

# Probabilistic model of the relationship between ENSO and the occurrence of agricultural droughts in southern Peru

Juan Cabrera \* , Jesús Mejía , Edilberto Guevara-Pérez 

*Programa de Doctorado en Recursos Hídricos, Escuela de Post Grado, Universidad Nacional Agraria La Molina. Lima, Perú*



<https://doi.org/10.54139/revinguc.v28i1.13>

**Abstract.-** In the article, the information on rainfall in the Province of Candarave, located in southern Peru, is analyzed and contrasted with different indices of the ENSO phenomenon to define whether there is influence of this phenomenon on the dry seasons in the region and build a probabilistic model using copula functions. The EMI (El Niño Modoki Index), ONI (Oceanic Niño Index), TNI (Trans-Niño Index), ICEN (Coastal El Niño Index), and the temperature anomalies ENSO34 (from zone 34) and ENSO 1 +2 (from zone 1 + 2) were analyzed and correlated with the three-month standardized precipitation index, SPI3, usually taken as an indicator of the occurrence of agricultural droughts. The results show that there is an association between the SPI3 index and the EMI, ENSO34, TNI and ONI indices at a significance level of 5 %. The goodness-of-fit analysis shows that the Gumbel-type copula is the most representative of the phenomenon being evaluated; therefore, the corresponding probabilistic model was built. The results allow inferring at a probabilistic level the possible occurrence of agricultural droughts based on the occurrence of the ENSO phenomenon.

**Keywords:** ENSO; probabilistic model; agricultural drought; copula functions.

## Modelación probabilística de la relación entre ENSO y sequías agrícolas en el sur de Perú

**Resumen.-** En el artículo se analiza la información de precipitaciones en la Provincia de Candarave, ubicada en el Sur del Perú y se contrasta con diferentes índices del fenómeno ENSO para definir si existe influencia de este fenómeno sobre las temporadas secas en la región y construir un modelo probabilístico usando funciones cópula. Se analizaron los índices EMI (El Niño Modoki Index), ONI (Oceanic Niño Index), TNI (Trans-Niño Index), ICEN (Índice Costero El Niño), y las anomalías de temperatura ENSO34 (de la zona 34) y ENSO 1+2 (de la zona 1+2), y se correlacionaron con el índice de precipitación estandarizada trimensual, SPI3, tomado usualmente como indicador de la ocurrencia de sequías agrícolas. Los resultados muestran que existe dependencia del índice SPI3 y los índices EMI, ENSO34, TNI y ONI a un nivel de significancia del 5 %. El análisis de la bondad de ajuste muestra que la cópula tipo Gumbel es el más representativo del fenómeno que se evalúa por lo que, se construyó el modelo probabilístico correspondiente. Los resultados permiten inferir a nivel probabilístico la posible ocurrencia de sequías agrícolas en base a la ocurrencia del fenómeno ENSO.

**Palabras clave:** ENSO; modelo probabilístico; sequía agrícola; funciones cópula.

Recibido: 02 de febrero, 2021.

Aceptado: 05 de abril, 2021.

### 1. Introducción

Durante las últimas décadas, la ocurrencia de eventos extremos pareciera haberse intensificado en Perú [1], con un incremento en las precipitaciones en el norte y reducción de estas en la zona sur del país, ocurrencias que podrían

estar asociadas al proceso de cambio climático que actualmente envuelve al planeta. Las fuertes precipitaciones en el norte están usualmente asociadas a la presencia de El Niño-Oscilación del Sur (ENSO, por sus siglas en inglés), mientras que las sequías en el sur se asocian a sistemas de circulación anticiclónica, como La Alta de Bolivia; sin embargo, no se descarta la influencia de ENSO sobre la ocurrencia de sequías, como demostró Vega [2], quien estableció que ambos fenómenos estaban relacionados y que se requerían más estudios especialmente en lo relacionado a

\* Autor para correspondencia:

Correo-e: 20180857@lamolina.edu.pe (J. Cabrera)

sequías agrícolas y meteorológicas.

El problema de las sequías en Perú afecta directamente a la población rural, cuya principal actividad económica es la agricultura; ocasiona pérdidas monetarias y trastoca el normal desarrollo de actividades socioeconómicas [3, 4, 5, 6], dando lugar a oleadas masivas de migración, como las ocurridas en la primera mitad de la década de 1980-1990 [7].

Con la finalidad de ampliar el conocimiento sobre el tema en el sur de Perú, en esta investigación se propone estudiar la probabilidad de ocurrencia conjunta de las sequías y el fenómeno ENSO, desarrollando un modelo probabilístico en base a una función cópula [8, 9, 4], utilizando el procedimiento sugerido por Rayens [10] y Salvadori [4]. Con este objetivo, se evalúan algunos índices El Niño, tales como anomalías de temperatura del agua superficial del océano en las zonas 1+2 (ENSO1+2) y 34 (ENSO34), índice oceánico El Niño (ONI), índice costero El Niño (ICEN), índice El Niño Modoki (EMI), índice de evolución El Niño (TNI); los cuales se contrastan con el índice de precipitación estandarizada de orden 3 (SPI3), seleccionado como indicador de ocurrencia de sequías. La finalidad es definir el nivel de influencia entre los mencionados índices, para luego construir el modelo en cópulas que represente la probabilidad conjunta de ocurrencia y permita estimar las probabilidades de ocurrencia de las sequías para diferentes escenarios.

La técnica propuesta ha sido utilizada en otros países [11, 12, 13, 14], y en el mismo Perú [6], para establecer relaciones entre estadísticos de sequía (duración, intensidad y magnitud) sin embargo estos estudios solamente analizan las sequías y no su relación con otros fenómenos climáticos. Recientemente, Cabrera y Mejía [15] realizaron un primer intento de construir un modelo probabilístico bivariado entre estas variables, pero utilizando menos estaciones y una menor longitud de registro, encontrando poca representatividad de las funciones marginales encontradas y recomendando el uso de probabilidades empíricas. De esta manera, el presente artículo representa el segundo avance de un modelo probabilístico mucho más complejo construido para comprender

las relaciones entre variables climáticas asociadas a la ocurrencia de sequía agrícola y servirá como referencia para futuras investigaciones y como insumo de posibles planes de gestión y/o mitigación de sequías en la región de estudio.

## 2. Metodología

### 2.1. Área de estudio y datos

Se ha seleccionado para la investigación la zona de Candarave ubicada en el Departamento de Tacna, en el sur de Perú. El área está definida por las Coordenadas UTM 350000 - 370000 Este y 8082000 - 8096000 Sur, con altitudes que varían entre 2800 y 3800 m s.n.m. La principal actividad económica es la agricultura; el riego es operado por cuatro comisiones principales: Cairani, Candarave, Huanuara y Quilahuani (ver Figura 1). La región tiene un clima moderadamente frío, con temperaturas que oscilan entre los 6,8 °C (temperatura media máxima en noviembre) y 1,9 °C (temperatura media mínima en julio). La precipitación total mensual en Candarave varía entre 0 y 55 mm, siendo diciembre, enero, febrero y marzo, los meses húmedos; y, mayo, junio y julio, los meses más secos; éstos últimos definen una escasez permanente de agua.

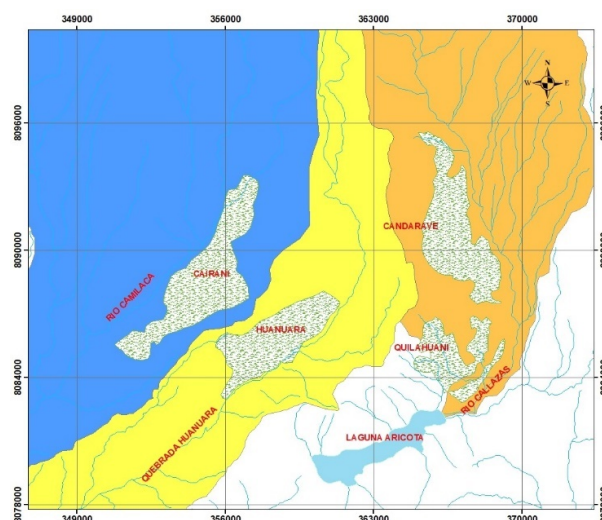


Figura 1: Zona de estudio. Los triángulos muestran la ubicación de las estaciones climáticas utilizadas

Para la investigación se utilizaron las precipitaciones mensuales del período 1964-2020

registradas en tres estaciones climáticas del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) ubicadas dentro del área de estudio. En la Tabla 1 se presentan las características geográficas de ubicación de dichas estaciones: latitud, longitud y altitud.

Tabla 1: Características de ubicación geográfica de las Estaciones climáticas utilizadas

| Estación  | Coordenadas geográficas<br>– Zona 19 Datum WGS 84 |                |                       |
|-----------|---------------------------------------------------|----------------|-----------------------|
|           | Longitud<br>(m)                                   | Latitud<br>(m) | Altitud<br>(m s.n.m.) |
| ARICOTA   | 70°14' O                                          | 17°19' S       | 2843                  |
| CANDARAVE | 70°15' O                                          | 17°16' S       | 3415                  |
| CAIRANI   | 70°22' O                                          | 17°17' S       | 3205                  |

La información correspondiente a los índices El Niño se descargó de la base de datos del Centro Climático APEC (APCC, por su acrónimo en inglés) y del Instituto Geofísico del Perú (IGP).

## 2.2. Identificación de sequías

Las sequías se definen como desviaciones climáticas negativas respecto a condiciones consideradas normales de una región [3, 16]. Esta condición de humedad insuficiente puede ser causada por un déficit en la precipitación o un retraso del inicio a lo largo de un cierto período de tiempo.

Dependiendo de la duración temporal, las sequías pueden afectar en mayor o menor medida la disponibilidad hídrica y las condiciones y actividades asociadas del entorno. De esta forma se habla de una sequía meteorológica, si el déficit de precipitaciones está por debajo de lo normal; sequía agrícola, si el déficit de precipitaciones afecta el normal crecimiento de las plantas; sequía hidrológica, si el déficit afecta el almacenamiento subterráneo y el ciclo hidrológico; y, sequía socioeconómica, si el déficit afecta la oferta y demanda de un cierto producto. El presente estudio se enfoca en la ocurrencia de sequías agrícolas y, por tanto, en lo sucesivo se utilizará el término “sequía” para referirse a este tipo en especial.

Como se ha mencionado, el análisis convencional de ocurrencia de sequías se realiza en base al estudio de cuantiles o en base a índices,

denominados “índices de sequía”. En este trabajo se eligió el Índice de Precipitación Estandarizada (SPI, por su acrónimo en inglés) [17] como indicador de la ocurrencia de sequías. La selección del índice SPI se debió al bajo requerimiento de información necesario para el cálculo en comparación con otros como, por ejemplo, el Índice de Severidad de Sequías de Palmer (PDSI, por su acrónimo en inglés) [18]; pero también por la eficiencia que ha mostrado en la caracterización y ocurrencia de sequías [19, 20, 3, 4, 6, 16].

El SPI se calcula “agregando precipitaciones mensuales” en escalas de 1, 3, 6, 12 y 24 meses, de acuerdo con el efecto que se quiere capturar. Para sequías meteorológicas, cuyos efectos se manifiestan a corto plazo, se recomienda la escala de 1 mes; para sequías agrícolas, escala 1 o 3 meses; para sequías hidrológicas, escalas de 6, 12 o 24 meses. Las series agregadas se ajustan a una distribución de probabilidad y luego de transformar a una distribución normal estándar, quedando definido el índice SPI como se muestra en la ecuación (1):

$$SPI = \frac{y - \mu}{\sigma}, \quad (1)$$

donde  $y$  es la precipitación agregada mensual;  $\mu$  es la media de la variable precipitación agregada mensual; y  $\sigma$  es la desviación estándar de la variable precipitación agregada mensual.

El modelo (1) arroja para el índice SPI valores positivos y negativos. Los valores positivos representan precipitaciones por encima de un valor medio; es decir, representan la ocurrencia de un período húmedo; los valores negativos representan precipitaciones por debajo de la media; es decir, un período seco. En base a esta caracterización, McKee [17] define cuatro categorías de sequía, las cuales se presentan en la Tabla 2.

De acuerdo con lo indicado, en el presente estudio se usa el SPI a escala de 3 meses (SPI3), como indicador de ocurrencia de sequía agrícola.

## 2.3. Identificación de las zonas El Niño

El fenómeno El Niño Oscilación del Sur (ENSO, por sus siglas en inglés) es un ciclo de desviaciones periódicas de la temperatura superficial del agua

Tabla 2: Categorías de sequías en base al índice SPI [17]

| SPI           | Categoría       |
|---------------|-----------------|
| 0,00 a -0,99  | Sequía leve     |
| -1,00 a -1,49 | Sequía moderada |
| -1,50 a -1,99 | Sequía severa   |
| -2,00 o menos | Sequía extrema  |

en el Océano Pacífico ecuatorial, y en la atmósfera, denominadas anomalías. Estas desviaciones representan incrementos o reducciones de temperatura que pueden afectar los patrones climáticos en todo el mundo, ocasionando precipitaciones intensas en algunas regiones y períodos muy secos en otras, datos que están disponibles en <https://psl.noaa.gov/data/climateindices/list/>).

El estudio de este fenómeno ha permitido definir regiones en el Pacífico Ecuatorial donde la ocurrencia de estas desviaciones está directamente asociada con los cambios climáticos en diferentes partes del planeta. En el caso de Perú, las anomalías que se presentan en la región 1+2 influyen fuertemente en la ocurrencia de precipitaciones extremas en el norte del país. La Figura 2 muestra las regiones “El Niño”.

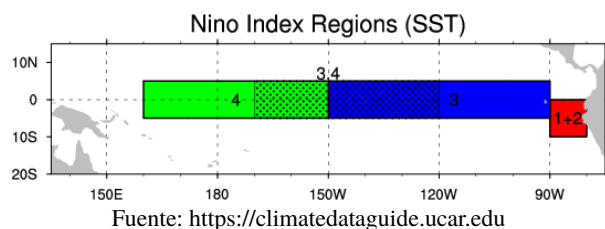


Figura 2: Regiones “El Niño”

Un primer indicador de la presencia del ENSO son las anomalías de temperatura del agua superficial del mar en cada una de las regiones “El Niño”; sin embargo, teniendo en cuenta la complejidad del fenómeno, se han desarrollado diferentes “índices” para definir su ocurrencia. Para el presente estudio se toman en cuenta las anomalías de temperatura superficial del mar en las zonas 1+2 (ENSO1+2) y 3+4 (ENSO3+4), el índice oceánico El Niño (ONI), el índice costero El Niño (ICEN), el índice El Niño Modoki (EMI), y el índice de evolución El Niño (TNI).

## 2.4. Medidas de asociación

La asociación entre el índice SPI3, considerado como indicador de sequía agrícola, y los índices ENSO indicados, se analizó utilizando enfoques cualitativos y cuantitativos. La dependencia cualitativa se evaluó mediante un análisis gráfico, y fue utilizado también para eliminar índices ENSO que muestran dependencia entre sí. La dependencia cuantitativa se evaluó mediante el coeficiente de correlación lineal  $r$  de Pearson, el coeficiente de correlación de rango de Spearman y el de Kendall.

## 2.5. Función cópula

Una cópula es una función multivariada que une o acopla funciones unidimensionales de distribución marginal correspondientes al grupo de variables que se busca analizar [10]. El concepto fue introducido por Sklar [9, 4], quien propuso que toda función de distribución  $F$ ,  $n$ -dimensional, puede descomponerse en dos partes, una función de distribución marginal  $F_i$  y una cópula  $C$ , que representa la dependencia de la distribución.

Sklar [10, 21] estableció la relación entre una función de distribución acumulada conjunta  $F_X$  y una cópula mediante el llamado Teorema de Sklar. En este teorema se propone que, para una función de distribución acumulada conjunta  $F_X$  con distribuciones marginales  $F_{X1}$  y  $F_{X2}$ , existe una cópula bidimensional  $C$  tal que  $\forall x \in R^2$  (ecuación (2)):

$$F_X(x_1, x_2) = C(F_{X1}(x_1), F_{X2}(x_2)). \quad (2)$$

El teorema puede generalizarse para cópulas  $n$ -dimensionales como se muestra en la ecuación (3):

$$F_X(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (3)$$

Asimismo, Sklar demostró que, si las distribuciones marginales son continuas, la cópula es única.

Para el presente estudio se seleccionaron dos cópulas elípticas (normal y  $t$ ), y tres cópulas arquimedianas (Clayton, Frank y Gumbel), por ser las más usuales en Hidrología y Climatología [8, 21]. En la Tabla 3 se presenta un resumen de las definiciones de cada una de las cópulas seleccionadas para el caso trivariado.

Tabla 3: Definición de cópulas trivariadas utilizadas para el estudio [4]

| Copula     | Cópula trivariada $C_\theta(u, v)$                                                                                                                                                                                                                                | Parámetros                                                                          |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Normal     | $C_N(u; \Sigma) = \int_{-\infty}^{\phi^{-1}(u_1)} \cdots \int_{-\infty}^{\phi^{-1}(u_3)} \frac{1}{\sqrt{(2\pi)^3  \Sigma }} \exp\left(-\frac{1}{2} X' \Sigma^{-1} X\right) dx$                                                                                    | $\Sigma$ es la matriz de correlación                                                |
| t(Student) | $C_t(u; \Sigma; \nu) = \int_{-\infty}^{t^{-1}(u_1)} \cdots \int_{-\infty}^{t^{-1}(u_3)} \frac{\Gamma\left(\frac{\nu+3}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \sqrt{(\pi\nu)^3  \Sigma }} \left(1 + \frac{x' \Sigma^{-1} x}{\nu}\right)^{-\frac{\nu+3}{2}} dx$ | $\Sigma$ es la matriz de correlación; $\nu$ es el número de grados de libertad      |
| Clayton(*) | $\left(\left(U_1^{-\theta_2} + U_2^{-\theta_2} - 1\right)^{\theta_1/\theta_2} + U_3^{-\theta_1} - 1\right)^{-1/\theta_1}$                                                                                                                                         | $\theta_2 > \theta_1$                                                               |
| Frank(*)   | $-\theta_1^{-1} \ln\left(C_1^{-1} \left(1 - \left(1 - C_2^{-1} \left(1 - e^{-\theta_2 U_1}\right) \left(1 - e^{-\theta_2 U_2}\right) \left(1 - e^{-\theta_1 U_3}\right)\right)^{\theta_1/\theta_2}\right)\right)$                                                 | $C_1 = 1 - e^{-\theta_1}$ ,<br>$C_2 = 1 - e^{-\theta_2}$ ;<br>$\theta_2 > \theta_1$ |
| Gumbel(*)  | $\exp\left(-\left[\left(-\ln U_1\right)^{\theta_2} + \left(-\ln U_2\right)^{\theta_2}\right]^{\theta_1/\theta_2} + \left(-\ln U_3\right)^{\theta_1}\right)^{1/\theta_1}$                                                                                          | $\theta_2 > \theta_1$                                                               |

(\*) Si la cópula es simétrica,  $\theta_2 = \theta_1 = \theta$ 

Para estimar los parámetros de cópulas, se utilizó el enfoque de máxima pseudo-verosimilitud (MPL) por ser el más recomendado para análisis multivariados [21]. El método consiste en estimar pseudo-observaciones y, en base a ellas, estimar, a su vez, la función de pseudo verosimilitud a maximizar.

Para determinar el mejor ajuste a los datos y en consecuencia seleccionar la cópula, se utiliza la prueba paramétrica de Cramér-von Mises ( $S_n$ ) definida mediante la ecuación (4):

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(u)^2 dC_n(u), \quad (4)$$

donde:  $d$ , es el número de variables en análisis;  $C_n$  es el valor calculado utilizando la cópula empírica; y,  $\mathbb{C}_n$  es la diferencia entre el valor calculado con la cópula empírica y el calculado con la familia de cópulas en evaluación como se observa en la ecuación (5):

$$\mathbb{C}_n = \sqrt{n} (C_n - C_{\theta_n}). \quad (5)$$

Un valor alto de estadístico  $S_n$  indica el rechazo de la hipótesis nula  $H_0$ , que propone que “la cópula empírica  $C$  pertenece a la familia de cópulas  $C_\theta$ ”. Para el análisis indicado, se utilizó el paquete “Copula” en lenguaje R disponible en <https://CRAN.R-project.org/package=copula>.

### 3. Resultados y discusión

#### 3.1. Relación entre variables

En una primera etapa, se procedió a la estimación del índice SPI3 para cada una de las estaciones incluidas en el estudio. En la Figura 3 se presentan los resultados obtenidos para el índice SPI3 calculado para cada una de las estaciones: a) Aricota; b) Cairani; c) Candarave (índice SPI3 en las ordenadas y meses en las abscisas). Las series de los índices así calculados se utilizan como variable de entrada en la construcción del modelo cópula. En la Tabla 4 se presentan las características promedio de las sequías evaluadas según el índice SPI3, extraídas de las de las gráficas dadas en la Figura 3. Los resultados muestran que el número de eventos registrados se incrementa con la altitud de la zona y, consecuentemente, el período entre eventos se reduce. Asimismo, se observa que la magnitud y la intensidad aumentan con la altitud pero el máximo SPI negativo registrado (pico) se reduce (Tabla 4). Este comportamiento indica que, a mayor altitud, mayor número de sequías, más severas pero de menor intensidad.

Con la serie generada de índices SPI3 se procedió a realizar un análisis gráfico para evaluar la dependencia con los índices ENSO. Los resultados se presentan en la Figura 4, en la cual se observa una fuerte dependencia entre los índices ENSO1+2 e ICEN, ENSO34 y ONI, TNI y EMI (ver Figura 4 para la estación Aricota); por lo

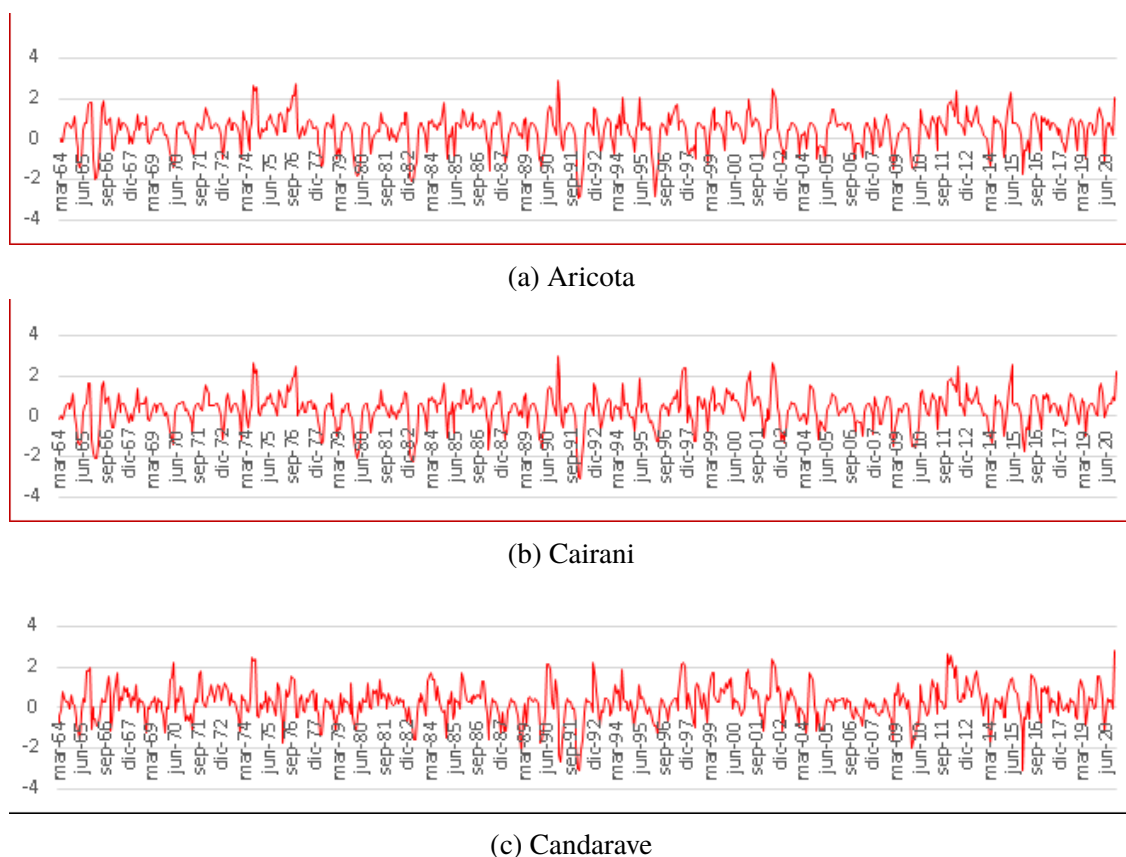


Figura 3: Resultados del índice SPI3 calculado para las estaciones usadas. Índice SPI3 vs. meses.

Tabla 4: Características promedio de las sequías evaluadas con el SPI3

| Características               | Aricota | Cairani | Candarave |
|-------------------------------|---------|---------|-----------|
| Numero de eventos             | 25      | 24      | 31        |
| Duración promedio (meses)     | 3,36    | 3,33    | 3,10      |
| Magnitud                      | -3,74   | -3,95   | -3,11     |
| Intensidad                    | -1,09   | -1,13   | -0,99     |
| Pico                          | -1,48   | -1,50   | -1,53     |
| Período entre eventos (meses) | 28      | 28      | 22        |

que se descartó el uso de los índices ICEN, ONI y TNI. Para determinar la bondad de ajuste de las correlaciones se calcularon los coeficientes de correlación entre los índices restantes y el índice SPI3, según se muestra en la Tabla 5; los valores calculados muestran que en las estaciones Aricota y Cairani existe una correlación muy baja pero significativa (con  $p - valor < 0,05$ ) entre el índice SPI3 y los índices ENSO34 y EMI; por tanto, se descartó el índice ENSO1+2.

En la estación Candarave, se encontró una correlación muy baja pero significativa solamente entre los índices SPI3 y EMI, esto implica que la influencia de las anomalías en la temperatura del mar de la zona 3-4 se pierde a una altitud de 3400 m s.n.m. aproximadamente; sin embargo, teniendo en cuenta que el  $p - valor$  es 0,06 (cercano a la significancia de 0,05), se tomará en cuenta para la construcción del modelo.

### 3.2. Aplicación de cópulas

Debido a que las variables que se analizan no se ajustan a distribuciones de probabilidad conocidas, se decidió utilizar la probabilidad empírica y se calcularon las pseudo-observaciones para cada variable. Las familias de cópulas elípticas y Arquímedeanas, descritas en la Tabla 3, se ajustaron utilizando el método de máxima pseudo-verosimilitud (MPL) y la performance de las diferentes familias de cópulas se evaluó mediante la prueba de Cramér-von Mises ( $S_n$ ). En la Tabla 6 se presenta el resumen de los estadísticos y

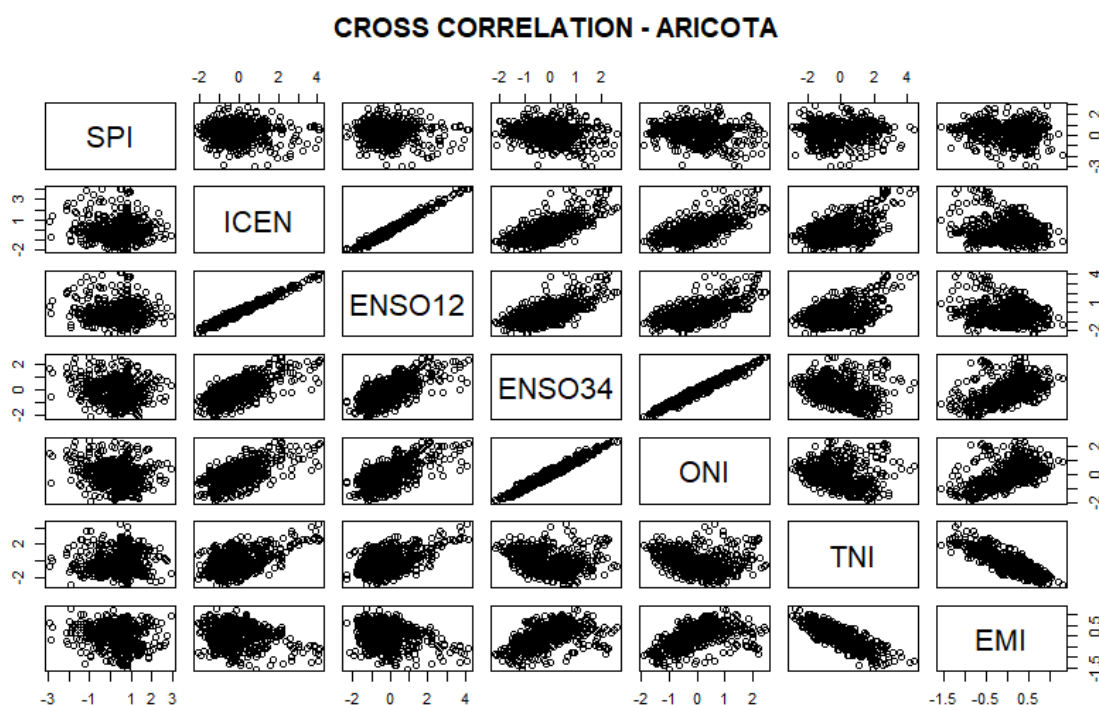


Figura 4: Análisis de correlaciones entre variables en estudio. Estación Aricota. Cada índice se encuentra graficado en el eje vertical de la fila correspondiente y en el eje horizontal de la columna correspondiente, como señalan los círculos rojos para el caso del SPI

parámetros estimados para cada familia de cópulas. En todas las estaciones, la cópula  $t$  (Student) mostró un menor valor del estadístico  $S_n$  y, por tanto, un mejor ajuste a los datos; asimismo, los valores  $p$  muestran significancia al 5 %. Por otro lado, se observa que las cópulas elípticas, normal y  $t$  (Student), son las que arrojan menor valor en el estadístico  $S_n$ , lo que indica que este tipo de cópulas tiene mejor performance para la evaluación de este tipo de variables climáticas.

Finalmente, se procedió a generar una serie aleatoria de 10.000 datos, cantidad mínima recomendada para mantener estables los valores del estadístico de bondad de ajuste de la cópula [22]. Los resultados de la simulación se muestran en la Figura 5. Estas series sintéticas tienen una longitud mucho mayor que la original, representando un espacio muestral más grande y, por tanto, con mayor probabilidad de encontrar las distribuciones marginales de cada variable.

La generación conjunta de estas series busca preservar la dependencia entre variables e incrementar el número de datos para facilitar el cálculo de las

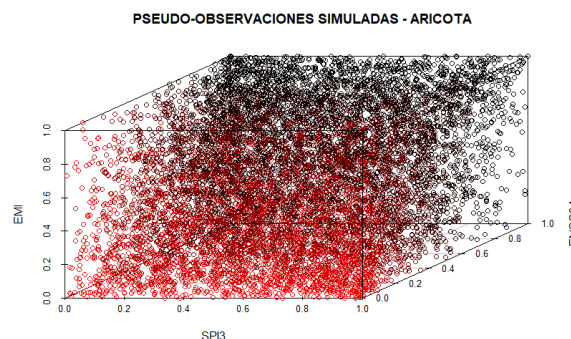


Figura 5: Simulación de 10000 pseudo observaciones para la estación Aricota.

funciones marginales.

El objetivo de generar estas series sintéticas en base a la función cópula es preservar la relación encontrada entre variables, la cual podría perderse si se generan series sintéticas para cada variable de forma independiente, así como definir un espacio muestral mayor que permita estimar valores de probabilidad empírica con mejor aproximación en para los casos en que no existan distribuciones

Tabla 5: Correlaciones calculadas para las estaciones en estudio. Se observa que solamente los índices EMI y ENSO34 tienen significancia al 5 % en todas las estaciones

| Estación  | Índice ENSO | Tipo de correlación | Coefficiente de correlación | Valor p | Significante al 5 % |
|-----------|-------------|---------------------|-----------------------------|---------|---------------------|
| ARICOTA   | EMI         | Kendall             | -0,0931                     | 0,0003  | Si                  |
|           |             | Spearman            | -0,134                      | 0,0005  | Si                  |
|           |             | Pearson             | -0,1231                     | 0,0013  | Si                  |
|           | ENSO1+2     | Kendall             | -0,0291                     | 0,2635  | No                  |
|           |             | Spearman            | -0,0433                     | 0,2592  | No                  |
|           |             | Pearson             | -0,0961                     | 0,0121  | Si                  |
|           | ENSO34      | Kendall             | -0,0861                     | 0,0009  | Si                  |
|           |             | Spearman            | -0,1287                     | 0,0008  | Si                  |
|           |             | Pearson             | -0,178                      | 0,0000  | Si                  |
| CAIRANI   | EMI         | Kendall             | -0,1237                     | 0,000   | Si                  |
|           |             | Spearman            | -0,1752                     | 0,000   | Si                  |
|           |             | Pearson             | -0,1622                     | 0,000   | Si                  |
|           | ENSO1+2     | Kendall             | -0,0178                     | 0,493   | No                  |
|           |             | Spearman            | -0,0284                     | 0,4591  | No                  |
|           |             | Pearson             | -0,0554                     | 0,1487  | No                  |
|           | ENSO34      | Kendall             | -0,0993                     | 0,000   | Si                  |
|           |             | Spearman            | -0,1449                     | 0,000   | Si                  |
|           |             | Pearson             | -0,164                      | 0,000   | Si                  |
| CANDARAVE | EMI         | Kendall             | -0,0615                     | 0,017   | Si                  |
|           |             | Spearman            | -0,0907                     | 0,0179  | Si                  |
|           |             | Pearson             | -0,121                      | 0,0015  | Si                  |
|           | ENSO1+2     | Kendall             | -0,0283                     | 0,2716  | No                  |
|           |             | Spearman            | -0,0424                     | 0,2685  | No                  |
|           |             | Pearson             | -0,0521                     | 0,1745  | No                  |
|           | ENSO34      | Kendall             | -0,0469                     | 0,0687  | No                  |
|           |             | Spearman            | -0,0708                     | 0,0648  | No                  |
|           |             | Pearson             | -0,0851                     | 0,0263  | Si                  |

Tabla 6: Prueba de bondad de ajuste y parámetros de la cópula multivariada. La cópula con mejor ajuste se indica en **negrita**

| Estación  | Cópula              | $S_n$           | Valor p          | Parámetros                                  |
|-----------|---------------------|-----------------|------------------|---------------------------------------------|
| ARICOTA   | Gumbel              | 0,65747         | 0,0004995        | 1,0596                                      |
|           | Clayton             | 0,60015         | 0,0004995        | 0,141                                       |
|           | Frank               | 0,60287         | 0,0004995        | 0,66602                                     |
|           | Normal              | 0,14599         | 0,002498         | -0,14221; -0,11902; 0,53578                 |
|           | <b>t Student(*)</b> | <b>0,12484</b>  | <b>0,0004995</b> | <b>-0,1312; -0,142; 0,5658; 8,8248</b>      |
| CAIRANI   | Gumbel              | 0,5676          | 0,0004995        | 1,0533                                      |
|           | Clayton             | 0,5374          | 0,0004995        | 0,1168                                      |
|           | Frank               | 0,54041         | 0,0004995        | 0,56528                                     |
|           | Normal              | 0,11195         | 0,01049          | -0,14750; -0,16195; 0,53580                 |
|           | <b>t Student(*)</b> | <b>0,093311</b> | <b>0,01548</b>   | <b>-0,1476; -0,1846; 0,5689; 7,0776</b>     |
| CANDARAVE | Gumbel              | 0,62131         | 0,0004995        | 1,0771                                      |
|           | Clayton             | 0,5254          | 0,0004995        | 0,17864                                     |
|           | Frank               | 0,53792         | 0,0004995        | 0,79468                                     |
|           | Normal              | 0,16219         | 0,0004995        | -0,067464; -0,110520; 0,535740              |
|           | <b>t Student(*)</b> | <b>0,13395</b>  | <b>0,0004995</b> | <b>-0,05478; -0,11335; 0,56548; 7,55966</b> |

(\*) Los tres primeros parámetros de la cópula t Student representan la matriz de correlaciones; el cuarto parámetro es el número de grados de libertad. El estadístico  $S_n$  no está definido para grados de libertad no enteros, por lo que se estimó para el entero más cercano.

marginales que se adecúen a los datos.

En una etapa posterior, estas distribuciones marginales (o las correspondientes probabilidades

empíricas), se utilizarán para la estimación de los períodos de retorno y probabilidad de ocurrencia de diferentes escenarios conjuntos, información

necesaria para la elaboración de planes de mitigación y riesgo de sequías.

#### 4. Conclusiones

En este estudio, se analizó la relación entre la ocurrencia del fenómeno ENSO, definida en base a los índices ENSO1+2, ENSI34, ICEN; TNI; EMI y ONI, y la ocurrencia de sequías agrícolas, definida en base al índice SPI3. Para el efecto, se compararon diferentes modelos probabilísticos basados en cópulas para construir la frecuencia conjunta entre variables. En primer lugar, se correlacionaron entre sí los diferentes índices mencionados para descartar aquellos que no tenían relevancia o aquellos que presentaban una fuerte dependencia respecto a otros, reduciendo el número de índice de interés a tres: ENSO34, EMI y SPI3. El siguiente paso fue evaluar cinco familias de cópulas para probar cuál de ellas presenta mejor ajuste a los datos, encontrándose que la cópula trivariada t de Student tiene mejor performance en todas las estaciones. Los resultados muestran que la cópula de Student es una función robusta y adecuada para el análisis del caso presentado, permitiendo a su vez evaluar la probabilidad de ocurrencia y períodos de retorno de sequías agrícolas en conjunto con la ocurrencia del ENSO, información que es relevante para desarrollar planes de gestión y mitigación de las sequías en la región. Asimismo, se encontró que la influencia de las anomalías de la temperatura del mar en la región ENSO34 se pierde alrededor de la cota 3400 m s.n.m., hecho que contrasta con el incremento en el número de sequías y la mayor severidad registrada por encima de dicha altitud, por lo que se sugiere evaluar la construcción de un modelo probabilístico que incluya la variable “temperatura del aire” para conseguir mayor representatividad en regiones de mayor altitud.

#### 5. Referencias

- [1] Intergovernmental Panel on Climate Change, “Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects.” Cambridge University, Cambridge, United Kingdom and New York, NY, USA, Reporte Técnico, 2014.
- [2] F. Vega, “Variabilidad espacio-temporal de las sequías en el Perú y el peligro asociado al Fenómeno del Niño.” Tesis para optar el grado de Magister Scientiae en recursos hídricos, Universidad Nacional Agraria La Molina, Lima, Perú, 2018.
- [3] C. Knutson, “Methods and Tools for Drought Analysis and Management,” *Eos*, vol. 89, no. 22, 2018.
- [4] G. Salvadori, C. De Michele, N. Kottegoda, and R. Rosso, *Extremes in Nature: An approach using Copulas*. Springer Netherlands, 2007.
- [5] L. Santos, I. Cordery, and I. Iacovides, *Coping with Water Scarcity Addressing the Challenges*. Springer, Dordrecht, 2009.
- [6] SENAMHI, “Análisis de riesgo de sequias en el Sur del Perú.” Ministerio del Ambiente del Perú, Reporte Técnico, 2016.
- [7] INEI, *Perú: Migraciones Internas*. Lima. INEI, 2009.
- [8] A. AghaKouchak, “Entropy-copula in hydrology and climatology,” *Journal of Hydrometeorology*, vol. 15, no. 6, pp. 2176–2189, 2014.
- [9] L. Rüschendorf, *Mathematical Risk Analysis Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer Berlin Heidelberg, 2013, ch. Copulas, Sklar’s Theorem, and Distributional Transform.
- [10] B. Rayens and R. Nelsen, “An Introduction to Copulas by Roger B. Nelsen,” *Technometrics*, vol. 42, no. 3, p. 317, 2000.
- [11] M. Azam, S. Maeng, H. Kim, and A. Murtazaev, “Copula-Based Stochastic Simulation for Regional Drought Risk Assessment in South Korea,” *Water*, vol. 10, no. 4, p. 359, 2018.
- [12] L. Hangshing and P. Dabral, “Multivariate Frequency Analysis of Meteorological Drought Using Copula,” *Water Resources Management*, vol. 32, pp. 1741–1758, 2018.
- [13] F. Serinaldi, B. Bonaccorso, A. Cancelliere, and S. Grimaldi, “Probabilistic characterization of drought properties through copulas,” *Physics and Chemistry of the Earth Parts A/B/C*, vol. 34, no. 10, pp. 596–605, 2009.
- [14] K. Xu, D. Yang, X. Xu, and H. Lei, “Copula based drought frequency analysis considering the spatio-temporal variability in Southwest China,” *Journal of Hydrology*, vol. 527, pp. 630–640, 2015.
- [15] J. Cabrera and J. Mejía, “Relationship between drought occurrence and ENSO in Southern Peru: a copulas analysis,” in *38th IAHR World Congress-“Water: Connecting the World”*, Panama, 2019.
- [16] D. A. Wilhite and M. Glantz, “Understanding the drought phenomenon: The role of definitions,” *Water International*, vol. 10, no. 3, pp. 111–120, 1985.
- [17] T. McKee, N. Doesken, and J. Kleist, “The relationship of drought frequency and duration to time scales,” in *Eighth Conference on Applied Climatology*, Anaheim, California, 1993, pp. 17–22.
- [18] L. Nkemdirim, *Encyclopedia of Atmospheric Sciences*,

2nd ed. Academic Press, 2015, ch. HYDROLOGY, FLOODS AND DROUGHTS | Palmer Drought Severity Index, pp. 224–231.

- [19] A. Cancelliere, G. Di Mauro, B. Bonaccorso, and G. Rossi, “Drought forecasting using the Standardized Precipitation Index,” *Water Resources Management*, vol. 2, no. 5, pp. 801–819, 2007.
- [20] A. Cancelliere and J. Salas, “Drought length properties for periodic-stochastic hydrologic data,” *Water Resources Research*, vol. 40, no. 2, 2004.
- [21] M. Sadegh, E. Ragno, and A. AghaKouchak, “Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework,” *Water Resources Research*, vol. 56, no. 6, pp. 5166–5183, 2017.
- [22] C. Genest, B. Rémillard, and D. Beaudoin, “Goodness-of-fit tests for copulas: a review and a power study,” *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 199–213, 2009.