificar el desempeño de nuevas alternativas de predicción kriging ([13], [27], [54]). En esta sección se analizan datos simulados con el propósito de comparar técnicas de kriging residual para datos circulares basadas, respectivamente, en pulimento de medianas y redes neuronales. Con el procedimiento descrito en la sección 2.7.3 se generan datos de campos aleatorios circulares con distintos niveles de dependencia espacial empleando simulaciones de distribuciones von Mises y a partir de estos se comparan las dos estrategias de análisis geoestadístico de datos circulares no estacionarios. Para determinar las diferencias entre las dos metodologías se utilizan dos estadísticas obtenidas a partir de los ángulos simulados y estimados: El error medio absoluto de la diferencia de cosenos (EMADC) y el error medio del coseno de las diferencias (EMCD) definidos respectivamente como:

$$\text{EMADC} = \frac{1}{n}\sum_{i=1}^{n} |\cos(\theta(s_i) - \cos(\hat{\theta}(s_i)| \tag{4-1}$$

$$\text{EMCD} = \frac{1}{n}\sum_{i=1}^{n} \cos(\theta(s_i) - \hat{\theta}(s_i)) \tag{4-2}$$

4.1. Resultados

Se obtienen simulaciones de tres campos aleatorios circulares a partir de distribuciones von Mises centradas en cero con κ =0.8, 3.1, 8.0 y estructura de correlación esférica (ver Sección 2.7.3). Siguiendo el ejemplo de simulación presentado en la sección 6.1 de [33] se emplea un enmallado regular de 121 puntos equiespaciados con coordenadas (x, y), $x, y = 1, 2, \ldots, 11$. Con base en los ángulos simulados (vectores en negro en las Figuras **4-2** y **4-3**) se hace comparación de los resultados obtenidos por kriging residual empleando pulimento de medianas (Figura **4-2**) y redes neuronales (Figura **4-3**), respectivamente. Inicialmente se obtienen las estimaciones de los cosenogramas y se hace el ajuste de modelos teóricos de covariograma (puntos y curvas en la Figura **4-1**)

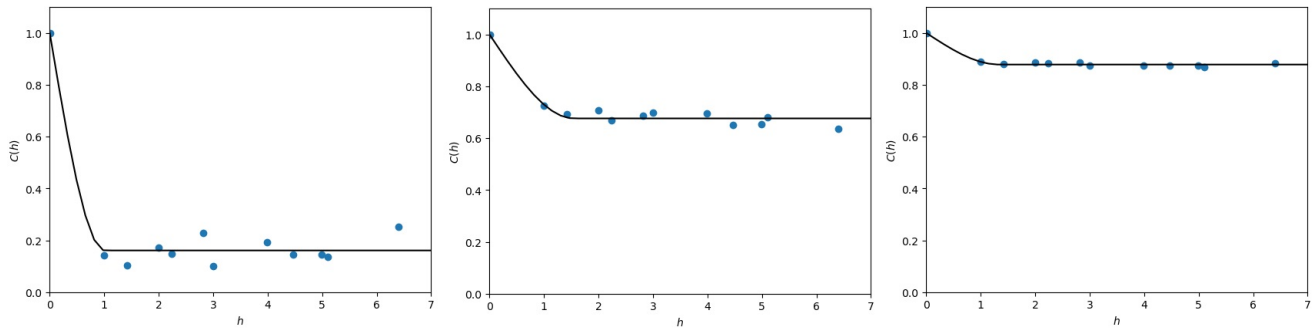

Figura 4-1: Funciones de covarianza ajustadas (curvas) a los cosenogramas experimentales (puntos) correspondientes a tres conjuntos de datos simulados a partir de distribuciones von Mises con parámetros κ=0.8 (izquierda), 3.1 (centro) y 8.0 (derecha).

Los modelos de covarianza ajustados (curvas en negro) a los cosenogramas (puntos azules) en los tres conjuntos de datos no involucran efecto pepita

τ^2, es decir que no se asume discontinuidad en el origen (Figura **4-1**). Se emplearon modelos de covarianza esféricos en los tres casos. Las estimaciones de los parámetros son dadas en la Tabla **4-1**. Es claro de las modelos ajustados a los cosenogramas que a mayor κ hay menor varianza (menor valor de $\hat{\sigma}^2$) y ligeramente mayor correlación (mayor $\hat{\phi}$). De lo anterior, es razonable que las predicciones basadas en datos simulados de distribuciones con mayor κ tengan un mejor desempeño, es decir que en el proceso de validación cruzada es esperable que se obtengan menores residuales (mayor similitud entre los ángulos simulados y los predichos) cuando incrementa κ.

κ	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
0.8	0.0	0.83	1.01
3.1	0.0	0.27	1.12
8.0	0.0	0.12	1.22

Tabla 4-1: Estimaciones de los parámetros de modelos de covarianza circular esféricos ajustados a datos simulados de distribuciones von Mises con tres valores de κ.

Inicialmente en la Figura **4-2** se muestran los resultados de validación cruzada obtenidos por kriging residual circular basado en pulimento de medianas. Puede notarse que en la medida en que aumenta κ mejoran las predicciones. Esto es debido a que disminuye la variabilidad de los procesos subyacentes. Un resultado muy parecido se observa en la Figura **4-3** correspondiente a las predicciones de validación cruzada por la técnica propuesta que se soporta en el uso de redes neuronales como método de estimación de la tendencia espacial en los datos angulares. Nuevamente a mayor κ menor error de predicción.

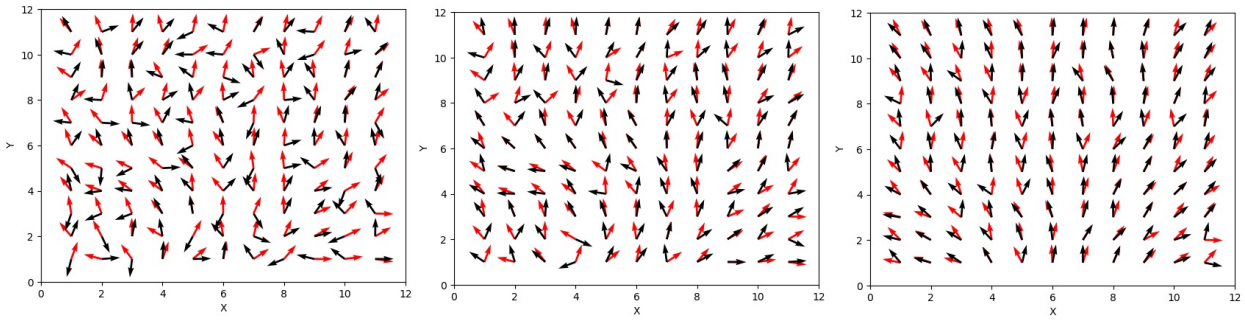

Figura 4-2: Representación de los datos angulares $\theta(s_i)$ simulados (vectores en negro) a partir de campos aleatorios circulares von Mises con κ=0.8 (izquierda), κ=3.1 (centro) y κ=8 (derecha) y de las predicciones obtenidas por medio del kriging residual con pulimento de medianas $\hat{\theta}(s_i)$, $i = 1, ..., 121$ (vectores en rojo)

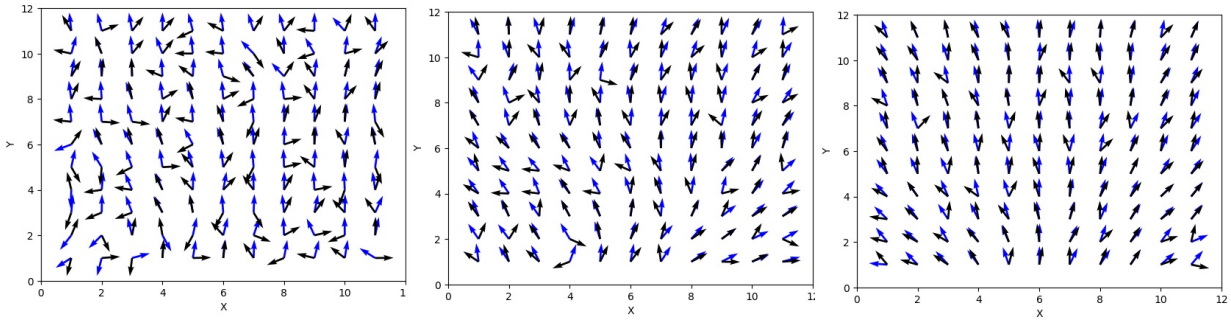

Figura 4-3: Representación de los datos angulares $\theta(s_i)$ simulados (vectores en negro) a partir de campos aleatorios circulares von Mises con κ=0.8 (izquierda), κ=3.1 (centro) y κ=8 (derecha) y de las predicciones obtenidas por medio de kriging residual con redes neuronales $\hat{\theta}(s_i)$, $i = 1, ..., 121$ (vectores en azul)

Los resultados gráficos parecen sugerir que las dos metodologías son semejantes. Para discernir numéricamente respecto a las posibles diferencias entre

las dos técnicas se emplean los criterios EMADC y EMCD definidos respectivamente en las ecuaciones 4-1 y 4-2. La evaluación del desempeño de cada metodología (kriging residual basado en pulimento de medianas y en redes neuronales) se realiza mediante un proceso de validación cruzada (dejando una observación fuera), en el que se calcula la predicción en cada sitio $\hat{\theta}(s_i)$, $i = 1, ..., 121$, con base en los 120 restantes. La Figura **4-2** presenta en color rojo los valores simulados y en negro las predicciones basadas en kriging residual circular usando pulimento de medianas. Es claro que las predicciones mejoran a medida que κ aumenta. Esto se debe nuevamente a que disminuye la variabilidad de los datos. Las predicciones con kriging residual de datos circulares empleando redes neuronales son mostradas en la Figura **4-3** (vectores en azul). Al igual que en el caso de pulimento de medianas, hay mejora en las predicciones cuando se incrementa κ. Los valores de las medidas EMADC y EMDC para ambas metodologías son presentados en las Tablas **4-2** y **4-3**. Los resultados revelan que la metodología basada en redes neuronales tiene mejor desempeño en todos los escenarios de simulación. Es también claro que a medida que aumenta κ se tiene menor error, puesto que como se ha mencionado anteriormente cuando hay un incremento de este parámetro se obtiene una menor dispersión.

EMADC	κ=0.8	κ =3.1	κ =8
Pulimiento de Medianas	0.389	0.113	0.047
Red Neuronal	0.335	0.086	0.037

Tabla 4-2: Valores del error medio absoluto de la diferencia de cosenos (EMADC) calculados con los errores de validación cruzada obtenidos por kriging residual circular basado en pulimento de medianas y redes neuronales en cada uno de tres campos aleatorios circulares con distribución von Mises de parámetro κ.

EMDC	κ=0.8	κ=3.1	κ=8
Puliminto de Medianas	0.726	0.444	0.300
Red Neuronal	0.636	0.386	0.270

Tabla 4-3: Valores del error medio de diferencia de cosenos (EMDC) calculados con los errores de validación cruzada obtenidos por kriging residual circular basado en pulimento de medianas y redes neuronales en cada uno de tres campos aleatorios circulares con distribución von Mises de parámetro κ.

Aunque no se ha realizado en este trabajo un estudio suficientemente amplio que permita determinar la significancia estadística de las diferencias entre las dos técnicas, los resultados de las simulaciones sugieren que la metodología propuesta basada en redes neuronales para datos circulares podría presentarse como una opción más efectiva en comparación con la soportada en el pulimento de medianas.

5 Análisis de datos reales: Predicción espacial de dirección del viento en una región de Colombia

El proyecto NASA POWER [45] es una iniciativa de la NASA (National Aeronautics and Space Administration) que ofrece conjuntos de datos solares y meteorológicos con el objetivo de respaldar investigaciones en áreas como energía renovable, eficiencia energética y necesidades socio-económicas y culturales. La NASA desempeña un papel fundamental en la investigación y la expansión de nuestro conocimiento sobre el planeta. Esta institución recopila y distribuye información precisa y detallada de las condiciones meteorológicas en todo el mundo y particularmente sobre la radiación solar. En aras de hacer una aplicación a un conjunto de datos reales de la metodología propuesta en el trabajo se consideró un conjunto de datos meteorológicos recopilados por esta fuente. Específicamente se colectaron 81 registros de dirección del viento del día 31 de octubre de 2010 a las 10 am en una subregión de Colombia delimitada por las coordenadas de longitud $(2.75°, 6.75°)$ y latitud $(-78.25°, -74.25°)$ (Figura **5-1**). Para realizar el análisis, los datos originales obtenidos en grados de latitud y longitud fueron transformados a coordenadas planas UTM.

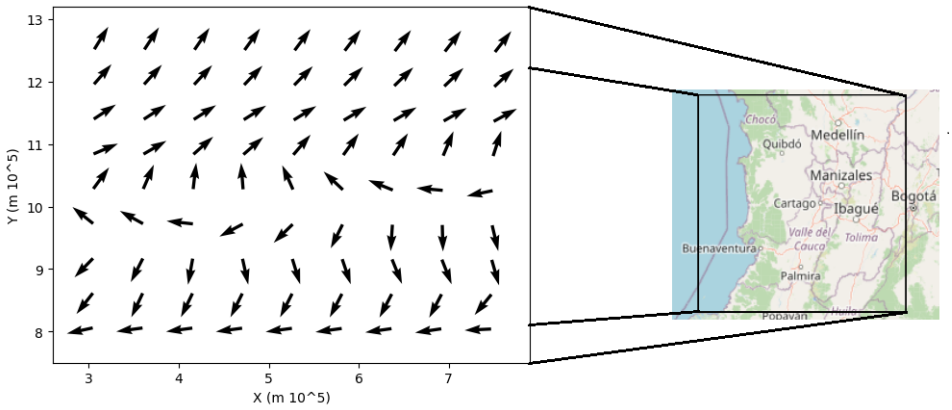

Figura 5-1: Representación de datos de dirección del viento tomados del proyecto NASA power en una región de Colombia el día 31 de octubre del 2010 a las 10 am.

5.1. Análisis de validación cruzada

Aquí se hace nuevamente una comparación entre kriging circular basado en pulimento de medianas y la propuesta fundamentada en el uso de redes neuronales. Las medidas de desempeño EMADC y EMDC (Ecuaciones 4-1 y 4-2) consideradas en el estudio de simulación fueron también empleadas en este caso para establecer las diferencias entre las dos estrategias de análisis. La Figura **5-2** muestra una comparación entre los resultados de la validación cruzada obtenidos con las dos metodologías (a la izquierda pulimento de medianas y a la derecha redes neuronales). A simple vista es claro que el rendimiento es mejor cuando se emplea la segunda opción. Puede percibirse que los vectores negros y rojos (observados y predichos) en el panel derecho (método basado en redes neuronales) están prácticamente superpuestos, es decir que las predicciones están muy cercanas a las observaciones, mientras que en el panel izquierdo hay muchas discrepancias, particularmente en la zona centro y sur de la región considerada. Las gráficas insinúan que cuando hay un cambio drástico en la tendencia el algoritmo basado en pulimento de medianas presenta mayor error e indican que en general la metodología

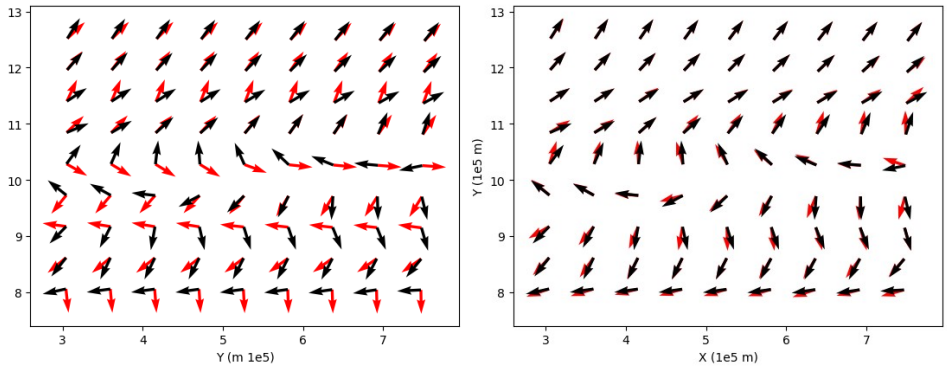

Figura 5-2: Representación de los datos angulares observados (direcciones del viento) $\theta(s_i), i = 1, \ldots, 81$ (vectores en negro en ambos paneles) y predicciones $\hat{\theta}(s_i)$ obtenidas por medio de kriging residual con pulimento de medianas (vectores en rojo en el lado izquierdo) y kriging residual con redes neuronales (vectores en rojo en el panel derecho).

propuesta, soportada en redes neuronales, brinda predicciones más precisas (hay menor error de predicción). Estas revelan también que el método fundamentado en redes neuronales tiene mayor capacidad de adaptación a cambios bruscos de tendencia. Nótese que en la zona centro y sur del área de estudio se presentan cambios bruscos de dirección que no logran ser capturados por el método de predicción cimentado en el uso de pulimento de medianas, mientras que estos mismos si son capturados por el construido a partir de redes neuronales. Note en los dos paneles de la Figura **5-2** que hay diferencias en las escalas de las direcciones sur-norte y este-oeste (las distancias en sentido norte son menores). Esta diferencia de escala puede afectar la capacidad del método de pulimento de medianas para estimar de manera adecuada la componente horizontal de los datos. En contraste, la metodología soportada en redes neuronales se adapta mejor a la estimación dada su capacidad de capturar relaciones y patrones complejos en datos de dirección del viento [42].

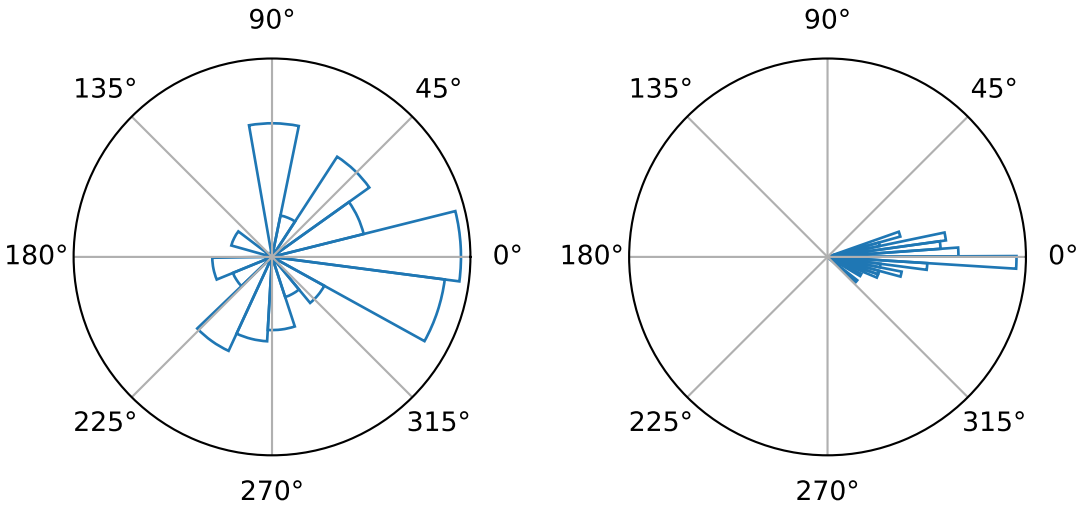

Figura 5-3: Histograma de residuales de validación cruzada correspondientes a los métodos basados en pulimento de medianas (izquierda) y redes neuronales (derecha).

Los histogramas de los residuales de predicción con ambas técnicas (Figura **5-3**) confirman que hay mucha menos dispersión cuando se aplican redes neuronales. En ambos casos el promedio angular está alrededor de cero, pero con más variabilidad en el caso de pulimento de medianas. Una comparación formal de los errores de predicción de las dos técnicas se obtiene con las medidas EMADC y EMDC (Ecuaciones 4-1 y 4-2). Estas son mostradas en las Tabla **5-1**. Nuevamente este insumo permite concluir que la combinación de redes neuronales y kriging residual circular resulta más favorable en el contexto estudiado. El coeficiente de concentración V (Sección 2.6.2) es otro

estadístico que permite inferir en este caso que hay mucha mayor variabilidad en los residuales resultantes de la predicción con la combinación de pulimento de medianas y kriging residual circular.

	EMADC	EMDC	V
Pulimiento de medianas	0.542	0.251	0.485
Red neuronal	0.071	0.006	0.014

Tabla 5-1: Errores EMADC y EMDC del proceso de validación cruzada aplicado a dos metodologías de kriging residual en datos reales. Se muestra también el valor del coeficiente de concentración (variabilidad de datos angulares).

En resumen, los resultados del análisis de validación cruzada de datos angulares reales están en la misma línea de lo establecido con datos procesos espaciales angulares simulados, es decir que la metodología que combina redes neuronales con kriging circular residual presenta un mejor rendimiento en la predicción espacial de los datos circulares que los hallados con el método soportado en el uso de pulimento de medianas. Los residuales de los análisis de datos simulados y reales revelan que cuando hay complejidad en la tendencia espacial (que cambia en distintos sentidos de forma no lineal) y alta variabilidad, la predicción con la opción de redes neuronales y kriging circular residual es mucho más llamativa. Los resultados presentados en la tabla **5-1** resaltan la superioridad de la metodología basada en redes neuronales. Esta misma permite que haya una reducción significativa del error y un aumento en el coeficiente de concentración para los residuales. Estos resultados indican que la metodología de redes neuronales es capaz de proporcionar estimaciones más precisas y consistentes en comparación con el método alternativo de estimación de la tendencia.

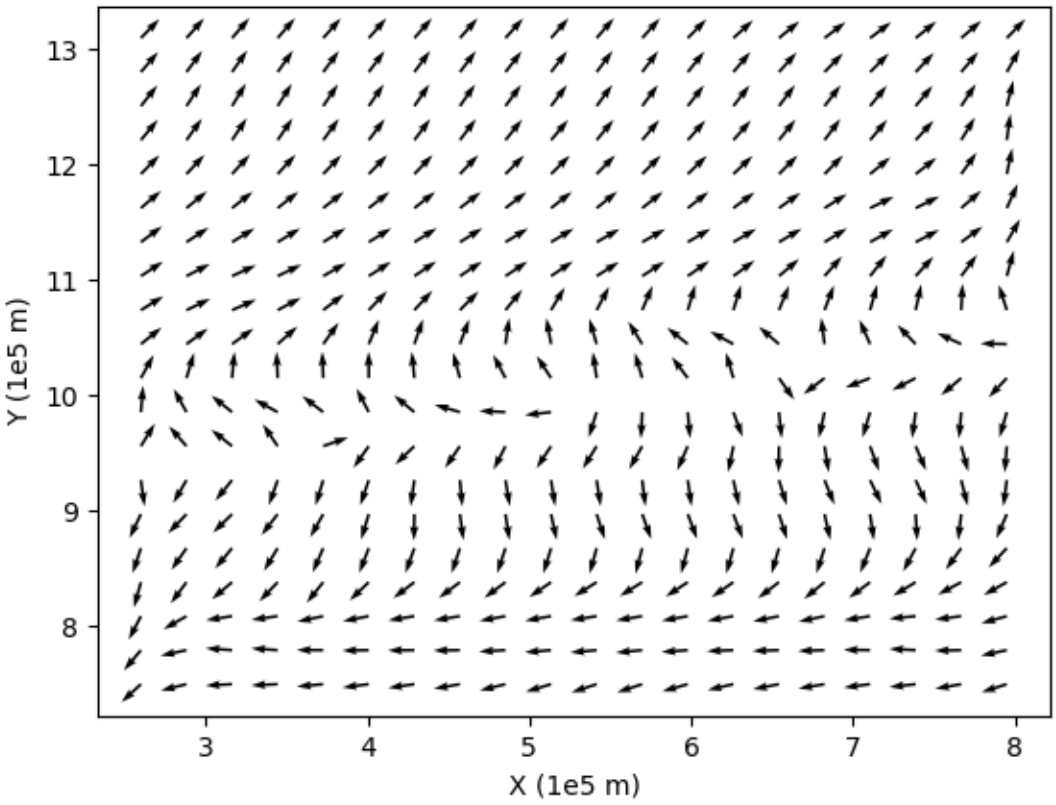

Figura 5-4: Predicción del viento usando kriging residual con redes neuronales de 400 puntos no medidos a partir de 81 mediciones

5.2. Predicción en sitios no muestreados y caracterización de los vientos

En el análisis de datos reales también se lleva a cabo predicción en sitios sin información (Figura **5-4**). Específicamente a partir de los 81 registros de dirección del viento tomados de la base de datos de estudio, se realizó predicción en 400 puntos del área de interés no considerados en la información inicial. Los vectores predichos permiten caracterizar el comportamiento de la dirección del viento el día de la captura de información en la región considerada. En la parte norte se perciben vientos con dirección nor-oriental y en la sur predominan vientos con dirección sur, sur-occidental y occidental. Este

patrón puede ser explicado por varias causas. Por un lado, el comportamiento de los vientos en la costa occidental de Colombia en octubre se caracteriza por una combinación de factores geográficos y meteorológicos. La topografía y la geografía de la región pueden influir en la dirección de los vientos [40]. En la parte norte de la costa, es posible que la presencia de cadenas montañosas o valles pueda canalizar los vientos hacia una dirección nor-oriental. En cambio, en la sur la topografía y la configuración geográfica pueden favorecer vientos que provienen del sur, sur-occidente y occidente [2]. La corriente de Panamá, una corriente oceánica cálida que fluye a lo largo de la costa del pacífico de América del sur, puede tener un impacto en los patrones de viento [41]. Esta corriente puede influir en la dirección de los vientos costeros, especialmente en la parte sur de la costa. La variación estacional también desempeña un papel importante en los patrones de viento. En octubre, Colombia experimenta cambios en la intensidad y dirección de los vientos debido a factores estacionales, como la posición del sol y la interacción entre sistemas de alta y baja presión. Los frentes atmosféricos que se mueven a través de la región pueden cambiar la dirección de los vientos. La interacción de masas de aire frío y cálido puede influir en la dirección y velocidad de los vientos costeros [15].

6 Conclusiones y trabajo futuro

Los datos circulares presentan una estructura matemática que requiere un tratamiento diferente al de los datos de variables reales comúnmente considerados en estadística. En el campo de la geoestadística, especialmente en la estimación de la estructura de correlación espacial y en la predicción mediante kriging, es esencial tener en cuenta estas particularidades. En este trabajo, se propone una alternativa para realizar predicción espacial cuando se dispone de datos de campos aleatorios no estacionarios. La propuesta consiste en utilizar redes neuronales para modelar la tendencia espacial y luego tratar los residuales circulares de manera análoga a como se hace con datos circulares estacionarios. La predicción final se obtiene sumando la tendencia estimada con la predicción circular residual.

La metodología se valida utilizando datos reales y simulados. Se realiza una comparación con una propuesta basada en el método de pulimento de medianas. Los resultados demuestran que la estrategia planteada ofrece mejores resultados en términos de los errores de predicción en la validación cruzada. Los hallazgos de este estudio son de gran relevancia ya que subrayan la efectividad y confiabilidad de la metodología basada en redes neuronales en el análisis de datos circulares. La disminución considerable en el error significa que las predicciones realizadas con esta metodología son más exactas y confiables, lo que puede tener un impacto positivo en diversos campos de aplicación, como la energía renovable, la eficiencia energética y otros estudios relacionados.

Este estudio plantea nuevas oportunidades para explorar y mejorar la metodología de redes neuronales en el contexto del análisis y modelado de datos circulares. Los resultados obtenidos no solo respaldan la validez de esta

técnica, sino que también resaltan su potencial y su aplicabilidad en una variedad de escenarios. La efectividad demostrada en la reducción del error y el aumento del coeficiente de concentración, sugiere que la metodología de redes neuronales tiene un papel prometedor en futuros estudios y aplicaciones prácticas en el campo de la estadística espacial y el análisis de datos espaciales. Se plantea como trabajo futuro la extensión del método de predicción kriging universal al caso de datos circulares georreferenciados y la predicción espacial de datos circulares en presencia de covariables empleando cokriging.

Bibliografía

[1] Albawi, S ; Mohammed, T ; Al-Zawi, S: Understanding of a convolutional neural network. En: *2017 international conference on engineering and technology (ICET)* Ieee, 2017, p. 1–6

[2] Arroyo, L: Estudio de la variabilidad espacio-temporal de la precipitación, el viento y la humedad en la región del Urabá antioqueño a través de resultados de WRF. (2023)

[3] Atkinson, P ; Lloyd, C: *geoENV VII–geostatistics for environmental applications*. Vol. 16. Springer Science & Business Media, 2010

[4] Breckling, J: *The analysis of directional time series applications to wind speed and direction*. Springer, 1989

[5] Breckling, J: *The analysis of directional time series: applications to wind speed and direction*. Vol. 61. Springer Science & Business Media, 2012

[6] Carrat, F ; Valleron, A: Epidemiologic mapping using the "kriging" method: application to an influenza-like epidemic in France. En: *American journal of epidemiology* 135 (1992), Nr. 11, p. 1293–1300

[7] Carrera, J ; Alcolea, A ; Medina, A ; Hidalgo, J ; Slooten, L: Inverse problem in hydrogeology. En: *Hydrogeology Journal* 13 (2005), p. 206–222

[8] Coble, K ; Mishra, A ; Ferrell, S ; Griffin, T: Big data in agriculture: A challenge for the future. En: *Applied Economic Perspectives and Policy* 40 (2018), Nr. 1, p. 79–96

[9] Cressie, N: Geostatistical analysis of spatial data. En: *Spatial statistics and digital image analysis* 1991 (1991), p. 87–108

[10] Cressie, N: *Statistics for spatial data*. John Wiley & Sons, 2015

[11] Cuador-Gil, J ; Quintero-Silverio, A: Simulación condicional de variables regionalizadas y su aplicación al comportamiento de la porosidad efectiva en un yacimiento fracturado-poroso. En: *Boletín de la Sociedad Geológica Mexicana* 54 (2001), Nr. 1, p. 19–27

[12] Demyanov, V ; Kanevsky, M ; Chernov, S ; Savelieva, E ; Timonin, V: Neural network residual kriging application for climatic data. En: *Journal of Geographic Information and Decision Analysis* 2 (1998), Nr. 2, p. 215–232

[13] Dowd, P ; Sarac, C: A neural network approach to geostatistical simulation. En: *Mathematical Geology* 26 (1994), p. 491–503

[14] Emery, X ; Séguret, S: *Geostatistics for the Mining Industry: Applications to Porphyry Copper Deposits*. CRC Press, 2020

[15] Eslava, J: Climatología y diversidad climática de Colombia. En: *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 18 (1993), Nr. 71, p. 507–538

[16] Fernholz, L: *Von Mises calculus for statistical functionals*. Vol. 19. Springer Science & Business Media, 2012

[17] Gill, J ; Hangartner, D: Circular data in political science and how to handle it. En: *Political Analysis* 18 (2010), Nr. 3, p. 316–336

[18] Grancher, D ; Bar-Hen, A ; Paris, R ; Lavigne, F ; Brunstein, D: Spatial interpolation of circular data: application to tsunami of December 2004. En: *Advances and Applications in Statistics* 30 (2012), Nr. 1, p. 19–29

[19] Gribov, A ; Krivoruchko, K: Empirical Bayesian kriging implementation and usage. En: *Science of the Total Environment* 722 (2020), p. 137290

[20] Handcock, M ; Wallis, J: An approach to statistical spatial-temporal modeling of meteorological fields. En: *Journal of the American Statistical Association* 89 (1994), Nr. 426, p. 368–378

[21] Jona, L ; Santoro, M ; Mastrantonio, G: CircSpaceTime: an R package for spatial and spatio-temporal modelling of circular data. En: *Journal of Statistical Computation and Simulation* 90 (2020), Nr. 7, p. 1315–1345

[22] Kanevski, M ; Maignan, M: *Analysis and modelling of spatial environmental data*. Vol. 6501. EPFL press, 2004

[23] Kanevski, M ; Timonin, V ; Pozdnukhov, A: *Machine learning for spatial environmental data: theory, applications, and software*. EPFL press, 2009

[24] Kovacs-Györi, A ; Ristea, A ; Havas, C ; Mehaffy, M ; Hochmair, H ; Resch, B ; Juhasz, L ; Lehner, A ; Ramasubramanian, L ; Blaschke, T: Opportunities and challenges of geospatial analysis for promoting urban livability in the era of big data and machine learning. En: *ISPRS International Journal of Geo-Information* 9 (2020), Nr. 12, p. 752

[25] Lamamra, A ; Neguritsa, D ; Mazari, M: Geostatistical modeling by the Ordinary Kriging in the estimation of mineral resources on the

Kieselguhr mine, Algeria. En: *IOP Conference Series: Earth and Environmental Science* Vol. 362 IOP Publishing, 2019, p. 012051

[26] Langevin, P: Magnétisme et théorie des électrons. En: *Ann. chim. et phys.* (1905), Nr. 8, p. 203

[27] Lantuéjoul, C: *Geostatistical simulation: models and algorithms.* Springer Science & Business Media, 2001 (1139)

[28] Mardia, K ; Jupp, P: *Directional statistics.* Wiley, 2000

[29] Martínez, F: *Modelización de la función de covarianza en procesos espacio-temporales: análisis y aplicaciones.* Universitat de Valencia (Spain), 2008

[30] Martínez, W ; Melo, C ; Melo, O: Median polish kriging for space–time analysis of precipitation. En: *Spatial statistics* 19 (2017), p. 1–20

[31] Matías, J ; Vaamonde, A ; Taboada, J ; González-Manteiga, W: Comparison of kriging and neural networks with application to the exploitation of a slate mine. En: *Mathematical geology* 36 (2004), p. 463–486

[32] McNeill, L: Interpolation and smoothing of mapped circular data. En: *South African Statistical Journal* 27 (1993), Nr. 1, p. 23–49

[33] Morphet, W: *Simulation, kriging, and visualization of circular-spatial data.* Utah State University, 2009

[34] Oliver, A ; Webster, R: A tutorial guide to geostatistics: Computing and modelling variograms and kriging. En: *Catena* 113 (2014), p. 56–69

[35] Oliver, M ; Webster, R: *Basic steps in geostatistics: The Variogram and Kriging.* Springer, 2015

[36] Padarian, J ; Pérez-Quezada, J ; Seguel, O: Modelling the distribution of organic carbon in the soils of Chile. En: *Proceeding of the fifth global workshop on digital soil mapping, Digital Soil assessments and beyond, Sydney*, 2012, p. 329–333

[37] Pewsey, A ; Neuhäuser, M ; Ruxton, G: *Circular statistics in R.* OUP Oxford, 2013

[38] Playfair, W: *Playfair's commercial and political atlas and statistical breviary.* Cambridge University Press, 2005

[39] Rao, T: Spatial statistics and spatio-temporal data. En: *Journal of Time Series Analysis* 34 (2013), Nr. 2, p. 280–280

[40] Rodriguez-Rubio, E ; Stuardo, J: Variability of photosynthetic pigments in the Colombian Pacific Ocean and its relationship with the wind field using ADEOS-I data. En: *Journal of Earth System Science* 111 (2002), p. 227–236

[41] Rueda, J ; Rodríguez, E ; Ortiz, J: Caracterización espacio temporal del campo de vientos superficiales del Pacífico colombiano y el Golfo de Panamá a partir de sensores remotos y datos in situ. (2007)

[42] Sareen, K ; Panigrahi, B ; Shikhola, T ; Sharma, R: An imputation and decomposition algorithms based integrated approach with bidirectional LSTM neural network for wind speed prediction. En: *Energy* 278 (2023), p. 127799

[43] Seber, G ; Wild, J. *Nonlinear Regression.* 2003

[44] Seo, Y ; Kim, S ; Singh, V: Estimating spatial precipitation using regression kriging and artificial neural network residual kriging (RKNNRK) hybrid approach. En: *Water Resources Management* 29 (2015), p. 2189–2204

[45] Sparks, A: nasapower: a NASA POWER global meteorology, surface solar energy, and climatology data client for R. En: *Journal of Open Source Software* 3 (2018), Nr. 30, p. 1035

[46] Stein, M: *Interpolation of Spatial Data: Some theory for Kriging*. Springer, 2013

[47] Surjotedjo, H ; Widyaningsih, Y ; Nurrohmah, S: Median polish kriging model for circular-spatial data. En: *Empowering Science and Mathematics for Global Competitiveness*. CRC Press, 2019, p. 377–385

[48] Wackernagel, H: *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media, 2003

[49] Wang, L ; Wong, P ; Shibli, S: Modelling porosity distribution in the A'nan Oilfield: Use of geological quantification, neural networks and geostatistics. En: *SPE International Oil and Gas Conference and Exhibition in China* SPE, 1998, p. SPE–48884

[50] Webster, R ; Oliver, M: *Geostatistics for environmental scientists*. John Wiley & Sons, 2007

[51] Weisstein, E: Bessel function of the first kind. En: *https://mathworld.wolfram.com/* (2002)

[52] Xiao, L ; Zhang, Y: Zhang neural network versus gradient neural network for solving time-varying linear inequalities. En: *IEEE transactions on neural networks* 22 (2011), Nr. 10, p. 1676–1684

[53] Yan, Q ; Wan, Z ; Yang, C: Flight Load Calculation Using Neural Network Residual Kriging. En: *Aerospace* 10 (2023), Nr. 7, p. 599

[54] Zakeri, F ; Mariethoz, G: A review of geostatistical simulation models applied to satellite remote sensing: Methods and applications. En: *Remote Sensing of Environment* 259 (2021), p. 112381