

CÓPULAS: UN NUEVO ENFOQUE PARA LA MODELIZACIÓN EN GEOESTADÍSTICA

DANNA LESLEY CRUZ REYES¹, EDILBERTO CEPEDA CUERVO²,

¹ *Universidad El Bosque, Universidad Nacional de Colombia, DLCRUZR@UNAL.EDU.CO,*

² *Universidad Nacional de Colombia, ECEPEDAC@UNAL.EDU.CO,*

Resumen

Es común en geoestadística utilizar métodos como el variograma o el coeficiente de correlación para describir la dependencia espacial y kriging para realizar interpolación y predicción, pero estos métodos son sensibles a valores extremos y están fuertemente influenciados por la distribución marginal del campo aleatorio. Por lo tanto, pueden conducir a resultados poco fiables. Como alternativa a los modelos tradicionales de geoestadística se considera el uso de las funciones cópula. La cópula es ampliamente usada en el campo de las finanzas y ciencias actuariales y debido a sus resultados satisfactorios empezaron a ser consideradas en otras áreas de aplicación de las ciencias estadísticas. En este trabajo se muestra el efecto de las cópulas como una herramienta que muestra un análisis geoestadístico bajo todo el rango de cuantiles y una estructura de dependencia completa, considerando modelos de tendencia espacial, distribuciones marginales continuas y discretas y funciones de covarianza. Además se presentan tres métodos de interpolación espacial: el primero corresponde al indicador kriging y kriging disyuntivo, el segundo método se conoce como el kriging simple y el tercer método es una predicción plug-in y la generalización del kriging trans-Gaussiano, estos métodos son utilizados con base en la función cópula debido a la relación que existe entre las cópulas bivariadas y los indicadores de covarianzas. Finalmente se presenta resultados obtenidos para un conjunto de datos reales de la ciudad de Gomel que contiene mediciones de isótopos radioactivos, consecuencia del accidente nuclear de Chernóbil.

Palabras clave: Cópulas, geoestadística, estadística espacial, estadística computacional, tendencia..

1. INTRODUCCIÓN

Las cópulas describen la estructura de dependencia entre variables aleatorias, no es extraño que la palabra cópula insinúa vínculo o unión, proviene del latín y su significado es conexión o lazo que une dos cosas distintas, fue utilizada por primera vez por Sklar en su célebre teorema en 1959 (Ayyad 2008), para describir funciones de distribución multivariadas definidas sobre el cubo unidad $[0; 1]^n$ enlazando variables aleatorias con funciones de distribución de una sola dimensión.

Geoestadística se concentra en el estudio de fenómenos que se desarrollan en el tiempo mientras que pasan por fluctuaciones aleatorias, más formalmente, analiza las realizaciones de un campo aleatorio $\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{D}\}$ donde $D \subset \mathbb{R}$, cuya realización, $z(\mathbf{x})$, representa el valor de interés registrado en la medición con respecto a cierto sistema de referencia $\mathbf{x}$.

Actualmente existen herramientas que para modelar la variabilidad espacial. La primera fue usada en principios de los cincuenta por Danie G. Krige, en Sudáfrica ampliando técnicas estadísticas para la estimación de las reservas de minerales. En los años sesenta el trabajo de Krige fue formalizado por el matemático Georges Matheron, desde entonces ha sido ampliamente utilizado en áreas como la minería, la industria petrolera, hidrología, meteorología, oceanografía, el control del medio ambiente, la ecología del paisaje y la agricultura. A pesar de estos desarrollos, el modelado espacial a menudo se basa en hipótesis Gaussianas que muchas veces no se considera realista para los tipos de datos y/o se reportan datos atípicos que causan problemas en las investigaciones.

Bárdossy en el año 2006 fue el pionero en proponer el uso de cópulas para describir variabilidad espacial, (Bárdossy 2008) realiza modelación de campos aleatorios continuos, (Kazianka 2010b), adopta la metodología de Bárdossy y realiza una extensión considerando modelos de tendencia y campos aleatorios discretos, (Kazianka 2010a) muestran cómo se pueden incorporar en un marco bayesiano mediante la asignación de probabilidades apriori de todos los parámetros del modelo, en esta tesis se propone una extensión a estos modelos incluyendo las cópulas radialmente asimétricas.

Con el fin de describir la estructura de dependencia en un campo aleatorio en el área de geoestadística se enfoca en los métodos clásicos como propiedades de suavizamiento de los campos aleatorios, la función de autocorrelación, variogramas y técnicas para la interpolación espacial, kriging simple, kriging universal, cokriging, kriging disyuntivo, kriging Bayesiano entre otros. En este trabajo se presenta el análisis de estructura de dependencia a través de cópulas proponiendo una familia de cópulas radialmente asimétricas.

1.1. Descripción del campo aleatorio usando cópulas

Bardossy (Bárdossy 2008) presentó un método diferente para el modelado de dependencia espacial por medio de cópulas que aquí adoptaremos, se pretende describir todas las distribuciones multivariadas necesarias del campo aleatorio por medio de cópulas.

Se asume un campo aleatorio estacionario $Z(x)|x \in S$ donde $S \in \mathbb{R}^2$ es el área de interés. Se nota h el vector de separación entre dos puntos.

Sea F_Z la distribución univariante del proceso espacial, debido a que el campo aleatorio es estacionario F_Z es la misma para cada localización $x \in S$.

Con el teorema de Sklar, se puede establecer un modelo multivariante del campo tomando $F_Z = F_1 = F_2 = \dots = F_n$ tal que,

$$H(x_1, \dots, x_n) = C(F_1, F_2, \dots, F_n), \quad (1)$$

Entonces, la relación entre dos localizaciones separadas por el vector h esta caracterizada por la distribución bivalente:

$$P(Z(x) \leq z_1, Z(x+h) \leq z_2) = C_h(F_Z(z_1), F_Z(z_2)),$$

por tanto, la estructura de dependencia quedaría descrita por la función copula C_h en función del vector h , esto implica que la copula podría describir la estructura completa de dependencia a diferencia de los variogramas que solo describen con respecto a la media. Por otro lado, la elección de la cópula C será determinada aplicando varias familias de cópulas al modelo y comparando los diferentes resultados.

Es de esperarse que no todas las familias de cópulas continuas sea apropiadas para este modelo, de manera natural se puede suponer una cópula simétrica, debido a que la dependencia entre dos localizaciones x_2 y x_1 es la misma que x_1 y x_2 , de manera general se tiene que:

$$C_h(u_1, \dots, u_n) = C_h(u_{\pi(1)}, \dots, u_{\pi(n)}), \quad (2)$$

para una permutación arbitraria π .

Además, tenemos que añadir las siguientes restricciones:

- Cuando $\|h\| \rightarrow \infty$ entonces, $C_h(u) \rightarrow \Pi^n(x)$, ya que se quiere independencia sobre localizaciones muy alejadas entre si,
- Cuando $\|h\| \rightarrow 0$ entonces, $C_h(u) \rightarrow M^n(x)$ o equivalentemente, en localizaciones muy próximas entre si, queremos dependencia muy fuerte.

Estas condiciones son fundamentales para la construcción de la cópula, permite realizar un filtro de las cópulas que se podrían proponer, por ejemplo la familia Farlie-Gumbel-Morgenstern, no son útiles.

Debido a que los campos aleatorios más importantes son los Gaussianos es de esperarse que las cópulas también lo sean, la cópula Gaussiana definida en la ecuación 3 con marginal $F_Z = \Phi_{\mu, \sigma^2}$, donde μ y σ denota la media y varianza respectivamente, es muy utilizada en campos aleatorios,

$$C_{\Sigma}^G = \Phi_{0, \Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)). \quad (3)$$

Las ventajas de trabajar con estas cópulas es que son invariantes bajo transformaciones estrictamente crecientes de las variables aleatorias además de cumplir con las condiciones anteriormente dadas para campos aleatorios. La cópula gaussiana se convierte en una función de h , suponiendo que la función de correlación sigue uno de los modelos paramétricos conocidos, por ejemplo, el modelo Matern.

1.2. Cópulas empíricas bivariadas

Las cópulas empíricas son usadas por primera vez en el área de geoestadística, (Haslauer 2010), quien implementó las cópulas bivariadas empíricas, describiendo la estructura de dependencia entre variables aleatorias. En esta tesis se consideran este tipo de cópulas como una metodología de exploración de datos, de tal forma que se pueda considerar la forma de la distribución que puedan tener, para esto el campo aleatorio debe cumplir la condición de estacionariedad fuerte y como consecuencia se pueden obtener las siguientes ventajas, (Haslauer 2010):

- La distribución marginal, que podría distorsionar la estructura de dependencia, se filtra usando cópulas. Así, quedaría definida únicamente por los datos.
- La cópula permiten una mejora en la cuantificación de la incertidumbre en la interpolación.
- Un modelo estocástico completo es la columna vertebral para el análisis geoestadístico.

1.2.1. Algoritmo para la aplicación de las cópulas empíricas bivariadas

Las cópulas son usadas para explorar la estructura de dependencia entre dos variables aleatorias sin considerar las distribuciones marginales de cada variable. Las cópulas empíricas es el caso más simple de construcción pero no es computacionalmente óptimo, aun así, estas cópulas se pueden evaluar en diferentes direcciones y ángulos para cada par de puntos y generar una idea de la forma de la estructura de dependencia del campo aleatorio.

Según Haslauer, ((Haslauer 2010)) se debe considerar el campo aleatorio estacionario, la construcción de la cópula correspondiente se puede realizar con el siguiente algoritmo:

1. Se calcula la cópula empírica marginal $F_Z(z)$ de las observaciones.
2. Para algún vector h dado, se calcula el conjunto $S(h)$:

$$S(h) = \{(F_Z(z(s_i)), F_Z(z(s_j))) | |s_i - s_j| \approx h\} \quad (4)$$

3. Debido a que $S(h)$ es un conjunto de pares de puntos definidos en el cuadrado unidad, se puede calcular la función de densidad de la cópula empírica dado un vector h usando la siguiente ecuación:

$$g_{i,j} = c^* \left(\frac{2i-1}{2k}, \frac{2j-1}{2k} \right) \quad (5)$$

$$= \frac{k^2}{|S(h)|}; (u, v) \in S(h) | \frac{i-1}{k} < u < \frac{i}{k} \quad \frac{j-1}{k} < v < \frac{j}{k} \quad (6)$$

con $|S(h)|$ el número de parejas que tienen como vector de separación h .

2. Cópula Gaussiana

La cópula gaussiana toma la forma:

$$\begin{aligned} C(u_1, u_2; \theta) &= \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left(\frac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)} \right) ds dt \end{aligned}$$

donde Φ es la función de distribución normal estándar y $\Phi_G(u_1, u_2)$ es la distribución normal bivariada con parámetro de correlación θ restringido al intervalo $(-1, 1)$.

La cópula normal permite por igual, grados de dependencia positiva o negativa, es por esto que es una de las más utilizadas. Pero a pesar de esto, esta cópula es simétrica y en muchos casos, los datos reales no cumplen esta propiedad, para solucionar este problema utilizamos una nueva familia de cópulas que permiten asimetría en los datos y se presenta en la siguiente sección.

3. Familia de cópulas no gaussianas

En este capítulo se presenta una familia de cópulas asimétricas multivariadas no gaussianas, debido a la necesidad que surge en geoestadística para solucionar la asimetría por naturaleza de este tipo de datos, la cópula gaussiana no sólo expresa simetría, sino también dependencia de simetría radial, tal que los cuantiles altos y bajos de la distribución tienen propiedades iguales dependencia. Este supuesto pocas veces se cumple con datos reales.

Una cópula bivariada expresa dependencia simétrica si:

$$C(u_1, u_2) = C(1 - u_1, 1 - u_2) - 1 + u_1 + u_2, \quad (7)$$

y para la densidad:

$$c(u_1, u_2) = c(1 - u_1, 1 - u_2), \quad (8)$$

la simetría se entiende con respecto al eje menor $u_2 = 1 - u_1$ del cuadrado unidad.

La construcción de una familia de cópulas no es trivial, actualmente existen numerosas cópulas pero no cumplen las condiciones necesarias para ser utilizadas en geoestadística y en otros casos la construcción es imposible debido a problemas computacionales.

Bárdossy y Li (Bárdossy 2008) propone una familia de distribuciones que se obtienen a través de una transformación no-monotónica de la cópula gaussiana multivariante, llamada cópula V-transformada. En este trabajo se propone dos extensiones de esta metodología: La primera es la inclusión de tendencia y la segunda es el método de indicador kriging e interpolación usando cópulas que se presentara mas adelante.

Definición 1. Sea $Y \sim N(0, \Gamma)$ una variable aleatoria n -dimensional con media $0^T = (0, \dots, 0)$ y matriz de correlación Γ . Todas las marginales se suponen con varianza unitaria. Sea $\mathbf{X}$ definida para cada coordenada $j = 1, \dots, n$ de tal forma:

$$\begin{cases} k(Y_j - m)^\alpha & Y_j \geq m, \\ m - Y_j & Y_j < m, \end{cases}$$

donde k es una constante positiva y α y m números reales arbitrarios, a manera de ejemplo, en la Figura (1), se presenta algunas transformaciones, se puede observar que la cópula recibe este nombre debido a la forma de V en la gráfica. La consecuencia más importante es que para valores menores que m , los cuales son los valores menores en el espacio original, se convierten, en el nuevo espacio en los valores mayores; tal que el valor m se convierte en el valor más pequeño en el nuevo espacio, en cuanto a los valores mayores que m en el espacio original, la transformación produce un concentramiento o una división dependiendo de la configuración de k y α . Este efecto produce que la dependencia asociada con los valores cercanos o iguales a m influya en la dependencia de los valores bajos en el nuevo espacio. (Jing 2010).

El efecto de la transformación también puede explicarse por el cambio en la distribución marginal. Figura (2) muestra cómo la densidad en la distribución cambia después de la transformación. Se puede notar las siguientes características:

- Por lo general, los datos se dispersan después de la transformación,
- Si los valores de k y α son invariables, pero el valor de m incrementa, la densidad está más concentrada a la mediana y la distribución es más simétrica.

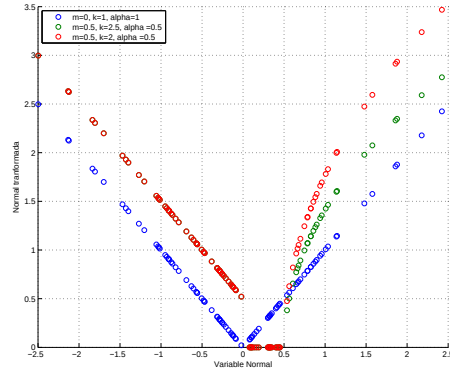
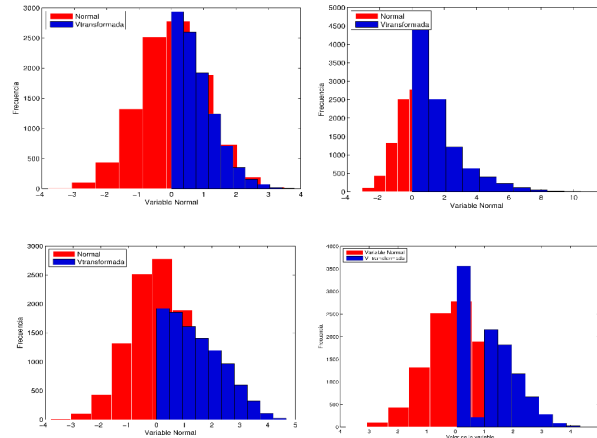


FIGURA 1: Transformaciones V-Normal

- Si $k = 1$, $\alpha = 1$ la transformada V-normal se aproxima a la distribución χ^2 con un grado de libertad.

FIGURA 2: Transformaciones V-Normal con parámetros: $m = 0$, $k = 1$ y $\alpha = 1$ (Esquina superior izquierda), $m = 0$, $k = 3$ y $\alpha = 1$ (Esquina superior derecha), $m = 0$, $k = 2.5$ y $\alpha = 0.5$ (Esquina inferior izquierda) y $m = 1$, $k = 2.5$ y $\alpha = 0.5$ (Esquina inferior derecha)

La función de distribución marginal de X es:

$$\begin{aligned}
 H(x) &= P(X \leq x) \\
 &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + P(Y > x - m) \\
 &= \Phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - \Phi(-x + m)
 \end{aligned}$$

y la función de densidad:

$$h(x) = \frac{1}{k\alpha} \left(\frac{x}{k}\right)^{\frac{1}{\alpha}-1} \phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + \phi(-x + m)$$

tal que $\Phi(\cdot)$ y $\phi(\cdot)$ son las funciones de distribución y densidad de la normal estándar, respectivamente.

Para el caso particular bivariado, se tiene la función de distribución:

$$\begin{aligned}
 V(x_1, x_2) &= P(X_1 < x_1, X_2 < x_2) \\
 &= P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &- P\left(Y_1 < -x_1 + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) - P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < -x_2 + m\right) \\
 &+ P(Y_1 < x_1 + m, Y_2 > -x_2 - m) \\
 &= \Phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &- \Phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, -x_2 + m\right) - \Phi\left(-x_1 + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) + \Phi(-x_1 + m, -x_2 + m)
 \end{aligned}$$

y la función de densidad:

$$\begin{aligned}
 v(x_1, x_2) &= \frac{1}{k^2 \alpha^2} \left(\frac{x_1 x_2}{k^2}\right)^{\frac{1}{\alpha}-1} \phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &+ \frac{1}{k \alpha} \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}-1} \phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, -x_2 + m\right) \\
 &+ \frac{1}{k \alpha} \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}-1} \phi\left(-x_1 + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &+ \phi(-x_1 + m, -x_2 + m)
 \end{aligned}$$

La Figura 3 se muestra la densidad de algunas cópulas bivariadas, en todos los casos $\rho = 0.8$ pero m va aumentando, tal que $m = 0, 1.3, 15$ y 50 . Se puede notar que las cópulas son asimétricas y similares a los resultados de la distribución empírica obtenidas en la (??). Además, cada vez que incrementa m la distribución es mas simétrica, así que se puede concluir que si $m \rightarrow \infty$ la cópula converge a la cópula gaussiana. Para $m = 0, k = 1$ y $\alpha = 1$ la cópula que se genera es la cópula χ^2 , en la figura se ubica en la esquina superior derecha.

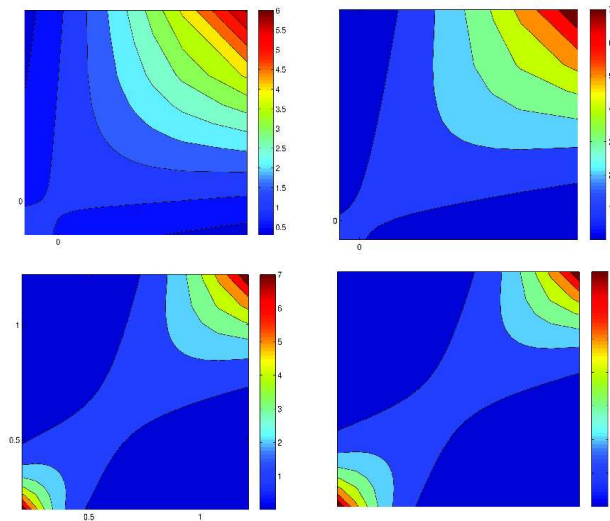


FIGURA 3: Cópula V -transformada Normal

4. RESULTADOS

Con base en los resultados teóricos propuestos en esta tesis, en este capítulo se aplica a un conjunto de datos llamados Gomel la teoría de cópulas en geoestadística. Los datos poseen este nombre ya que pertenecen a un conjunto de datos de la región de Gomel, Bielorrusia.

En abril de 1986, ocurrió el peor accidente nuclear en Chernóbil en la ex USSR (ahora Ucrania). La central nuclear de Chernóbil, situada a 100 kilómetros al norte de Kiev, tenía 4 reactores. En un día de abril a las 1:23 am la reacción en cadena en un reactor perdió el control, la creación de explosiones y una bola de fuego, que voló el acero pesado del reactor y la tapa de concreto. El desastre destruyó el reactor Chernóbil-4 y mató a 30 personas, entre ellas 28 por exposición a la radiación, 209 más fueron tratados por envenenamiento agudo por radiación y entre estos, 134 casos fueron confirmados. Grandes áreas de Bielorrusia, Ucrania, Rusia y más allá estaban contaminados en diversos grados. El desastre de Chernobyl fue un evento único y el único accidente en la historia de la energía nuclear comercial. Ahora 26 años después del accidente todavía más de 3 millones de niños sufren de estos efectos, la zona alrededor del reactor todavía está muy contaminada, la naturaleza está muerta y no hay vida silvestre.

4.1. Catástrofe en Chernóbil

la región de Gomel el gobierno ruso construyó una red donde se estudia la concentración de la cantidad de radiactividad. El conjunto de datos que utilizamos son mediciones de Cs137, un isótopo radiactivo.

Se analiza el conjunto de datos que corresponde a 148 localizaciones $x_i = (x_{1i}, x_{2i})^T$, $i = 1, \dots, 148$ en la región de Gomel. Los datos son observados diez años después del accidente de Chernobyl. En la Figura (4) se muestra las realizaciones donde se encontró el isótopo radiactivo (cruces rojas), se puede ver que la mayoría de los valores son pequeños, sin embargo, en la parte noreste, noroeste y sur de la región algunos valores relativamente grandes se producen.

El análisis de datos en copulas bivariadas permite realizar predicción, en la Figura 4 se muestra las localizaciones observadas marcadas con una x roja, y los puntos azules la grilla de interpolación, donde se realizarán las respectivas predicciones.

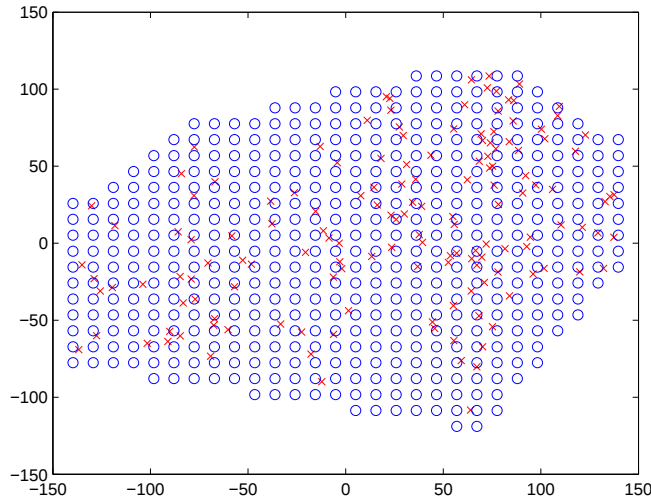


FIGURA 4: Observaciones y datos interpolados de los datos de Gomel

Para encontrar un marginal univariada apropiada, se prueba las distribuciones univariadas normal, gamma, el valor generalizado extremo (GEV) y la distribución de log-normal, para todas ellas se calcula las estimaciones de máxima verosimilitud, asumiendo que las observaciones son independientes. De esta forma, la elección que se toma será de la distribución marginal Box- cox con parámetro $\gamma = 0.0032$.

La selección de una cópula Gaussiana se puede justificar por medio del ajuste de bondad propuesto en el capítulo 5, es recomendable realizar un numero grande de simulaciones, pero en este caso es casi imposible ya que la complejidad computacional es muy alta, por tanto, según Genest (Genest 2008), se pueden realizar 150 simulaciones, los resultados de la prueba muestran que:

Resultado p -valor con 95 %			
h	0 – 10	10 – 20	90 – 100
T_n	0.33	0.999	0.99
S_n	0.99	0.99	0.999

TABLA 1

Los valores de la prueba de Kolmogorov–Smirnov parecen ser menores que los valores de p para el de Cramer–von Mises prueba. Esto es debido a que la prueba de Kolmogorov–Smirnov es insensible a valores extremos. Además, no existe un p -valor significativamente pequeño, por tanto, se puede asumir que el modelo propuesto se ajusta a los datos. A pesar de esto, se puede comparar en la Figura (5) la cópula empírica con la teórica, mostrando de manera gráfica que el modelo no puede ser simétrico.

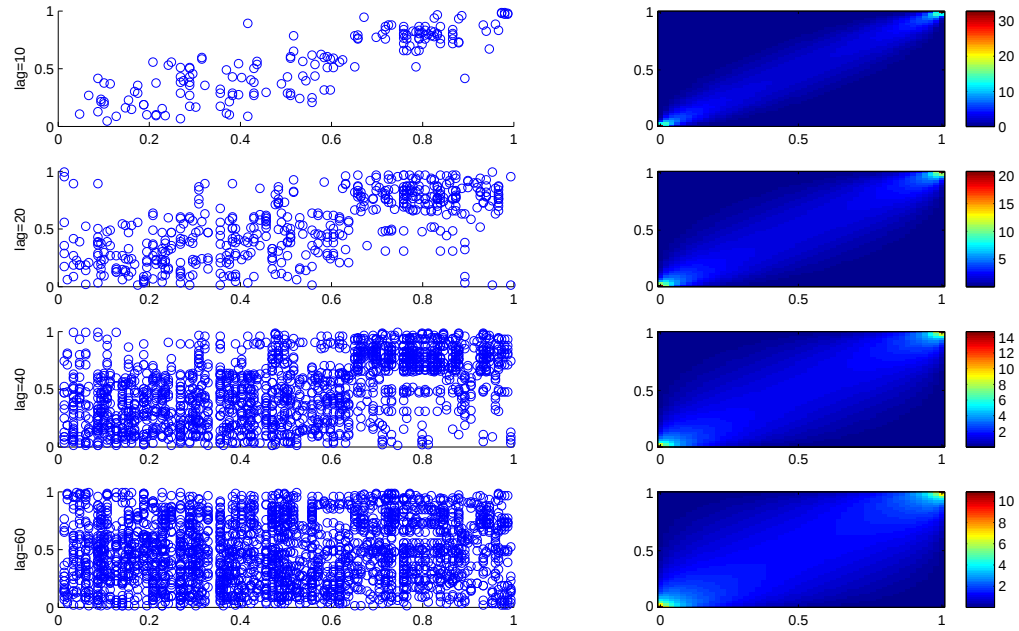


FIGURA 5: Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada Gaussiana (columna de la derecha)

Aun así, se realiza la estimación de los parámetros, se utiliza el modelo de correlación Matérn incluyendo un término efecto de pepita,

$$\rho(h) = \nu_2 I(h=0) + \frac{1-\nu_2}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{2\kappa^{1/2}h}{\nu_1} \right)^{\kappa} K_{\kappa} \left(\frac{2\kappa^{1/2}h}{\nu_1} \right) \quad (9)$$

tal que $\nu_1 \in [0, \infty]$ corresponde al parámetro de rango, $\nu_2 \in [0, 1]$ el parámetro del efecto pepita y κ es el parámetro de suavizamiento. Entonces se debe calcular cinco parámetros, los parámetros correspondientes a la transformación Box–cox, γ , y los parámetros de la función de correlación $\theta = (\nu_1, \nu_2, \kappa)$. Los resultados son:

γ	0,090
ν_1	61.92
ν_2	0.0539
κ	0.8650

TABLA 2: Estimación de parámetros para la cópula Gaussiana

Para la cópula no Gaussiana, se estudiará la cópula V –transformada, se selecciona la distribución marginal log Normal, con esto se pretende tomar en cuenta la propiedad fundamental de las cópula que permite a la estructura de dependencia no ser influenciada con las distribuciones marginales. De la misma forma que en la anterior sección, la selección de la cópula V se puede justificar por medio del ajuste de bondad propuesto en el capítulo 5, los resultados de la prueba muestran que:

Resultado p –valor con 95 %			
h	0 – 10	10 – 20	90 – 100
T_n	0.99	0.999	0.99
S_n	0.99	0.99	0.99

TABLA 3

Realizando una comparación con los p – valores de la cópula Gaussiana, se puede notar que son mas bajos que los de la cópula V – transformada. Por tanto, se puede asumir que el modelo propuesto se ajusta a los datos. Se puede comparar en la Figura (7) la cópula empírica con la teórica, mostrando un mejor ajuste que la cópula Gaussiana de la (5).

En este caso, se debe calcular 6 parámetros, los parámetros correspondientes a la marginal Log Normal μ y σ^2 , a la función de correlación, que en este caso se usara Matern $\theta = (\nu_1, \nu_2, \kappa)$ y el parámetro correspondiente a la cópula V , $m = 1, 27$, $k = 1$ y $\alpha = 1$. Los resultados son:

μ	0,595
σ	1,37
ν_1	100,023
ν_2	0,0576
κ	10

TABLA 4: Estimación de parámetros para la cópula V

En ausencia de datos de prueba que se utiliza para realizar una adecuada validación cruzada como un método cuantitativo para evaluar el desempeño del modelo, utilizamos el valor de MSE :

$$MSE = \frac{1}{2} \sum_{i=1}^n (\hat{Z}_{-i}(x_i) - z(x_i))^2 \quad (10)$$

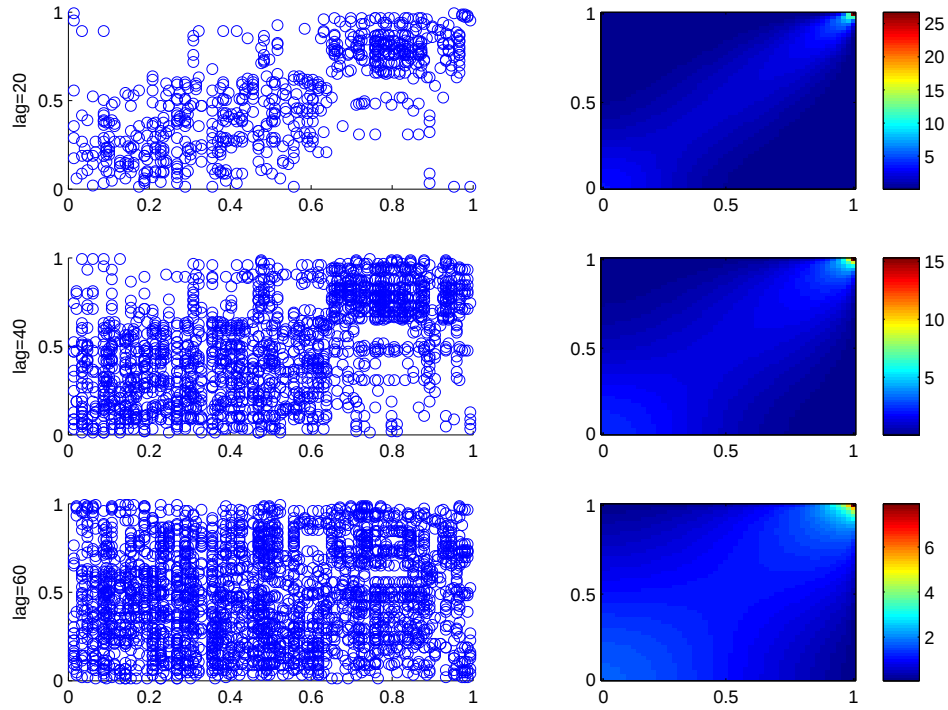


FIGURA 6: Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada χ^2 -cópula (columna de la derecha)

MSE	V -transformada	Gaussiana
	16.835	17.9995

TABLA 5: Valores de MSE para modelos de dependencia

y las respectivas predicciones para cada modelo se pueden observar en la Figura

El valor de: $MSE = 16,835$, resulta menor para la cópula no Gaussiana, esto demuestra un mejor desempeño del modelo, Además en la Figura se observa que los intervalos de confianza son mucho más cortos para los modelos basados en la cópula V - transformada, un hecho que también se refleja en las predicciones de las desviaciones estándar.

5. Conclusiones

Las cópulas permitieron generar una estructura de dependencia, para esto fue necesario realizar un análisis exploratorio de los datos con base en la cópula empírica, la cual permitió generar una idea de la forma y las propiedades que debe cumplir la cópula teórica, demostrando así la necesidad de formular una cópula asimétrica con propiedades adecuadas para el manejo geoestadístico.

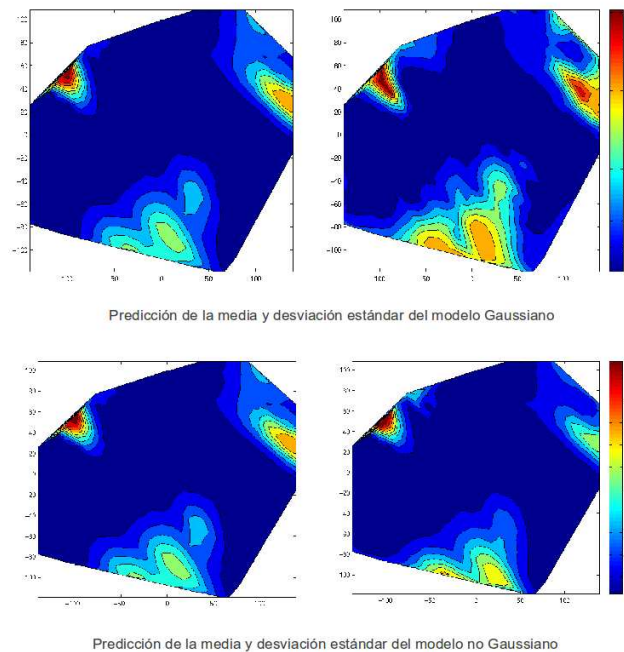


FIGURA 7: Datos Gomel: Predicción de la media y desviación estándar para modelos Gaussianos y no Gaussianos

Se desarrolla una metodología para la elección de las copulas, en este caso, la copula gaussiana y V probaron ser adecuadas para este tipo de datos y demostró que son adecuadas para el modelado espacial.

Después de generar la cópula empírica, se presentaron copulas teóricas Gaussiana y V transformada para la correcta estimación de parámetros por medio de máxima verosimilitud pero con algoritmos computacionales diferentes para cada tipo de cópula ya que la complejidad para la cópula V es 2^n .

La validación cruzada mostró que las copulas generan una completa estructura de dependencia, implicando que los resultados de predicción son más fiables que los métodos tradicionales, además de mostrar que las copulas asimétricas se comportan mejor para datos que cumplen estas características

Referencias

- Ayyad, C. Mateu, J. P. E. (2008), *Inferencia y modelización mediante cópulas*, Universidad Jaume.
- Bárdossy, A. Li, J. (2008), 'Geostatistical interpolation using copulas', *Water Resources Research* 44(7), W07412+.
[*http://dx.doi.org/10.1029/2007WR006115](http://dx.doi.org/10.1029/2007WR006115)
- Genest, C. Remillard, B. (2008), 'Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models', *Ann Inst Henri Poincaré* 44(6), 1096–1127.

- Haslauer, C. Li, J. B. A. (2010), ‘Application of Copulas in Geostatistics’, **16**, 395–404.
- Jing, L. (2010), Application of copulas as a new geostatistical tool, PhD thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart.
*<http://elib.uni-stuttgart.de/opus/volltexte/2010/5133>
- Kazianka, H. Pilz, J. (2010*a*), ‘Bayesian spatial modeling and interpolation using copulas’, *Computers & Geosciences* .
*<http://dx.doi.org/10.1016/j.cageo.2010.06.005>
- Kazianka, H. Pilz, J. (2010*b*), ‘Copula based geostatistical modeling of continuous and discrete data including covariates’, *Stochastic environmental research and risk assessment* **24**(5), 661–673.