

# Regresión cuantil funcional basada en cópulas

Lic. Anselmo Daniel Suarez Muñoz, Ing. Luciano Andres Juárez López  
Centro de Investigación en Matemáticas. Unidad Monterrey  
anselmo.suarez@cimat.mx, luciano.juarez@cimat.mx

## I. INTRODUCCIÓN

En la actualidad, el fenómeno de las islas de calor urbanas representa un desafío significativo para las ciudades, ya que afecta tanto a la salud pública como a la sostenibilidad ambiental. En este contexto, resulta crucial desarrollar herramientas que permitan pronosticar la *temperatura de la superficie terrestre* (*Land Surface Temperature*, LST), un indicador que mide el calor emitido por la superficie terrestre. Este se diferencia de la temperatura del aire, informada en los reportes meteorológicos, ya que la tierra se calienta y enfría más rápidamente que la atmósfera.

Este proyecto propone un modelo de pronóstico del LST basado en *regresión cuantil funcional con cópulas*, utilizando datos históricos del LST y del *Índice de Vegetación de Diferencia Normalizada* (*Normalized Difference Vegetation Index*, NDVI) correspondientes a la zona metropolitana de Monterrey entre los años 2018 y 2023. El NDVI, es uno de los índices de vegetación más utilizados, se calcula a partir de imágenes satelitales y describe la relación entre la luz roja y el infrarrojo cercano reflejados por la superficie terrestre.

A continuación, empezamos a definir la teoría que se necesita para realizar este proyecto.

## II. MARCO TEÓRICO.

Para poder introducir formalmente el concepto de una **cópula**, es necesario primero comprender qué es una función de distribución acumulada (FDA), ya que las cópulas se derivan de estas funciones al modelar la dependencia entre variables aleatorias. Comenzamos definiendo la FDA y las distribuciones marginales.

**Definición de Función de Distribución Acumulada (FDA):** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un vector aleatorio. Denotamos con  $F$  la FDA conjunta del vector aleatorio y con  $F_j$  la FDA marginal de  $X_j$ , para  $j = 1, \dots, p$ . Estas funciones están definidas como:

$$F_j(x_j) = \mathbb{P}(X_j \leq x_j), \quad F(x_1, \dots, x_p) = \mathbb{P}(X_1 \leq x_1, \dots, X_p \leq x_p),$$

donde  $\forall \mathbf{x} = (x_1, \dots, x_p)^T \in \text{Ran}(\mathbf{X})$ .

En el caso de que  $X_j$  sea continua, podemos aplicar la **Transformación Integral de Probabilidad (PIT)**, que permite transformar cada variable aleatoria  $X_j$  mediante su FDA marginal  $F_j$ , obteniendo una nueva variable:

$$U_j = F_j(X_j) \sim \mathcal{U}(0, 1), \quad j = 1, \dots, p.$$

Esto genera un vector transformado  $\mathbf{U} = (U_1, \dots, U_p)^T$ , en el que todas las variables tienen distribuciones marginales uniformes en el intervalo  $[0, 1]$ . Este resultado es fundamental para introducir el concepto de cópula.

Una **cópula** se define como una FDA en el hipercubo unitario  $[0, 1]^p$ , cuyas distribuciones marginales son uniformes en  $[0, 1]$ . Intuitivamente, la cópula describe únicamente la estructura de dependencia entre las variables, separándola de las características propias de las distribuciones marginales. Formalmente, el concepto de cópula se establece mediante el **Teorema de Sklar**.

**Teorema de Sklar:** Sea  $F$  una FDA  $p$ -dimensional con distribuciones marginales unidimensionales  $F_1, \dots, F_p$ . Entonces, existe una cópula  $C$  tal que:

$$F(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)).$$

Además, si las variables  $X_1, \dots, X_p$  son continuas, la cópula  $C$  es única. Este teorema establece que cualquier FDA conjunta puede descomponerse en sus distribuciones marginales y una cópula, lo que permite modelar la dependencia entre las variables de forma independiente de sus características individuales.

Una vez que hemos dado el concepto de cópulas, podemos extender su uso a estructuras más complejas que permitan modelar relaciones jerárquicas de dependencia entre múltiples variables. Estas estructuras se denominan **cópulas D-vine**, un caso especial dentro de las cópulas tipo vine, y son herramientas fundamentales para descomponer dependencias multivariadas en términos de relaciones bivariadas más manejables. A continuación damos la definición formal de una cópula D-vine.

**Definición de Cópula:** Una cópula  $d$ -dimensional  $C$  es una función de distribución de  $d$  variables en el hipercubo unitario  $[0, 1]^d$ , con distribuciones marginales uniformes. De acuerdo con el **Teorema de Sklar**, existe una cópula  $C$  asociada a un vector aleatorio multivariado  $\mathbf{X} = (X_1, \dots, X_d)^T$ , tal que:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

donde  $F_1, \dots, F_d$  son las distribuciones marginales y  $C$  captura la estructura de dependencia.

Cuando el vector  $\mathbf{X}$  es absolutamente continuo, la densidad conjunta  $f(x_1, \dots, x_d)$  puede descomponerse de manera similar, utilizando las densidades marginales  $f_1, \dots, f_d$  y la densidad de la cópula  $c(u_1, \dots, u_d)$ :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d),$$

donde

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \cdots \partial u_d}$$

es la densidad de la cópula.

Las siguientes funciones definidas en  $I^2 = [0, 1] \times [0, 1]$  y que tienen como rango el intervalo  $I = [0, 1]$ , son conocidas como cópulas fundamentales debido a sus propiedades y características particulares:

### 1. Cópula de Fréchet-Hoeffding Superior:

$$M(u, v) = \min(u, v).$$

Esta cópula describe el caso de dependencia máxima positiva entre las variables. Es decir,  $u$  y  $v$  están perfectamente relacionadas, de modo que valores altos de  $u$  corresponden a valores altos de  $v$  y viceversa. Es útil como límite superior en el conjunto de todas las cópulas posibles.

### 2. Cópula de Fréchet-Hoeffding Inferior:

$$W(u, v) = \max(u + v - 1, 0).$$

Esta cópula representa el caso de dependencia mínima o negativa máxima permisible entre las variables. En este caso, los valores altos de  $u$  corresponden a valores bajos de  $v$ , y viceversa. Sirve como límite inferior en el conjunto de todas las cópulas posibles.

### 3. Cópula Independiente:

$$\Pi(u, v) = u \cdot v.$$

Esta cópula modela el caso en que  $u$  y  $v$  son independientes, es decir, no existe ninguna dependencia entre las variables. Su estructura es simple y se utiliza como referencia en muchos análisis.

Cada una de estas cópulas es fundamental en el estudio de la teoría de cópulas, ya que define extremos teóricos en la estructura de dependencia entre variables. La cópula  $M(u, v)$  establece el límite superior para la dependencia positiva,  $W(u, v)$  establece el límite inferior para la dependencia negativa, y  $\Pi(u, v)$  define la independencia total.

Un **diagrama de contornos** para una cópula bivariada es una representación gráfica de sus conjuntos de nivel, definidos como:

$$\{(u_1, u_2) \in [0, 1]^2 : C(u_1, u_2) = a\},$$

donde  $a$  es una constante en  $[0, 1]$ . En este contexto,  $C(u_1, u_2)$  denota la función de distribución acumulada conjunta, es decir, la cópula que captura la dependencia entre las variables uniformes  $U_1$  y  $U_2$ .

El diagrama de contornos permite visualizar cómo la cópula distribuye la dependencia entre las variables  $U_1$  y  $U_2$ . Cada línea de contorno representa un conjunto de puntos donde la cópula toma un valor constante  $a$ . Este tipo de gráfico es útil para identificar patrones en la dependencia, tales como:

- **Simetría:** La simetría del diagrama puede indicar que la relación entre las variables es uniforme en ambas direcciones.

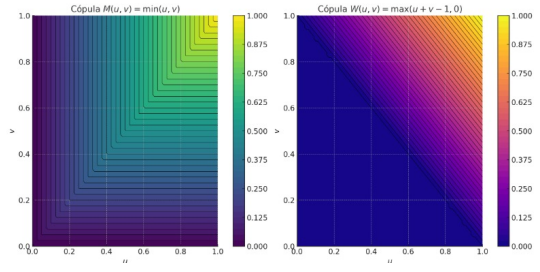


Figura 1: Diagrama de contornos para las copulas de Fréchet-Hoeffding

- **Asimetría:** Una forma asimétrica sugiere que la dependencia no es uniforme y que una variable puede tener mayor influencia sobre la otra.
- **Concentración en las colas:** Los contornos cercanos a las esquinas del cuadrado  $[0, 1]^2$  muestran la presencia de dependencia en los extremos (colas) de las distribuciones marginales.

En resumen, los diagramas de contornos son herramientas visuales que ofrecen una interpretación geométrica de las propiedades de una cópula, permitiendo un análisis intuitivo de la estructura de dependencia entre las variables.

**Cópula Gaussiana Bivariada** Sea  $\Phi$  la función de distribución acumulada (FDA) estándar normal univariante, y  $\Phi_\rho$  la FDA normal bivariada con coeficiente de correlación  $\rho$ . La **cópula Gaussiana bivariada** se define como:

$$C_G(u_1, u_2; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)),$$

donde  $u_1, u_2 \in [0, 1]$ , y  $\Phi^{-1}$  es la inversa de la FDA normal estándar.

Es simétrica respecto a  $u_1$  y  $u_2$ , es decir,  $C_G(u_1, u_2; \rho) = C_G(u_2, u_1; \rho)$ . Captura únicamente dependencia lineal, debido a la naturaleza de la distribución normal. Su aplicación se limita a casos donde las variables tienen una relación aproximadamente elíptica.

Sea  $t_\nu$  la FDA t-Student univariante con  $\nu$  grados de libertad, y  $t_{\nu, \rho}$  la FDA t-Student bivariada con  $\nu$  grados de libertad y coeficiente de correlación  $\rho$ . La **cópula t-Student bivariada** se define como:

$$C_t(u_1, u_2; \nu, \rho) = t_{\nu, \rho}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)),$$

donde  $u_1, u_2 \in [0, 1]$ , y  $t_\nu^{-1}$  es la inversa de la FDA t-Student univariante.

Es más flexible que la cópula Gaussiana para capturar dependencia en los extremos de las distribuciones (colas pesadas). Puede modelar mejor la dependencia no lineal y los eventos extremos conjuntos. Incluye la cópula Gaussiana como caso límite cuando  $\nu \rightarrow \infty$ .

- **Dependencia en las colas:** La cópula t-Student puede capturar dependencia en las colas (tanto superior como inferior), mientras que la cópula Gaussiana no.
- **Parámetros:** La cópula Gaussiana tiene un único parámetro ( $\rho$ ), mientras que la cópula t-Student requiere  $\nu$  (grados de libertad) además de  $\rho$ .

- **Aplicaciones:** La cópula Gaussiana es más común en problemas donde la relación es aproximadamente lineal y no hay eventos extremos significativos. La cópula t-Student es preferida en problemas financieros y de riesgos extremos.

La concordancia y la dependencia son conceptos fundamentales en el análisis de la relación entre dos variables aleatorias  $X_1$  y  $X_2$ . Estas se cuantifican a través de medidas derivadas de la cópula  $C(u, v)$  que las une.

La **concordancia** mide cómo las variables  $X_1$  y  $X_2$  tienden a moverse en la misma dirección o en direcciones opuestas. Dos medidas comunes de concordancia son:

El **coeficiente de Spearman**:

$$\rho_S(X_1, X_2) = \frac{1}{2} \int_{[0,1]^2} [C(u, v) - uv] du dv.$$

Este coeficiente se interpreta como una medida de asociación basada en los rangos de  $X_1$  y  $X_2$ .

El **coeficiente de Kendall**:

$$\tau_K(X_1, X_2) = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1.$$

El coeficiente de Kendall captura la probabilidad de concordancia menos la probabilidad de discordancia entre pares de observaciones de  $X_1$  y  $X_2$ .

La **dependencia** mide qué tan alejadas están las variables  $X_1$  y  $X_2$  de ser independientes. Una medida típica de dependencia es:

El **coeficiente de dependencia de Schweizer-Wolff**:

$$\sigma_{SW}(X_1, X_2) = \sigma_{SW}(C) = \frac{1}{2} \int_{[0,1]^2} |C(u, v) - uv| du dv.$$

Este coeficiente se basa en la diferencia absoluta entre la cópula  $C(u, v)$  y la cópula de independencia  $uv$ . Un valor cercano a cero indica independencia, mientras que valores mayores reflejan una dependencia más fuerte.

Aunque ambos conceptos están relacionados, la concordancia se centra en la dirección de la relación (positiva o negativa), mientras que la dependencia captura la intensidad y estructura de la relación.

Por otro lado, otro concepto clave es el de **STtD (Empirical Conditional  $(1 - \alpha)$ -functional Skewness Total-to-Dispersion)** que es una medida diseñada para capturar el comportamiento asimétrico de las distribuciones condicionales de una variable respuesta  $Y$ , dado un conjunto de predictores  $\mathbf{X} = (X_1, \dots, X_p)$  y otras variables adicionales  $\mathbf{Z} = (Z_1, \dots, Z_r)$ . Esta métrica permite analizar cómo las distribuciones se sesgan en relación con sus bandas de cuantiles condicionales, lo que resulta especialmente útil en contextos funcionales y en situaciones donde las distribuciones no son simétricas.

Para formalizar esta idea, consideremos un nivel de significancia  $\alpha \in (0, 0.5]$ . Las bandas de cuantiles condicionales, que representan intervalos específicos entre los  $(1 - \alpha)$ -cuantiles empíricos, se definen de la siguiente manera:

La primera banda de cuantiles condicionales, denotada por  $CQB_n^{12}(Y|X = x, Z = z)$ , contiene los valores de  $Y$  en el rango comprendido entre el  $\alpha$ -cuantil condicional y el cuantil central (mediana) condicional. Esta banda se formaliza como:

$$\begin{aligned} CQB_n^{12}(Y|X = x, Z = z) &= \\ &= B(\{y \in \mathbb{R} \mid q_{n,\alpha}(x, z) \leq \text{STtD}_n(y) \leq q_{n,0.5}(x, z)\}). \end{aligned}$$

Por otro lado, la segunda banda de cuantiles condicionales, denotada por  $CQB_n^{23}(Y|X = x, Z = z)$ , representa los valores de  $Y$  entre el cuantil central y el  $(1 - \alpha)$ -cuantil condicional, y se define como:

$$\begin{aligned} CQB_n^{23}(Y|X = x, Z = z) &= \\ &= B(\{y \in \mathbb{R} \mid q_{n,0.5}(x, z) < \text{STtD}_n(y) \leq q_{n,1-\alpha}(x, z)\}). \end{aligned}$$

Para estas bandas de cuantiles, se consideran los límites inferiores y superiores. En el caso de  $CQB_n^{12}$ , los límites inferior y superior se denotan como  $\Psi_1^\alpha(t; x, z)$  y  $\Psi_2^\alpha(t; x, z)$ , respectivamente. Para  $CQB_n^{23}$ , los límites son  $\Psi_2^\alpha(t; x, z)$  y  $\Psi_3^\alpha(t; x, z)$ . Estos límites permiten capturar la amplitud y la variación dentro de cada banda condicional.

Para evaluar la dispersión en estas bandas, definimos las distancias entre los límites. La distancia entre el límite inferior y superior de la primera banda ( $CQB_n^{12}$ ) se denota por  $\Delta_{21}^{1-\alpha}(x, z)$  y se calcula como:

$$\Delta_{21}^{1-\alpha}(x, z) := \sqrt{\sum_{t \in \mathcal{T}} (\Psi_2^\alpha(t; x, z) - \Psi_1^\alpha(t; x, z))^2}.$$

De manera similar, la distancia en la segunda banda ( $CQB_n^{23}$ ) se denota como  $\Delta_{32}^{1-\alpha}(x, z)$  y está dada por:

$$\Delta_{32}^{1-\alpha}(x, z) := \sqrt{\sum_{t \in \mathcal{T}} (\Psi_3^\alpha(t; x, z) - \Psi_2^\alpha(t; x, z))^2}.$$

A partir de estas distancias, podemos definir los ratios que cuantifican la contribución relativa de cada banda al total de la dispersión condicional. Estos se denotan como  $\Psi_{21}^{1-\alpha}(x, z)$  y  $\Psi_{32}^{1-\alpha}(x, z)$ , y se calculan como:

$$\Psi_{21}^{1-\alpha}(x, z) := \frac{\Delta_{21}^{1-\alpha}(x, z)}{\Delta_{21}^{1-\alpha}(x, z) + \Delta_{32}^{1-\alpha}(x, z)},$$

$$\Psi_{32}^{1-\alpha}(x, z) := \frac{\Delta_{32}^{1-\alpha}(x, z)}{\Delta_{21}^{1-\alpha}(x, z) + \Delta_{32}^{1-\alpha}(x, z)}.$$

Finalmente, el **STtD** se define como una medida del sesgo funcional condicional de  $Y|\mathbf{X} = x, \mathbf{Z} = z$ . Este se expresa como:

$$\text{SKW}_n^{1-\alpha}(Y|X = x, Z = z) := (2\Psi_{32}^{1-\alpha}(x, z) - 1) \mathbf{1}\{\Psi_{32}^{1-\alpha}(x, z) \geq 0.5\}.$$

El STtD permite evaluar la asimetría condicional de  $Y$  de manera cuantitativa, identificando si la distribución condicional está sesgada hacia la derecha o la izquierda. Esto resulta útil en análisis funcionales y aplicaciones donde las relaciones condicionales no son simétricas, brindando una herramienta poderosa para describir la dependencia y la dispersión.

### III. METODOLOGÍA

#### Regresión Cuantil Funcional

Sea  $\mathbf{X} = (x_1, \dots, x_p)'$  un vector aleatorio de dimensión  $p$ , y  $Y$  una variable aleatoria de valores reales y sea  $\alpha \in (0, 1)$  y considerando  $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^p$ . Una regresión  $\alpha$ -cuantil multivariada se define como:

$$y = q_\alpha(\mathbf{x}), \quad (1)$$

la cual es una solución de la ecuación:

$$F_{Y|\mathbf{X}}(y|\mathbf{x}) = P(Y \leq y|\mathbf{X} = \mathbf{x}) = \alpha. \quad (2)$$

#### Regresión Cuantil Funcional Basada en Cópulas

Las **cópulas** son funciones que permiten modelar la dependencia entre variables aleatorias, separando las distribuciones marginales de su estructura de dependencia. Según el teorema de Sklar, cualquier distribución conjunta  $F_{Y,\mathbf{X}}$  puede representarse como:

$$F_{Y,\mathbf{X}}(y, \mathbf{x}) = C(F_Y(y), F_{\mathbf{X}}(\mathbf{x})),$$

donde:

- $C(u, v)$  es la cópula que modela la dependencia entre  $Y$  y  $\mathbf{X}$ .
- $F_Y(y)$  y  $F_{\mathbf{X}}(\mathbf{x})$  son las distribuciones marginales de  $Y$  y  $\mathbf{X}$ , respectivamente.

La función de distribución condicional  $F_{Y|\mathbf{X}}$  se puede expresar en términos de la cópula como:

$$F_{Y|\mathbf{X}}(y|\mathbf{x}) = \left. \frac{\partial C(F_Y(y), F_{\mathbf{X}}(\mathbf{x}))}{\partial u} \right|_{u=F_Y(y), v=F_{\mathbf{X}}(\mathbf{x})}.$$

Por lo tanto, el cuantil condicional  $q_\alpha(\mathbf{x})$  se puede calcular como:

$$q_\alpha(\mathbf{x}) = F_Y^{-1} \left( C_{V|U}^{-1}(\alpha | F_{\mathbf{X}}(\mathbf{x})) \right),$$

donde  $C_{V|U}^{-1}$  representa la inversa de la cópula condicional entre  $Y$  y  $\mathbf{X}$  dado  $\mathbf{x}$ .

### IV. PREPARACIÓN DE DATOS

Se cuenta con dos matrices principales de datos satelitales obtenidos para la zona metropolitana de Monterrey, que abarcan desde enero de 2018 hasta diciembre de 2023:

- **LST**: Representa los valores mensuales de la Temperatura de la Superficie Terrestre (Land Surface Temperature, LST), un indicador del calor emitido por la superficie terrestre.
- **NDVI**: Representa los valores mensuales del Índice de Vegetación de Diferencia Normalizada (Normalized Difference Vegetation Index, NDVI), un indicador de la actividad de la vegetación, calculado a partir de imágenes satelitales que miden la luz roja y el infrarrojo cercano reflejados por la superficie terrestre.

Ambas matrices tienen dimensiones  $72 \times 158$ , donde las filas corresponden a los meses (enero de 2018 a diciembre de 2023) y las columnas representan diferentes ubicaciones geográficas dentro de la región de estudio.

#### Análisis Exploratorio de las Series de Tiempo

Se generaron las series de tiempo de la Temperatura de la Superficie Terrestre (LST) para explorar patrones temporales en el conjunto de datos  $D_{LST}$ . Cada línea en la gráfica representa una ubicación geográfica distinta, mientras que el eje temporal abarca desde enero de 2018 hasta diciembre de 2023.

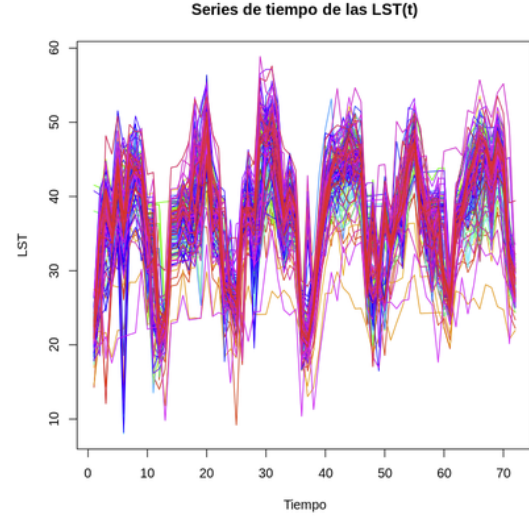


Figura 2: Series de tiempo de  $LST(t)$ .

La Figura 2 muestra un comportamiento estacional claro, con valores máximos en los meses cálidos (verano) y mínimos en los meses fríos (invierno). Además, hay diferencias entre las ubicaciones, reflejando variabilidad espacial en los valores de LST.

#### Construcción de Variables

Definiremos la matriz de diseño y la variable de respuesta del modelo de la siguiente manera:

- **Variable de respuesta:**

$$STtD(LST_T)$$

Representa los valores  $LST_T$  en un tiempo  $T$  generico, en nuestro caso  $T$  corresponde a los años 2023 y 2022.

- **Covariables:**

$$STtD(LST_{T-1})$$

$$STtD(NDVI_{T:T-4})$$

Donde:

- $STtD(LST_{T-1})$ : Valores de  $LST_{T-1}$  para los años 2022 y 2021.
- $STtD(NDVI_{T:T-4})$ : Combina los valores del  $NDVI_{T:T-4}$  de 2019 a 2023 y de 2018 a 2022.

El modelo queda definido como:

$$STtD(LST_T) \sim STtD(LST_{T-1}) + STtD(NDVI_{T:T-4})$$

Esta relación establece cómo la temperatura superficial terrestre en el tiempo  $T$  puede ser explicada por la misma variable en el periodo anterior ( $T - 1$ ) y por la vegetación rezagada ( $NDVI_{T:T-4}$ ).

#### Correlaciones entre Variable de Respuesta y Predictores

Tras la construcción del modelo, se analizaron las correlaciones:

- Correlación de Spearman entre  $STtD(LST_T)$  y  $STtD(LST_{T-1})$ : 0.5999532
- Correlación entre  $STtD(LST_T)$  y  $STtD(LST_{T-1})$ : 0.7002661
- Correlación de Spearman entre  $STtD(LST_T)$  y  $STtD(NDVI_{T:T-4})$ : -0.6584819
- Correlación entre  $STtD(LST_T)$  y  $STtD(NDVI_{T:T-4})$ : -0.6803726

#### Representación Gráfica de las Correlaciones

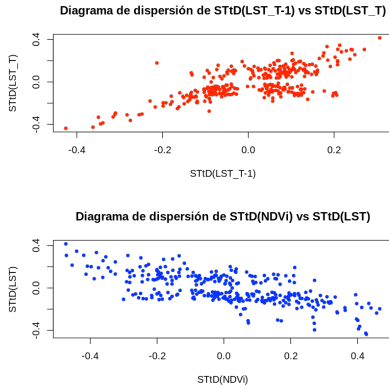


Figura 3: Diagrama de dispersión que muestra las correlaciones entre las variables del modelo.

Estos resultados indican una fuerte relación positiva entre  $STtD(LST_T)$  y  $STtD(LST_{T-1})$ , así como una relación negativa entre  $STtD(LST_T)$  y  $STtD(NDVI_{T:T-4})$ .

#### Relación entre NDVI y LST

En la siguiente grafica dado un punto de la  $STtD(NDVI)$ , se obtiene el valor correspondiente a  $STtD(LST)$ .

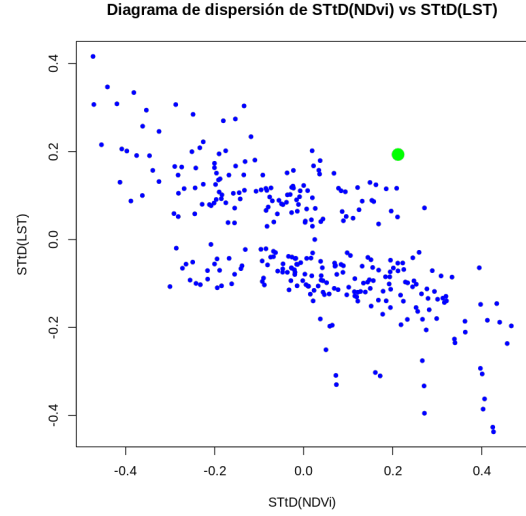


Figura 4: Relación entre NDVI y LST.

Este punto hace referencia a una curva de  $LST$ , el cual corresponde al año 2022, graficando el comportamiento del  $LST$  y  $NDVI$ , se tiene lo siguiente:

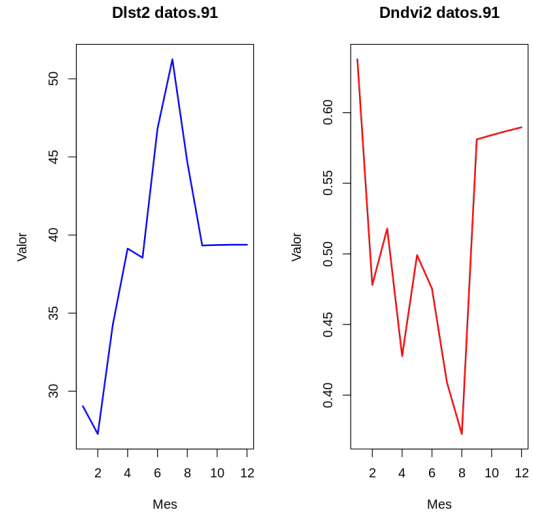


Figura 5: Relación inversa entre LST (grafico izquierdo) y NDVI (grafico derecho).

Se observa una relación inversa: a mayor NDVI (vegetación más densa), menor LST. Esto sugiere que la vegetación contribuye a mitigar el calor superficial.

#### V. REGRESIÓN CUANTIL FUNCIONAL BASADA EN CÓPULAS

Se empleó el enfoque de regresión cuantil funcional basada en cópulas utilizando la función `vinereg`. El modelo ajustado es:

$$STtD(LST_T) \sim STtD(LST_{T-1}) + STtD(NDVI_{T:T-4})$$

Este enfoque permite capturar relaciones no lineales y heterogéneas entre la respuesta y las covariables. La función prueba distintas familias de cópulas y selecciona la que mejor se ajusta según el AIC.

Las cópulas empleadas pueden ser, por ejemplo:

- Clayton: captura dependencias fuertes en eventos bajos.
- Gumbel: captura dependencias en eventos altos.
- t-Student: para relaciones complejas y colas pesadas.

#### V-A. Resumen del Modelo Ajustado

tree	edge	conditioned	conditioning	var_types	family	rotation	parameters	df	tau	loglik
<int>	<int>	<list>	<list>	<chr>	<chr>	<int>	<list>	<dbl>	<dbl>	<dbl>
1	1	1, 2		c,c	t	0	0.6713671, 2.0000010	2	0.4685852	135.84551
1	2	2, 3		c,c	t	0	-0.6100861, 4.5812513	2	-0.4177303	84.49518
2	1	1, 3	2	c,c	gaussian	0	-0.453274	1	-0.2994882	34.46329

Se observa que:

El cuadro anterior resume el modelo ajustado mediante cópulas *vine* a través de varios elementos:

- **tree**: Indica el nivel o “árbol” dentro de la estructura *vine*. En el primer árbol se capturan dependencias directas entre pares de variables, mientras que en árboles posteriores se modelan dependencias condicionales.
- **edge**: Representa el número de la arista o vínculo dentro de cada árbol. Cada arista conecta uno o más pares de variables (o nodos) y define una relación de dependencia.
- **conditioned**: Especifica cuáles variables están siendo condicionadas. Estas conforman el par de variables cuya dependencia se está modelando en esa arista.
- **conditioning**: Indica sobre qué variables se está condicionando la relación. Cuando aparece un guión (“-”), significa que no existen variables condicionantes (primer árbol); si hay variables listadas, la dependencia es condicional al conjunto especificado.
- **var\_types**: Muestra el tipo de variables involucradas (por ejemplo, “c,c” para variables continuas).
- **family**: Indica la familia de la cópula elegida (por ejemplo, *t* o *gaussian*). Cada familia capta distintos patrones de dependencia, como colas más pesadas (*t*) o dependencias más simétricas (*gaussiana*).
- **parameters**: Lista los parámetros estimados de la cópula. En el caso de la cópula *t*, se incluyen el coeficiente de correlación y los grados de libertad. En la cópula *gaussiana*, el parámetro principal es el coeficiente de correlación.
- **df**: Representa los grados de libertad (en el caso de cópulas *t*), lo que impacta la forma de las colas de la distribución.
- **tau**: Corresponde al coeficiente de correlación de Kendall ( $\tau$ ), una medida no paramétrica de dependencia que describe la fuerza y dirección de la relación entre las variables.
- **loglik**: Es la log-verosimilitud asociada a esa arista del *vine*. Valores más elevados suelen indicar un mejor ajuste del modelo a los datos.

#### Visualización de las Dependencias

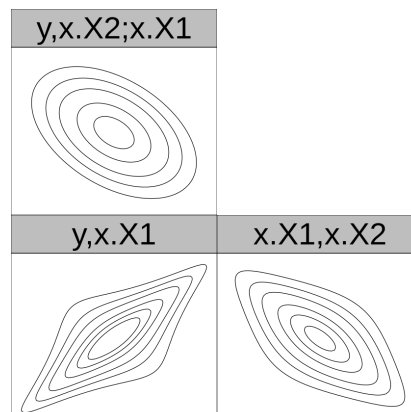


Figura 6: Contornos de cópulas que representan las dependencias entre las variables del modelo.

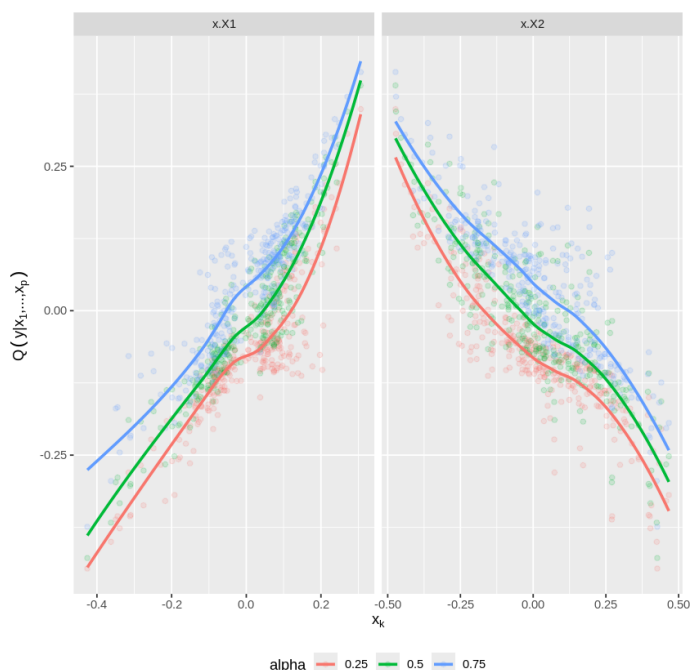


Figura 7: Efectos marginales: la distribución condicional de la respuesta para diferentes cuantiles en función de los predictores.

#### Análisis por Bandas de NDVI

A continuación se presentan tres bandas basadas en los valores  $STd(NDVI)$  del 2023, mostrando el comportamiento de las curvas de LST en dicho año:

Además, aborda eficazmente el desafío de la alta dimensionalidad al transformar un espacio complejo en uno más manejable, lo que resulta en una representación más eficiente y práctica para el análisis.

## REFERENCIAS

- [1] Ambriz, E. (2013). *Desarrollo humano e intensidad migratoria: modelado conjunto vía cópulas*. Tesis de Doctorado, Universidad Nacional Autónoma de México.
- [2] Czado, C. (2019). *Analyzing dependent data with vine copulas*. Lecture Notes in Statistics, Springer.
- [3] Kraus, D., & Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, 110, 1–18.

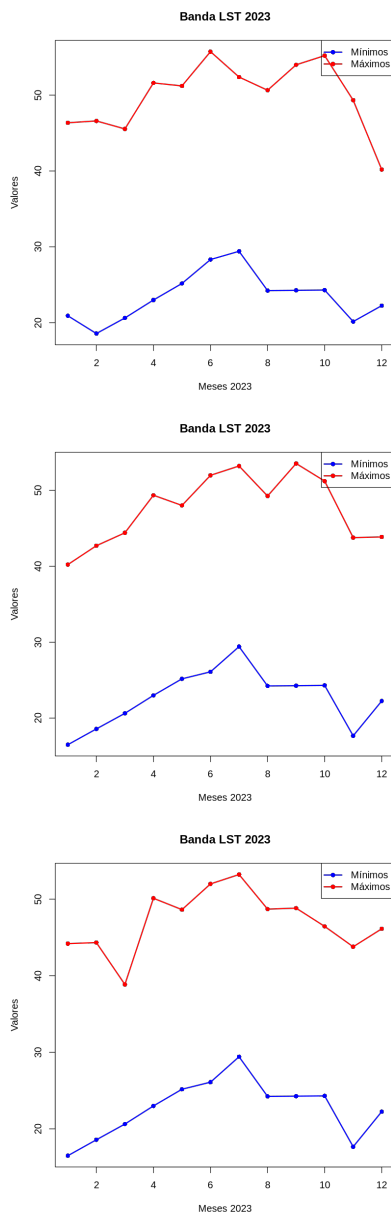


Figura 8: Bandas que muestran el comportamiento del LST 2023 en función de distintos valores de NDVI aplicado el STtD.

Tomando un punto del NDVI transformado con STtD, este corresponde a un valor de LST transformado con STtD, el cual nos da dos puntos en LST. De esta forma se obtienen un conjunto de curvas. Tomando los valores máximos y mínimos de cada mes, se generan las bandas, como se muestra en la imagen 8.

## VI. CONCLUSIÓN

Este enfoque es altamente innovador, ya que combina diferentes metodologías para integrar perspectivas funcionales y escalares. Su flexibilidad permite capturar y analizar el comportamiento funcional de forma simplificada, facilitando su interpretación.