UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
Diciembre 2013, Vol. 6, No. 2, pp. 139–156

Cópulas en geoestadística o lo que se puede hacer con coordenadas y estructuras de dependencia

Copulas in geostatistic or what can be done with coordinates and dependency structures

Danna Lesley Cruz Reyes[a]
dlcruzr@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 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 presenta 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. 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. Se presentan 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. Finalmente, se estudian las cópulas discretas y se aplican a un conjunto de datos simulados, esto permite realizar una extensión a los trabajos usuales de cópulas en geoestadística.

***Palabras clave*:** cópulas, geoestadística, estadística espacial, estadística computacional, tendencia.

[a]Investigadora Semillero IPREA. Departamento de Matematicas. Universidad Distrital Francisco José de Caldas. Colombia

Abstract

It is common in geostatistics to use methods such as the variogram or the correlation coefficient to describe spatial dependence, and kriging to make interpolation and predictions, but these methods are sensitive to extreme values and are strongly influenced by marginal distribution of the random field. Hence they can lead to unreliable results. As an alternative to traditional models in geostatistics are considered the use of the copula functions. Copula is widely used in the finance and actuary fields and due to satisfactory results they started to be considered in other areas of application of statistical sciences. This work shows the effect of copulas as a tool that presents a geostatistical analysis under the range of quantiles and a dependence structure, considering models of spatial tendency, continuous and discrete marginal distributions and covariance functions. Three interpolation methods are shown: the first is the kriging indicator and disjunctive kriging, the second method is known as the simple kriging and the third method is a plug-in prediction and the generalization of the trans-Gaussian kriging, these methods are used based on the copula function due to the existing relationship between bivariate copulas and covariance indicators. Results are presented for a set of actual data in the city of Gomel that contains measurements of radioactive isotopes, consequence of the Chernobyl nuclear accident. Finally, discrete copulas are studied and applied to a set of simulated data, this allows an extension of the usual works of copulas in Geostatistics.

***Keywords*:** copulas, geostatistics, spatial statistics, computational statistics, trend.

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, 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 (Ayyad et al. 2008).

En geoestadística se analizan las realizaciones de un campo aleatorio $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ donde $D \subset \mathbb{R}^n$, cuya realización, $z(\mathbf{s})$ representa el valor de interés registrado en la medición con respecto a cierto sistema de referencia $\mathbf{s}$.

En la actualidad existen herramientas para modelar la variabilidad espacial. La primera fue usada en principios de los cincuenta por Danie G. Krige, en Sudáfrica, para ampliar técnicas estadísticas para la estimación de las reservas de minerales (Bárdossy & Li 2008). 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 consideran realistas para los tipos de datos y se reportan datos atípicos que causan problemas en las investigaciones (Bárdossy & Li 2008).

Bárdossy en el año 2006 fue el pionero en proponer el uso de cópulas para describir variabilidad espacial, Bárdossy & Li (2008) realizan modelación de campos aleatorios continuos, Kazianka & Pilz (2010) adoptan la metodología de Bárdossy y realiza una extensión considerando modelos de tendencia y campos aleatorios discretos, Kazianka & Pilz (2011) muestran cómo se incorporan en un marco bayesiano mediante la asignación de probabilidades apriori de todos los parámetros del modelo, en este trabajo se propone una extensión a estos modelos incluyendo las cópulas radialmente asimétricas que resultan más eficientes que las hasta ahora usadas.

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 (Diggle & Ribeiro 2007). En este trabajo se presenta el análisis de estructura de dependencia a través de cópulas proponiendo una familia de cópulas radialmente asimétricas.

Bárdossy & Li (2008) proponen 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 proponen 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 presentarán más adelante.

El trabajo se organiza de la siguiente manera. La sección 2, describe cómo las cópulas serán implementas en los campos aleatorios, en la sección 3, se presenta un método de exploración de datos utilizando las funciones cópula. La sección 4 se refiere a la cópula gaussiana mientras que la Sección 5 se presenta la familia de las cópulas no gaussianas. Los resultados de la estimación y la cópula son presentados en la Sección 6, para analizar el conjunto de datos llamados Gomel, en la Sección 7 se presentan cópulas discretas y finalmente, en la última sección se presentan las conclusiones.

2. Descripción del campo aleatorio usando cópulas

Bardossy (Bárdossy & Li 2008) presentó un método diferente a los métodos clásicos mencionados en la introducción para el modelado de dependencia espacial por medio de cópulas, se pretende describir todas las distribuciones multivariadas necesarias del campo aleatorio por medio de cópulas.

Se asume un campo aleatorio estacionario $\{Z(x)|x \in D\}$, donde $D \in \mathbb{R}^2$ es el

área de interés. Se nota $\mathbf{h}$ 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 $\mathbf{x} \in D$.

Con el teorema de Sklar, se puede establecer un modelo multivariante del campo tomando $F_Z = F_1 = F_2 = \cdots = F_n$ tal que, su destitución conjunta $H(x_1, \ldots, x_n)$ de las variables $x_1, \ldots, x_n$,

$$H(x_1, \ldots, x_n) = C(F_1, F_2, \ldots, F_n), \tag{1}$$

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

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

Por tanto, la estructura de dependencia quedaría descrita por la función cópula $C_{\mathbf{h}}$ en función del vector $\mathbf{h}$, esto implica que la cópula 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 sean apropiadas para este modelo, de manera natural se puede suponer una cópula simétrica, debido a que la dependencia entre dos localizaciones x_1 y x_2 es la misma que x_2 y x_1, de manera general se tiene que:

$$C_h(u_1, \ldots, u_n) = C_h(u_{\pi(1)}, \ldots, u_{\pi(1)}), \tag{2}$$

para una permutación arbitraria π.

Además, se deben añadir las siguientes restricciones:

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

Donde, la $\Pi^n(x)$ representa la cópula de independencia y $M^n(x)$ la cópula mínima. 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.

3. Cópulas empíricas bivariadas

Las cópulas empíricas son usadas por primera vez en el área de geoestadística, por Haussler quien implementó las cópulas bivariadas empíricas, describiendo la estructura de dependencia entre variables aleatorias. En este artículo 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 et al. 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.

3.1. 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 son 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 et al. 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 $S(h)$:
$$S(h) = \{(F_Z(z(s_i)), F_Z(z(s_j)))||s_i - s_j| \approx h\} \tag{3}$$
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:
$$\begin{aligned} g_{i,j} &= c^* \left(\frac{2i-1}{2k}, \frac{2j-1}{2k} \right) \\ &= \frac{k^2}{|S(h)|}; (u,v) \in S(h)| \end{aligned}$$

donde $\frac{i-1}{k} < u < \frac{i}{k}$, $\frac{j-1}{k} < v < \frac{j}{k}$, y $|S(h)|$ denota el número de parejas que tienen como vector de separación h.

4. Cópula gaussiana

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 4 con marginal $F_Z = \Phi_{\mu,\sigma^2}$, donde μ y σ denota la media y varianza respectivamente, es muy utilizada en campos aleatorios,

$$C_{\Sigma}^{G} = \Phi_{0,\Sigma}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)). \tag{4}$$

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.

La cópula gaussiana toma la forma:

$$C(u_1, u_2; \theta) \quad = \quad \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2)),$$

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, y 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 se utiliza una nueva familia de cópulas que permiten asimetría en los datos y se presenta en la siguiente sección.

5. 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 solo expresa simetría, sino también dependencia de simetría radial, tal que los cuantiles altos y bajos de la distribución tienen propiedades iguales de dependencia. Este supuesto pocas veces se cumple con datos reales.

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. Por tanto, se presenta la siguiente definición de cópula V-transformada:

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

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

donde k es una constante positiva y α y m números reales arbitrarios, a manera de ejemplo, en la Figura 1, se presentan algunas transformaciones, se puede observar que la cópula recibe este nombre debido a la forma de V en la gráfica.

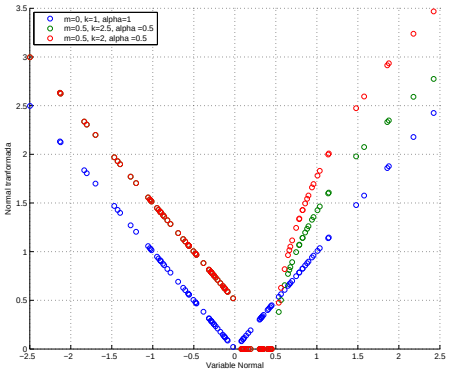

Figura 1: *Transformaciones V-normal. Fuente: elaboración propia.*

La función de distribución marginal de X es:

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

y la función de densidad:

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

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

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 para 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. La Figura 2 muestra cómo la densidad en la distribución cambia después de la transformación. Se pueden 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.
- Si $k = 1$, $\alpha = 1$ la transformada $V-$normal se aproxima a la distribución χ^2 con un grado de libertad.

La Figura 3 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. Además, cada vez que incrementa m la distribución es más simétrica, así que se puede concluir que si $m \to \infty$ la cópula converge a la cópula gaussiana (Jing 2010). 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.

6. Catástrofe en Chernóbil

En abril de 1986, ocurrió el peor accidente nuclear en Chernóbil en la antigua Unión Soviética (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 a.m. la reacción en cadena en un reactor perdió el control, la creación de explosiones y una bola de fuego voló el acero pesado del reactor y la tapa de concreto. El

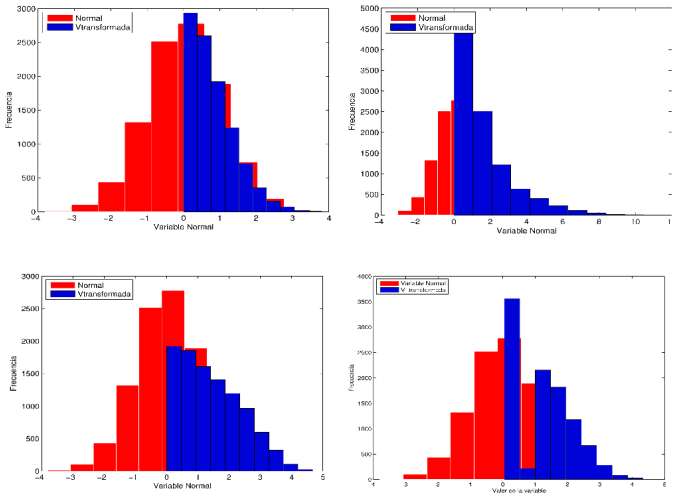

Figura 2: *Transformaciones V-Normal con parámetros:* $m = 0$, $k = 1$ *y* $\alpha = 1$ *(esquina superior izquierda),* $m = 0$, $k = 3$ *y* $\alpha = 1$ *(esquina superior derecha),* $m = 0$, $k = 2.5$ *y* $\alpha = 0.5$ *(esquina inferior izquierda) y* $m = 1$, $k = 2.5$ *y* $\alpha = 0.5$ *(esquina inferior izquierda). Fuente: elaboración propia.*

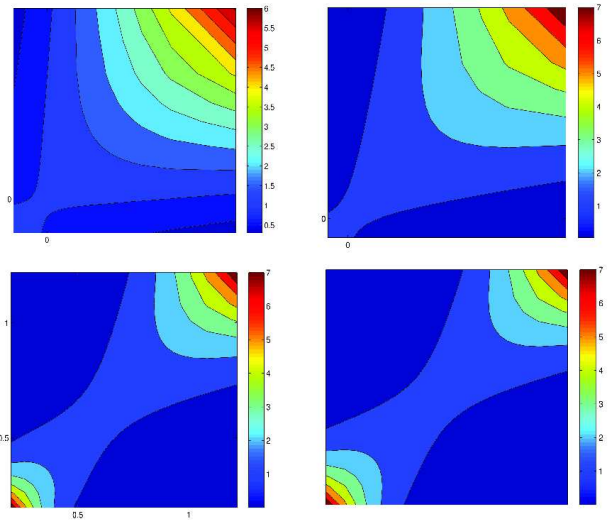

Figura 3: *Cópula V−transformada normal. Fuente: elaboración propia.*

desastre destruyó el reactor Chernóbil-4 y mató a 30 personas, entre ellas 28 por exposición a la radiación, 209 más fueron tratadas por envenenamiento agudo por radiación y entre estos, 134 casos fueron confirmados. Grandes áreas de Bielorrusia,

Ucrania, Rusia y más allá estaban contaminadas en diversos grados. El desastre de Chernóbil 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. En 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 se utiliza son mediciones de Cs137, un isótopo radiactivo.

Se analiza el conjunto de datos que corresponde a 148 localizaciones $x_i = (x_{1i}, x_{2i})^T$, $i = 1, \ldots, 148$ en la región de Gomel. Los datos son observados diez años después del accidente de Chernóbil. En la Figura 4 se muestran 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 cópulas bivariadas permite la predicción, en la Figura 4 se muestran las localizaciones observadas marcadas con una x roja, y los puntos azules la grilla de interpolación, donde se realizarán las respectivas predicciones.

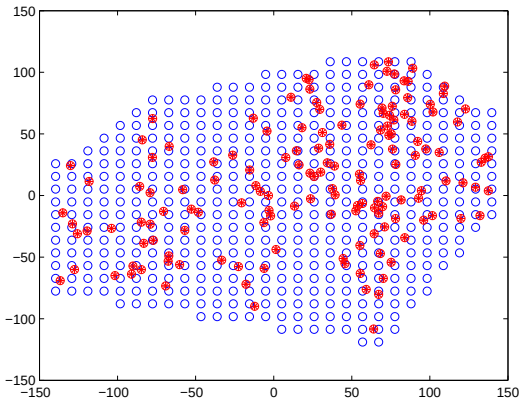

Figura 4: *Observaciones y datos interpolados de los datos de Gomel. Fuente: elaboración propia.*

Para encontrar una marginal univariada apropiada, se prueban las distribuciones univariadas normal, gamma, transformación box-cox, el valor generalizado extremo (GEV) y la distribución de log-normal, para todas ellas se calculan 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 que se presenta en Genest (Genest & Rémillard 2008), es recomendable realizar un número grande de simulaciones, pero en este caso es casi imposible, ya

que la complejidad computacional es muy alta, por tanto, según Genest (Genest & Rémillard 2008) se pueden realizar 150 simulaciones, los resultados de la prueba muestran que:

Tabla 1: *Valor p para el estadístico de elección de la cópula gaussiana. Fuente: elaboración propia*

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

Los valores de la prueba de Kolmogorov−Smirnov parecen ser menores que los valores de p para el de Cramer−von Mises prueba puesto que la prueba de Kolmogorov−Smirnov es insensible a valores extremos. Además, no existe un p− valor significantemente 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.

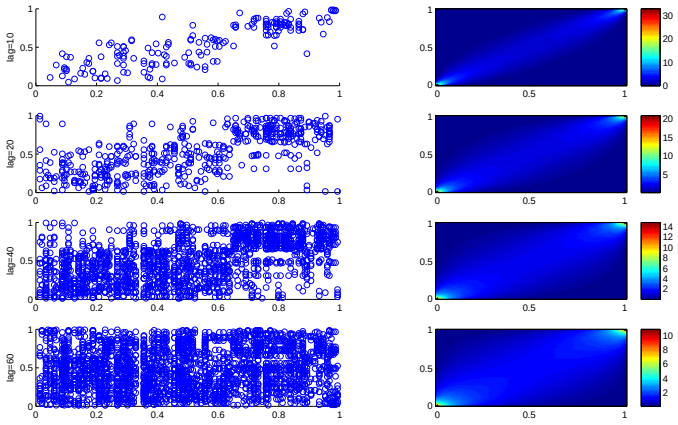

Figura 5: *Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada gaussiana (columna de la derecha). Fuente: elaboración propia.*

Aun así, se realiza la estimación de los parámetros, se utiliza el modelo de correlación Matérn incluyendo un término efecto de pepita, con los valores de $\nu_1 \in [0, \infty]$ que corresponde al parámetro de rango, $\nu_2 \in [0, 1]$ el parámetro del efecto pepita y κ es el parámetro de suavizamiento. Entonces, se deben 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: $\hat{\gamma} = 0,090$, $\hat{\nu}_1 = 61.92$, $\hat{\nu}_2 = 0.0539$ y $\hat{\kappa} = 0.8650$.

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 de Genest (Genest & Rémillard 2008), los resultados de la prueba muestran que:

Tabla 2: *Valor p para el estadístico de elección de la cópula V- Normal. Fuente: elaboración propia*

h	$0-10$	$10-20$	$90-100$
T_n	0.99	0.999	0.99
S_n	0.99	0.99	0.99

Al realizar una comparación con los valores p de la cópula gaussiana, se puede notar que son más 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 las Figura 7) y 8 la cópula empírica con la teórica, mostrando un mejor ajuste que la cópula gaussiana de la 5.

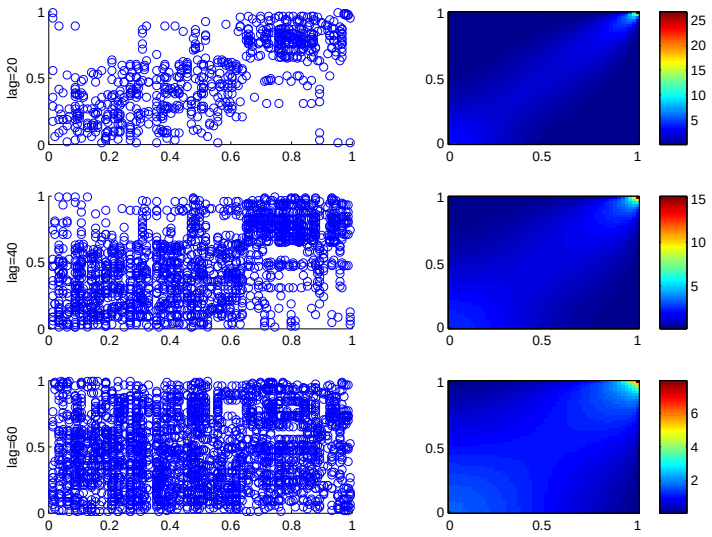

Figura 6: *Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada χ^2-cópula (columna de la derecha). Fuente: elaboración propia.*

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 usará 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: $\hat{\mu} = 0,595$, $\hat{\sigma} = 1,37$, $\hat{\nu}_1 = 100,023$, $\hat{\nu}_2 = 0,0576$ y $\hat{\kappa} = 10$.

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, se utiliza el valor de MSE y las respectivas predicciones para cada modelo se pueden observar en la Figura 7.

Tabla 3: *Valores de MSE para modelos de dependencia. Fuente: elaboración propia*

$V-$transformada	gaussiana
16.835	17.9995

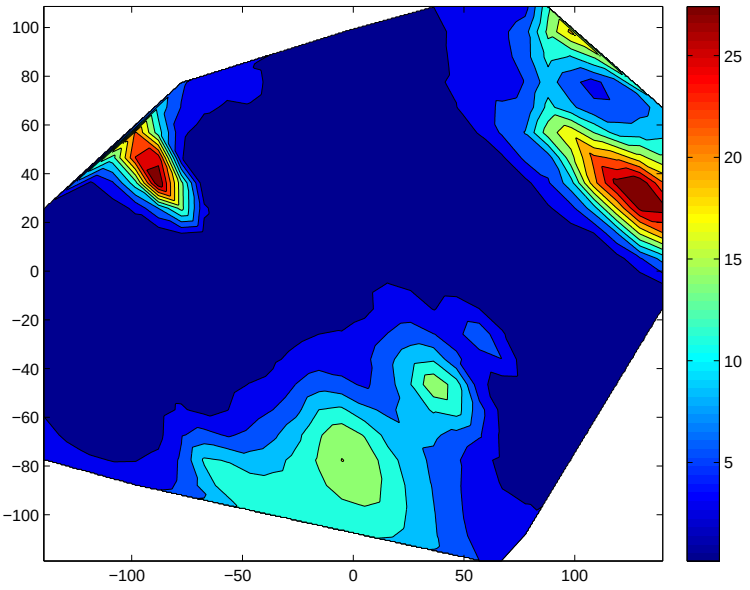

Figura 7: *Datos Gomel: Predicción de la media para modelos gaussianos. Fuente: elaboración propia.*

El valor de $MSE = 16,835$ resulta menor para la cópula no gaussiana, esto demuestra un mejor desempeño del modelo. Además en la Figura 8 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

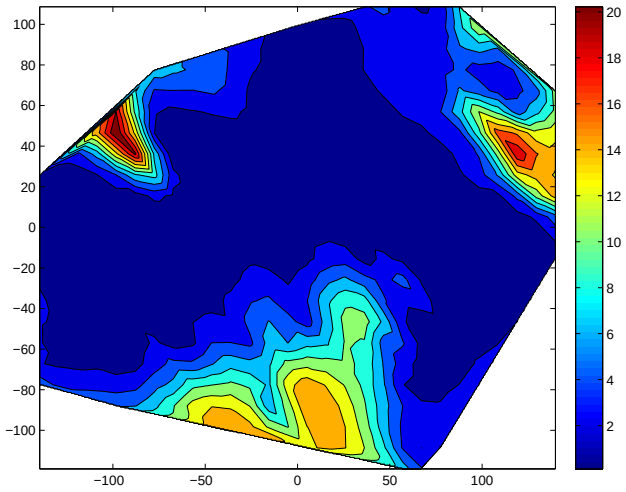

Figura 8: *Datos Gomel: Predicción de la desviación estándar para modelos gaussianos. Fuente: elaboración propia.*

las desviaciones estándar.

7. Cópula discreta

En las secciones anteriores se ha descrito la forma como se utilizan las cópulas para construir estructuras de dependencia, pero únicamente se consideran variables aleatorias con distribuciones marginales continuas y en algunos casos se puede presentar la necesidad de implementar otro tipo de cópulas.

En algunos casos geoestadísticos, se pueden tener variables aleatorias cuyas realizaciones pertenezcan al conjunto de los números naturales, de tal forma que, lo más conveniente es utilizar distribuciones discretas. En el marco de este artículo, donde el principal objetivo es utilizar cópulas para datos geoestadísticos y siguiendo el teorema de Sklar (1) el cual garantiza la existencia de la cópula para una función de distribución conjunta $H(X, Y)$ de las variables aleatorias X e Y, se propone una extensión a este teorema considerando X y Y variables aleatorias discretas. Kazianka (Kazianka & Pilz 2010) introduce cópulas en geoestadística para marginales discretas, con base en este último artículo se realiza este capítulo, sin embargo, se considera por una metodología diferente para realizar la inferencia y estimación de parámetros, ya que la complejidad computacional de Kazianka es alta.

7.1. Inferencia para cópulas con marginales discretas en geoestadística

Las cópulas con marginales discretas no difieren demasiado para el caso continuo, se deben tener en cuenta los anteriores resultados, pero se aplican de la misma forma.

La estructura de dependencia esta caracterizada por las familias de cópulas paramétricas, por ejemplo, la cópula gaussiana definida como:

$$C_{\mu,\Sigma}(u,v) = \Phi_{\Sigma}(\phi^{-1}(u), \phi^{-1}(v)) \tag{5}$$

Por tanto, es posible realizar una estimación de máxima verosimilitud, considerando un modelo generativo donde los marginales uniformes se generan a partir de la densidad de la cópula, y a su vez, se utilizan para generar las variables discretas con el uso de la distribución inversa de la marginal de las funciones de distribución. Esta marginal puede ser de cualquier familia parametrizada de distribuciones univariantes discretas.

A manera de ejemplo, se utiliza la simulación realizada por Diggle (Diggle & Ribeiro 2007) con respuestas Poisson y un proceso gaussiano en el programa R. Se consideró un proceso gaussiano para simular las localizaciones con $\mu = 0.5$m $\sigma^2 = 3$ y función de correlación Matérn con $\kappa = 1.5$ y $\phi = 0.2$. Las realizaciones son exponenciales con media Poisson, $\mu_i = exp(0.5 + z(x_i))$. EL resultado y el computo de la cópula se realiza en MATLAB. En la Figura 9 se muestra la simulación realizada.

De la misma forma que procedió anteriormente, se realiza una exploración de datos de la estructura de dependencia utilizando la cópula empírica, en la Figura 10 se muestra la estructura de dependencia para $lags = 0.15, 0.19, 0.198, 0.21$, se puede notar una clara simetría en los puntos lo que puede indicar una cópula gaussiana. Esta cópula se presenta en la Figura 11.

Los resultados muestran que es posible construir una estructura de dependencia para cópulas con marginales discretas, para el caso bivariado y que además resulta flexible para la estructura de dependencia, sin embargo, matemáticamente, la generalización de este método no es una tarea trivial debido a la complejidad computacional que se puede presentar para un caso multivariado ($n > 2$).

8. Conclusión

Si la estructura de dependencia entre una población es relativamente homogénea, entonces, las cópulas pueden ser útiles, en el sentido de que se puede estimar a partir de una muestra mucho menor que la necesaria, por ejemplo, para una matriz de covarianza completa. Por otra parte, si las dependencias dentro de una población varían notablemente para diferentes pares de datos, la cópula gaussiana carece de la flexibilidad para capturar las dependencias extremas. En tales casos,

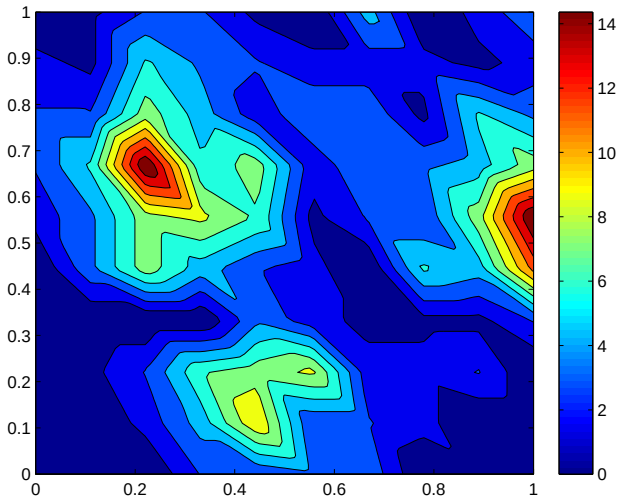

Figura 9: *Simulación proceso gaussiano con marginales Poisson. Fuente: elaboración propia.*

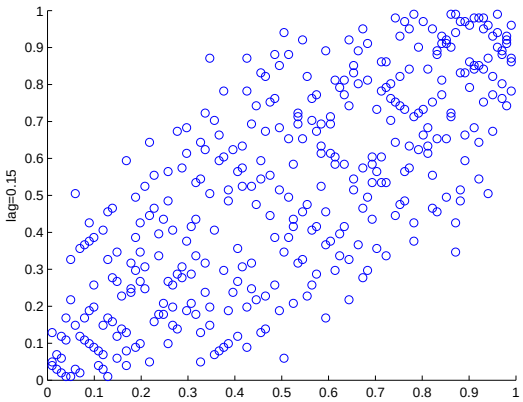

Figura 10: *Cópula empírica para simulación proceso gaussiano con marginales Poisson,* $h = 0.15$. *Fuente: elaboración propia.*

se puede aplicar otra cópula de la familia elíptica, ya que está parametrizada por la misma matriz de covarianza de la cópula gaussiana. Sin embargo, la cópula gaussiana se prohíbe para dimensiones altas, ya que la evaluación de la probabilidad requiere un número exponencial de evaluaciones de la gaussiana multivariada, que se debe calcular numéricamente convirtiendo el análisis en una labor imposible.

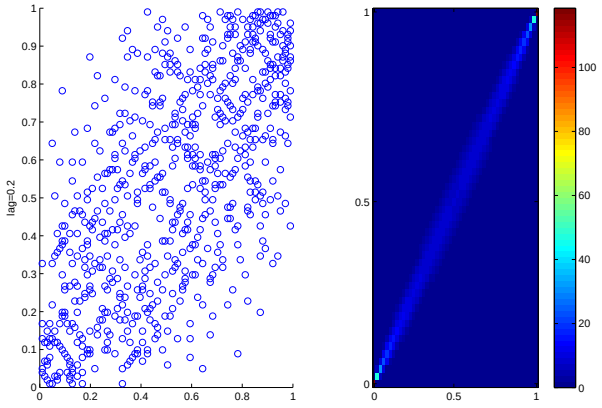

Figura 11: *Cópula gaussiana con marginales Poisson. Fuente: elaboración propia.*

Este artículo es producto de la tesis de maestría dirigida por el profesor Edilberto Cepeda de la Universidad Nacional de Colombia.

Recibido: 04 de junio de 2013
Aceptado: 20 de septiembre de 2013

Referencias

Ayyad, C., Mateu, J. & Porcu, E. (2008), *Inferencia y modelización mediante cópulas*, Universidad Jaume.

Bárdossy, A. & Li, J. (2008), 'Geostatistical interpolation using copulas', *Water Resources Research* **44**(7).

Diggle, P. & Ribeiro, P. (2007), *Model-based Geostatistics*, Springer Series in Statistics, Springer.

Genest, C. & Rémillard, B. (2008), 'Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models', *Annales de I'institut Henri Poincaré (B) Probabilités et Statistiques* **44**(6), 1096–1127.

Haslauer, C., Li, J. & Bárdossy, A. (2010), 'Application of copulas in geostatistics', *geoENV VII Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics* **16**, 395–404.

Jing, L. (2010), *Application of copulas as a new geostatistical tool*, PhD thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart.

Kazianka, H. & Pilz, J. (2010), 'Copula based geostatistical modeling of continuous and discrete data including covariates', *Stochastic environmental research and risk assessment* **24**(5), 661–673.

Kazianka, H. & Pilz, J. (2011), 'Bayesian spatial modeling and interpolation using copulas', *Computers & Geosciences* **37**(3), 310–319.