

**UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA CIVIL
SECCIÓN DE POSGRADO**



**PREDICCIÓN DE SEQUÍAS CON REDES NEURONALES
ARTIFICIALES Y ALGORITMOS GENÉTICOS UTILIZANDO
PRECIPITACIÓN POR PERCEPCIÓN REMOTA**

TESIS
PARA OPTAR EL GRADO DE MAESTRO EN CIENCIAS CON MENCIÓN
EN INGENIERÍA HIDRÁULICA

ELABORADO POR:
IVÁN ARTURO AYALA BIZARRO

ASESOR:
MSC. ING. MIGUEL ANGEL ZUBIAUR ALEJOS

LIMA - PERÚ

2016

PREDICCIÓN DE SEQUÍAS CON REDES NEURONALES ARTIFICIALES Y
ALGORITMOS GENÉTICOS UTILIZANDO PRECIPITACIONES POR
PERCEPCIÓN REMOTA

IvÁN ARTURO AYALA BIZARRO
Ingeniero Civil

Presentado a la Sección de Posgrado de la Facultad de Ingeniería Civil en
cumplimiento parcial de los requerimientos para el grado de:

MAESTRO EN CIENCIAS CON MENCIÓN EN INGENIERÍA HIDRÁULICA
DE LA UNIVERSIDAD NACIONAL DE INGENIERÍA

2016

Autor : Iván Arturo Ayala Bizarro

Recomendado : MSc. Ing. Miguel Ángel Zubiaur Alejos
Asesor de Tesis

Aceptado por : Dr. Victor Sánchez Moya
Director de la Unidad de Posgrado
Facultad de Ingeniería Civil - UNI



... Todas las ciudades de Judá están tristes y desanimadas; la gente se sienta en el suelo, y en Jerusalén todos lloran. Los gobernantes piden agua, y sus sirvientes van a buscarla, pero los pozos están secos. Confundidos y llenos de vergüenza se agarren la cabeza, pues regresan con sus baldes vacíos. Los campesinos se preocupan y se agarren la cabeza, porque el suelo está reseco y no ha llovido en el país. Tan escasos están los pastos que los venados, en el campo, dejan abandonadas a sus crías. Los asnos salvajes parecen chacales: se paran en las lomas desiertas y desde allí olfatean el aire; pero se desmayan de hambre porque no tienen pastos....

(Jeremías 14:2-6 Antiguo Testamento)

A mis hijos, Jhierald, Yasiel y el pequeño Jhamild, Aún a sus cortas edades, aprendo muchas cosas de la vida. Han sido mi motivación más grande para concluir con éxito este proyecto de tesis.

Agradecimientos

Al Msc. Ing. Miguel Ángel Zubiaur, asesor de la presente tesis, por la paciencia, apoyo incondicional, apreciaciones y buena disposición durante la formulación del presente trabajo de investigación.

A mis jurados la Msc. Ing. Patricia Calderón Hijuma, Msc. Ing. Juan Walter Cabrera Cabrera, por todo el apoyo, recomendaciones y consejos brindados durante la revisión de este trabajo.

A los docentes de la Universidad Nacional de Ingeniería - Facultad de Ingeniería Civil, Dr. Julio Kuroiwa Zevallos, Msc. Roberto Campaña Toro, Msc. Nicolás de Piérola, Msc. Ada Arancibia Samaniego, Dr. Hugo Scaletti, por su contribución durante mi desarrollo académico y profesional.

Al Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), por su atención prestada a mi solicitud de requerimiento de la información pluviométrica de las distintas estaciones que se considera dentro de la tesis.

A mis compañeros de la Maestría de la Universidad Nacional de Ingeniería: Edwing Arapa, Juan Carlos Rodríguez, Freddy Velazquez, grandes amigos y compañeros de carpeta.

UNIVERSIDAD NACIONAL DE INGENIERÍA
LIMA, JUNIO DE 2016

Iván Arturo Ayala Bizarro

Índice general

Dedicatoria	II
Agradecimientos	III
Índice general	IV
Índice de tablas	VII
Índice de figuras	IX
Resumen	XII
Introducción	1
Capítulo I: Generalidades	3
1.1. Antecedentes bibliográficos	3
1.1.1. Antecedentes en la variable Redes Neuronales Artificiales y Sequías	4
1.1.2. Antecedentes en la variable Algoritmos Genéticos y Sequías . . .	5
1.1.3. Antecedentes en la variable Sensoramiento Remoto y Sequías . .	5
1.2. Descripción de la realidad problemática	6
1.3. Formulación del problema	7
1.3.1. Problema principal	7
1.3.2. Problemas secundarios	7
1.4. Justificación e importancia de la investigación	7
1.5. Objetivos	8
1.5.1. Objetivos Generales	8
1.5.2. Objetivos Específicos	8
1.6. Hipótesis	8
1.7. Variables e indicadores	9
1.8. Tipo y nivel de investigación	9
1.8.1. Tipo de Investigación	9
1.8.2. Nivel de Investigación	9
1.9. Período de análisis	10
1.10. Fuentes de información	10
1.10.1. Fuentes primarias	10
1.10.2. Fuentes secundarias	10
1.11. Instrumentos utilizados	11
1.12. Técnicas de recolección	11



Capítulo II: Marco teórico y marco conceptual, desarrollo de la teoría y variable	12
2.1. Precipitación por percepción remota	12
2.1.1. Precipitación CHIRPS	12
2.1.2. Precipitación PISCO	13
2.2. Definición de sequía	14
2.3. Análisis Regional de Frecuencias L-Momentos para Sequías	16
2.3.1. Definición de L-Momentos	17
2.3.2. Procedimiento de Análisis Regional de sequías L-Momentos.	20
2.3.3. Software L-Momentos	24
2.4. Índices de sequía	26
2.4.1. Índice de Condición de Vegetación (VCI)	26
2.4.2. Índice de Precipitación Estandarizada (SPI)	28
2.4.3. Índice de Precipitación Evapotranspiración Estandarizada (SPEI)	30
2.4.4. Índices Climáticos Alternos	32
2.5. Redes Neuronales Artificiales	33
2.5.1. Características de las Redes Neuronales Artificiales	34
2.5.2. Estructura Básica de una Red Neuronal	35
2.5.3. Función de activación	36
2.5.4. Entrenamiento de las Redes Neuronales Artificiales	37
2.5.5. El perceptron Multicapa	38
2.5.6. Arquitectura de la Red	39
2.5.7. Modelo Self Organizing Map	40
2.5.8. Modelo No Lineal Auto regresivas con entradas Externas (NARX)	42
2.6. Algoritmos Genéticos	43
2.6.1. Bases Biológicas	44
2.6.2. Terminologías y Operadores de AG	44
2.6.2.1. Codificación	44
2.6.2.2. Operador de Selección	45
2.6.2.3. Operador de Cruce	45
2.6.2.4. Operador de Mutación	46
2.6.3. Diagrama de Flujo Algoritmos Genéticos	47
2.6.4. Algoritmos Genéticos Computacional	48
Capítulo III: Desarrollo de la Investigación	50
3.1. Identificación de la zona de estudio	50
3.2. Metodología	51
3.2.1. Optimización de los parámetros función Gamma para estimar los SPIs utilizando AG	55
3.2.2. Información Climatológica Global	56
3.2.3. Redes Neuronales Artificiales para predicción de sequías	60
3.2.3.1. Diseño de la arquitectura del modelo	60
3.2.3.2. Tipo de Red	60
3.2.3.3. Función de Entrenamiento	61
3.3. Evaluación de desempeño	62
Capítulo IV: Análisis y resultado de la investigación	65
4.1. Análisis exploratorio de las variables regionales	65



4.2. Clasificación Regional cuenca del Mantaro mediante SOM	70
4.3. Análisis regional de frecuencias de la sequía en la cuenca del Mantaro. .	74
4.3.1. Identificación de regiones homogéneas	74
4.3.2. Identificación de la función de distribución de probabilidades .	78
4.3.3. Determinación de Cuantiles	79
4.3.4. Mapeo de precipitaciones para periodos de retorno dados	80
4.4. Análisis estadístico de la precipitación en la región de análisis.	83
4.5. Resultado de la escala e índices de sequía	85
4.5.1. Índices de Vegetación	85
4.5.2. Índices de precipitación estandarizada	88
4.5.3. Índices de precipitación evapotranspiración estandarizada . .	91
4.5.4. Correlación y estimación del mejor índices de sequía	91
4.6. Predicción de sequías en la región de análisis.	93
4.6.1. Selección de datos de ingreso	94
4.6.2. Arquitectura Redes Neuronales Artificiales	95
4.6.3. Resultados Redes Neuronales Artificiales	96
4.6.4. Prueba de hipótesis	100
Conclusiones y recomendaciones	101
Conclusiones	101
Recomendaciones	101
Glosario	103
Acrónimos	105
Símbolos	108
Referencia Bibliográfica	110
Precipitación Percepción Remota	110
Sequías	110
Algoritmos Genéticos	113
Redes Neuronales Artificiales	114
Metodología de Investigación y Estadística	115
Fuentes Electrónicas	115
Anexo A. Análisis Estadístico de precipitaciones	116
Anexo B. Resultados Índices de condición de Vegetación cuenca Mantaro	121
Anexo C. Precipitaciones Mensuales PISCO	126
Anexo D. Índices de Precipitación-Evapotranspiración Estandarizada	130
Anexo E. Resultados L Momentos	136

Índice de tablas

1.1.	Variables e Indicadores del estudio	9
2.2.	Definiciones de sequías comunes	15
2.3.	Valores críticos de discordancia D_i	21
2.4.	Valores críticos de Homogeneidad H_i	22
2.5.	Clasificación de sequías según los Índice de Condición de Vegetación (VCI)	27
2.6.	Clasificación de sequías según los Índices de Precipitación Estandarizada (SPI)	30
2.7.	Clasificación del índice Índice de Aridez (IAR)	32
2.8.	Clasificación del índice Índice de Fournier Modificado (IFM)	33
2.9.	Clasificación del índice Índice de Estacionalidad (IE)	33
2.10.	Comparación entre Redes Neuronales Biológicas y Artificiales	36
2.11.	Técnicas para establecer número de nudos ocultos	40
2.12.	Similitud con Sistemas biológicos	44
2.13.	Librerías Python para la solución de los Algoritmos Genéticos (AG)	49
3.14.	Ubicación de las estaciones meteorológicas	51
3.15.	Distribuciones de probabilidades empíricas	55
3.15.	Continuación de la tabla	56
3.16.	Clasificación de los eventos según el valor del ONI - NOAA en la región del Pacífico tropical 3.4	56
3.16.	Continuación de la tabla	57
3.17.	Clasificación de los eventos según el valor del SOI - Western Regional Climate Center NOAA	57
3.18.	Escala correlación de Pearson	62
4.19.	Correlación de variables regionales	65
4.20.	Variables climáticas regionales y asignación de pesos	75
4.21.	Principales características L-Momentos Regionales de las regiones homogéneas	76
4.22.	Estadísticos H para evaluar la homogeneidad de cada región	76
4.23.	Estadístico Z para evaluar la distribución de probabilidades	78
4.24.	Distribuciones de probabilidades para cada región	79
4.25.	Cuantiles Regionales asociados a distintos períodos de retorno	79
4.26.	parámetros de ajuste curvas regionales	81
4.27.	Estadísticos para las distribuciones Normal y Gamma: Región de análisis	84
4.28.	Valores de parámetros estadísticos función Gamma por AG y MV, escala 6 meses	88
4.29.	Performance del modelo para distintas arquitecturas	95
4.29.	Continuación de la tabla	96



E.1. Resultados L-Momentos	136
E.1. Continuación de la tabla	137
E.1. Continuación de la tabla	138
E.1. Continuación de la tabla	139
E.1. Continuación de la tabla	140
E.1. Continuación de la tabla	141
E.2. Discordancia D_i asociado a las regiones homogéneas	141
E.2. Continuación de la tabla	142

Índice de figuras

2.1.	Climatología mensual de datos grillados - precipitación PISCO 1981-2013	13
2.2.	Estaciones disponibles para la interpolación PISCO	14
2.3.	Características de la sequía para un nivel de truncamiento X_0	17
2.4.	Software IC-RAFT para análisis regional de frecuencias	24
2.5.	Software REFRAN-CV para análisis regional de frecuencias	25
2.6.	Estructura física de una neurona biológica	35
2.7.	Estructura física de una neurona artificial	36
2.8.	Tipos de función de transferencia (a) hard limit, (b) lineal, (c) log sigmoidal y (d) tangente sigmoidal	37
2.9.	Estructura RNA Multicapa	38
2.10.	Arquitectura RNA Multicapa	39
2.11.	Estructura del modelo Self Organizing Map	41
2.12.	Estructura de vecindades	41
2.13.	Codificación Comunes	45
2.14.	Operación Cruce	46
2.15.	Operación Mutación	47
2.16.	Algoritmo genético diagrama de flujo	48
3.17.	Delimitación de la cuenca del Mantaro y ubicación de estaciones pluviométricas	50
3.18.	Esquema metodológico: cálculo de los SPIs óptimos a partir de precipitaciones por sensoramiento remoto	52
3.19.	Esquema metodológico: predicción de sequía	54
3.20.	Localización de Regiones EL NIÑO	58
3.21.	Series de tiempo de los Índices Climáticos Globales	59
3.22.	Diagrama de Taylor	64
4.23.	Distribución altitudinal cuenca del Mantaro	66
4.24.	Distribución de Precipitaciones media anual	67
4.25.	Distribución de Temperatura mínima	67
4.26.	Distribución de Temperatura máxima	68
4.27.	Distribución de Índice de Aridez	69
4.28.	Distribución del Índice de Fournier Modificado	69
4.29.	Distribución del Índice de Estacionalidad	70
4.30.	Topología y conexión de vecindades del modelo auto organizado	71
4.31.	Ventana gráfica Matlab usando SOM	72
4.32.	Distancias y pesos entre neuronas vecinas de los parámetros de ingreso y salida	72
4.33.	Hits del modelo auto organizado y posición de patrones clasificados .	73



4.34. SOM, clasificación regional según el número de neuronas	74
4.35. Ubicación de las regiones homogéneas con sus respectivas estaciones	77
4.36. L Momentos Región 1	78
4.37. Curvas de mejor ajuste de Precipitación media anual regional versus L-CV, L-Skewness, L-Kurtosis	80
4.38. Mapas L-Momentos regionales L-CV, L-Skewness y L-Kurtosis cuenca Mantaro	81
4.39. Mapas de períodos de retorno y PMA	82
4.40. Diagrama de cajas región de análisis de sequías	83
4.41. Gráfico de Cullen y Frey enero-febrero región de análisis	84
4.42. Índice de Condición de Vegetación períodos 2005 y 2006	86
4.43. Identificación de la sequía agrícola mediante el índice VCI cuenca del Mantaro	87
4.44. Índices de precipitación estandarizada por Máxima Verosimilitud y Algoritmos Genéticos	89
4.45. Performance Algoritmos Genéticos Población = 1000 y generación = 100	90
4.46. Índices de precipitación evapotranspiración Escala 6 meses	91
4.47. Correlación entre los índices de sequía SPIMV, SPIAG, SPEI y VCI para diferentes escalas	92
4.48. Diagrama de Taylor SPIMV, SPIAG y VCI para diferentes escalas	92
4.49. Correlaciones cruzadas, variables climáticas globales mes de diciembre y SPIs mes de enero	94
4.50. Arquitectura Redes Neuronales Artificiales Modelo NARX	96
4.51. Curvas de desempeño del modelo NARX, arquitectura 5-3-1-1	97
4.52. Análisis de regresión para el entrenamiento, validación, prueba y toda la serie, arquitectura 5-3-1-1	98
4.53. Autocorrelación de los errores desplazados, arquitectura 5-3-1-1	99
4.54. Series de tiempo entre los SPEIs y errores entre los valores entrenados, arquitectura 5-3-1-1	99
4.55. Serie de tiempo entre los valores simulados y esperados, arquitectura 5-3-1-1	100
A.1. Gráfico de Cullen y Frey región de análisis	117
A.2. Distribuciones Empírica y Teórica Acumulada región de análisis	119
B.1. Índices de condición de Vegetación: 2000-2003	122
B.2. Índices de condición de Vegetación: 2004-2007	123
B.3. Índices de condición de Vegetación: 2008-2012	124
B.4. Índices de condición de Vegetación: 2013-2015	125
C.1. Precipitaciones mensuales PISCO (mm) 1981-01 al 1986-12	127
C.2. Precipitaciones mensuales PISCO (mm) 1987-01 al 1992-12	127
C.3. Precipitaciones mensuales PISCO (mm) 1993-01 al 1998-12	128
C.4. Precipitaciones mensuales PISCO (mm) 1999-01 al 2004-12	128
C.5. Precipitaciones mensuales PISCO (mm) 2005-01 al 2010-12	129
C.6. Precipitaciones mensuales PISCO (mm) 2011-01 al 2015-12	129
D.1. Índices de precipitación y evapotranspiración estandarizada	132



E.1. L Momentos para las 10 regiones clasificadas	143
---	-----

Resumen

El objetivo de la investigación, es determinar el desempeño de la predicción de sequía agrícola regional, mediante el modelo de Redes Neuronales Artificiales tipo NARX, utilizando los índices SPI, SPEI, VCI e Índices Climáticos Globales. Se han determinado 10 regiones homogéneas mediante el análisis regional de frecuencias y L-Momentos, definiendo la región más árida y el índice representativo con su respectiva escala temporal (SPEI 6 meses) que responde al crecimiento y desarrollo de la vegetación en la cuenca con correlación de Pearson igual a 0.58. Las precipitaciones y temperaturas mensuales corresponden a los datos PISCO elaborados por SENAMHI-Perú, con resolución espacial de 0.05°. Para la predicción se han determinado 2 grupos, el primero para la construcción del modelo con 80 % del registro y para la validación del modelo y la hipótesis con el 20 % restante. Los resultados de la predicción han sido satisfactorios aceptándose la hipótesis nula planteada.

Palabras clave : Sequía, redes neuronales artificiales, SPI, SPEI, VCI, análisis regional de frecuencias, L-Momentos.

Abstract

The aim of the study is to determine the performance of the regional agricultural drought prediction by the model of Artificial Neural Networks type NARX, using the SPI, SPEI, VCI and Global Climate Indexes. There have been determined 10 homogeneous regions through regional frequency analysis and L-moments, defining the most arid region and the index representing their respective time scale (SPEI 6 months) which responds to the growth and development of vegetation in the basin correlation Pearson equal to 0.58. Monthly rainfall and temperatures correspond to PISCO data prepared by SENAMHI-Peru, with space resolution of 0.05 degrees. For prediction they have been determined two groups, the first to build the model with 80 % of the registration and validation of the model and the hypothesis with the remaining 20 %. The results have been satisfactory prediction accepting the null hypothesis.

keywords : Drought, artificial neural networks, SPI, SPEI, VCI, regional frequency analysis, L-moments.

Introducción

Actualmente, en muchas partes del planeta, se habla con mucha frecuencia del cambio climático y sus anomalías, dentro de ello se considera las bajas precipitaciones que normalmente se percibe según la estación temporal y espacial que conlleva al fenómeno llamado sequía meteorológica. Éste fenómeno no es reciente, por el contrario, existen evidencias muy marcadas de la presencia de éste en muchas culturas antiguas con resultados muy nefastos en muchos de ellos. Ciertamente, este fenómeno consigna directamente en el ámbito social, económico y político.

Según el Servicio Nacional de Meteorología e Hidrología (SENAMHI), la predicción climática en el Perú, es un trabajo muy complejo, debido a que su territorio está formado alrededor de 27 zonas climáticas que varían de 0 a 6000 metros sobre el nivel del mar. Cada región o cuenca, tiene características propias que varían según los Índices Climáticos Globales (ICG) Oceánicas, Atmosféricas, Telemétricas, la influencia de la Zona de Convergencia Intertropical (ZCI), entre otros. Asimismo, las costas del Perú sufren fenómenos extremos (EL NIÑO y LA NIÑA) bastante recurrentes en estas últimas décadas. Durante EL NIÑO, se tiene precipitaciones muy elevadas provocando inundaciones en la zona Norte y ausencia de precipitaciones en el sur.

En la sequía meteorológica, el parámetro importante corresponde a las precipitaciones mensuales y las estimaciones por sensoramiento remoto es una fuente alterna de información siendo calibradas y validadas debido a la naturaleza indirecta de la medición. Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) y Peruvian Interpolation data of the SENAMHI's Climatological and Hidrological Observations (PISCO), actualmente brindan esta información en una resolución de 0.05 grados latitud y longitud. Funk et al. (2015, p. 10), menciona que CHIRPS valida las precipitaciones para el Perú con 403 estaciones que le brinda la Autoridad Nacional del Agua (ANA) Perú.

SENAMHI, divide en tres zonas a la cuenca del Mantaro, norte, centro y sur, cuya actividad principal es la agricultura y presencia del sector energético debido a su variación en el potencial hídrico. Demostraron los escenarios climáticos en las tres zonas y concluyen que las precipitaciones disminuyen en 10 %, 19 % y 14 %, norte, centro y sur respectivamente, asimismo que en las décadas de los 70 y 80 se presentaron los períodos más lluviosos y durante la década de los 90, las lluvias mensuales disminuyeron significativamente. Por lo mismo, el impacto de las sequías en la cuenca del Mantaro demanda una especial atención.

En la actualidad, el uso de las técnicas de Inteligencia Artificial (IA) en la ingeniería son frecuentes y han demostrado una gran capacidad en el modelado y predicción de



series de tiempo y para la resolución de problemas complejos no estacionarios no lineal. Por ejemplo los Algoritmos Genéticos (AG) son metodologías para la optimización probabilística basada en la teoría de la evolución en sus tres etapas, selección, cruzamiento y mutación. Por otro lado, las Redes Neuronales Artificiales (RNA), especialmente en la predicción de los fenómenos, han mostrado un rendimiento muy bueno (Dastorani & Afkhami, 2011).

Para llegar a la etapa de predicción de sequías en la unidad de estudio, se ha realizado el Análisis Regional de Frecuencias (ARF), empleando la técnica L-Momentos previa regionalización o clustering con metodologías de RNA-Self Organizing Map (SOM) para la reducción de estaciones discordantes y heterogéneas. Asimismo, la determinación del indicador representativo de sequías con su correspondiente escala temporal, incluyendo 4 índices: Índices de Precipitación Estandarizada Máxima Verosimilitud (SPIMV), Índices de Precipitación Estandarizada Algoritmos Genéticos (SPIAG), Índices de Precipitación Evapotranspiración Estandarizada (SPEI) y el Índice de Condición de Vegetación (VCI), éste último considerado como índice patrón que representa la respuesta del crecimiento vegetativo en la cuenca del Mantaro. Finalmente, para la aplicación de la predicción se construye el modelo RNA-Nonlinear Autoregressive Network with Exogenous Inputs (NARX), con característica propia de ser recurrente o red dinámica para la predicción de series de tiempo. El tipo de red corresponde a *NARX* (Red dinámica) y función de entrenamiento *Levenberg-Marquardt*, derivado en la técnica de optimización del método de Newton. Los resultados de entrenamiento, validación y prueba, presentan coeficientes de correlación 0.92, 0.91 y 0.90 respectivamente.

El trabajo de investigación está organizado de la siguiente manera:

- *Capítulo uno - las generalidades*: contiene los antecedentes bibliográficos, descripción de la realidad problemática, formulación del problema, justificación e importancia de la investigación, objetivos, hipótesis, variables e indicadores, tipo y nivel de investigación, periodo de análisis, fuentes de información e instrumentos utilizados, técnicas de recolección.
- *Capítulo dos - marco teórico y marco conceptual*: se desarrolla la fundamentación teórica concerniente a la tesis; definición de precipitación por percepción remota, definición de sequías, análisis regional de sequías, y las metodologías para la etapa de predicción.
- *Capítulo tres - desarrollo del trabajo de tesis*: corresponde a la identificación de la población, muestra de estudio y la metodología para llegar al objetivo planteado en la investigación.
- *Capítulo cuatro - análisis y resultado de la investigación*: presentan los resultados del análisis exploratorio de las variables regionales, resultados de la regionalización y L-Momentos, región de análisis, parámetros óptimos de la función gamma, selección del indicador representativo de sequías, correlaciones entre los ICG y índice representativo y finalmente el modelo de predicción RNA-NARX.
- *Capítulo cinco - conclusiones y recomendaciones*: se presenta las conclusiones de la investigación, en relación a los objetivos establecidos.
- *Anexos*: contiene los resultados gráficos y tabulares documentados.

Capítulo I: Generalidades

1.1. Antecedentes bibliográficos

Los cambios climáticos que desarrolla nuestro planeta y el aumento de la población junto a la explotación de los sectores como la industria, energía y agricultura, han dado lugar una gran demanda de recurso natural del agua.

La sequía es un peligro natural que se vive en todo territorio terrestre, siendo una necesidad urgente para prevenir y preparar estrategias de control como lo vienen haciendo muchos países en Europa, África, América del Norte y Sur.

En Europa se creó un proyecto DROUGHT-R&SPI (Fortering European Drought Research and Science Policy Interface, 2011-2015) (Andreu, Solera, Paredes Arquiola, Haro Monteagudo & Van Lanen, 2015). El proyecto, aborda la necesidad de mejorar la preparación para la sequía a través de un mayor conocimiento, los planes de gestión y una mejor interfaz científico-normativo que reducirá la vulnerabilidad a las futuras sequías y los riesgos que suponen para Europa (Haro Monteagudo et al., 2015).

Desde finales del siglo pasado se han creado y puesto en marcha en Estados Unidos una serie de instituciones, leyes, planes y programas que tienen como finalidad prevenir y mitigar los impactos de la sequía

En América del Norte, se han creado y puesto en marcha una serie de instituciones, leyes, planes y programas que tienen como finalidad prevenir y mitigar los impactos de la sequía. El equipo de gestión de la sequía de Oklahoma 1997, muestra su plan de estrategia según el registro de 1995-1996, que estimuló el desarrollo de este plan, siendo unos de los más graves de la historia.

Son diversos los trabajos de investigación que se han realizado y se vienen realizando a las sequías, caracterizándose cada uno de ellos en la utilización de indicadores, metodologías propuestas como determinísticos, probabilísticos y estocásticos, estudios para sequías puntuales o regionales, etc. Investigadores como Nathaniel B. Guttman (1999), Thomas B. McKee, Doesken y Kleist (1993), Mishra y Singh (2010), Palmer (1965), Serianldi, Bonaccorso, Cancelliere y Grimaldi (2009), Yevjevich (1967), entre otros, aportaron con las definiciones e indicadores para el estudio de las sequías.



1.1.1. Antecedentes en la variable Redes Neuronales Artificiales y Sequías

A nivel internacional, Morid, Smakhtin y Bagherzadeh (2007), realizaron pronósticos de sequías utilizando el modelo de las RNA, para los Índices Efectivos de Sequías (IES) y los SPI. Evaluaron más de 25 modelos para cada indicador de sequía en seis estaciones de control en la Provincia de Tehran de Irán.

Keskin, Terzi, Taylan y Kucukyaman (2011), realizaron predicciones de sequías para el embalse District en Turquía. Usaron RNA, partiendo de los datos de las precipitaciones medias mensuales y luego determinan las series de los índices de SPI. Desarrollaron diferentes modelos de las RNA para predecir los índices del SPI para períodos de 3, 6, 9, y 12 meses.

(Dastorani & Afkhami, 2011), se construyeron diferentes arquitecturas de las RNA, así como diversas combinaciones de parámetros meteorológicos, como la precipitación, las temperaturas máximas, las temperaturas medias, humedad relativa, velocidad media del viento, dirección del viento dominante y evaporación, utilizando como entradas de los modelos desde 1966 hasta 2000. Se encontró que las RNA es una herramienta eficiente para modelar y predecir eventos de sequía.

Belayneh, Adamowski, Khalil y Ozga-Zielinski (2013), compara la eficiencia de 5 modelos para la predicción a largo plazo (sequías de 6 y 12 meses de escala temporal) en la cuenca del río Awash en Etiopía: ARIMA, RNA, Máquinas de Vector Soporte (MVS), Transformaciones Wavelet como pre-proceso de las entradas para los modelos RNA y Transformaciones Wavelet para MVS. Se pronostican los SPI de 12 y 24 meses utilizando el modelo estocástico tradicional ARIMA y se comparan con las técnicas de aprendizaje automático como las RNA y MVS. Los resultados indican que los modelos de predicción de la Red Neuronal con Wavelet fueron los mejores modelos para la predicción de los valores de SPI en múltiples plazos de entrega en la Cuenca del Río de Awash en Etiopía.

Wambua, Mutua y Raude (2014), identificaron los efectos de la sequía en la cuenca alta del río Ta na en Kenia, empleando RNA. El trabajo concluye recomendando establecer una herramienta integrada sobre la base de los datos hidrometeorológicos disponibles para implantar guías mediante programas de mitigación de planificación dentro de la cuenca.

A nivel nacional, Sánchez Romero (2009), propone un modelo de predicción basado en RNA que determina los índices de sequías de los meses de enero a marzo (temporada de lluvias) a partir de las variables climáticas globales observadas durante el mes de diciembre. Se trata de una red multicapa con funciones de activación no lineales, y entrenada con el algoritmo Levenberg-Marquardt.



1.1.2. Antecedentes en la variable Algoritmos Genéticos y Sequías

Hassanzadeh, Abdi Kordani y Fakheri Fard (2012), emplearon los AG, para minimizar los factores de forma y escala en la distribución teórica Gamma que representa la función de densidad de probabilidad. Los autores comparan los resultados con el método de estimación de máxima verosimilitud y los AG para los SPIs de 3, 6, 12, 24 y 48 meses. Los resultados demuestran que los AG obtienen mejor performance, según los coeficientes de eficiencia calculados. Finalmente para la predicción emplean las RNA con transformada wavelet, tomando el 75 % de los datos para el entrenamiento y el resto 25 % para la validación y prueba respectiva.

A nivel nacional no existe ninguna investigación respecto a la optimización metaheurística de parámetros para el cálculo de los SPI.

1.1.3. Antecedentes en la variable Sensoramiento Remoto y Sequías

X. Li, Zhang y Ye (2013), emplearon las precipitaciones satelitales Tropical Rainfall Measuring Mission (TRMM) como alternativas de uso a las estaciones de control pluviométricas de la cuenca Poyang, China, para determinar las condiciones húmedas y secas. Los datos (precipitaciones diarias y mensuales desde 1998 al 2010) fueron validados respecto a 14 estaciones de control. Los resultados demostraron que las precipitaciones TRMM diarias no presentan buenas correlaciones, por lo que no recomiendan el uso respectivo, por otro lado, las precipitaciones TRMM mensuales presentan buenas correlaciones respecto a las estaciones de control. Finalmente concluyen que las precipitaciones TRMM mensuales se pueden usar para el control de la variación y distribución espacial de las condiciones de humedad y sequía en la cuenca Poyang.

Zhang y Jia (2013), integra tres variables satelitales para el monitoreo des sequías meteorológicas a corto plazo. La primera variable corresponde a la integración de la precipitación mensual TRMM-3B43, el segundo es el Advanced Microwave Scanning Radiometer que mide la humedad del suelo y el tercero es AMSR-E, temperatura de la superficie de la tierra, a bordo del satélite EOS.

Mutuga, Nyadawa y Home (2014), realizan el re-escalamiento del TRMM, basándose en el parámetro Índice de Vegetación de Diferencia Normalizada (NDVI) y la precipitación para calcular los SPIs de 3, 6, 9, 12 y 24 meses utilizando los datos del TRMM a escala reducida. Mediante el tratamiento de varianza (ANOVA), se determinan la similitud de los SPIs de tres conjuntos de datos (TRMM re-escalado, TRMM original y Precipitaciones de las estaciones pluviométricas de la cuenca Narumoru-África).

A nivel nacional, Rojas Acuña, Rufasto Campos, Eche Llenque y Carrillo Gomero (2008), propusieron modelos de predicción del inicio de la sequía en la costa norte del Perú empleando la serie de tiempo de los promedios mensuales de las imágenes satelitales NDVI de los sensores Advanced Very High Resolution Radiometer (AVHRR) a bordo de la serie de satélites National Oceanic and Atmospheric Administration (NOAA), El Niño Southern Oscillation (ENSO) y la técnica de regresión lineal múltiple.



Por otro lado, recientemente Climate Hazards Group (CHG), realizan estimaciones cuasi-globales (50S-50N) por satélite y precipitación basados en observación (estaciones) sobre el planeta desde 1981 hasta la actualidad. Apoyados en estos instrumentos el SENAMHI elabora la investigación conocida como PISCO.

1.2. Descripción de la realidad problemática

El fenómeno de la sequía es un factor determinante dentro del desarrollo humano, siendo la agricultura el primer sector económico vinculado a éste. El déficit de agua ofrece muchos factores limitantes para la producción de las cosechas. Lamentablemente y como consecuencia del cambio climático, las sequías se dan con mayor frecuencia en distintas partes del mundo y Perú no está ajena a esta realidad.

Las sequías pueden tener efectos devastadores en el suministro de agua, la producción de cultivos y cría de ganado. Pueden dar lugar a la hambruna, desnutrición, epidemias y desplazamientos de grandes poblaciones de un área a otra.

Las sequías agrícolas de África de la década de 1980 y principios de 1990 han afectado a muchos países y personas y eran probablemente algunas de las situaciones de emergencia de hambruna más impactantes en la historia reciente (Gommes y Petrassi, 1994) citado por (Rojas, Vrieling & Rembold, 2011, p. 343).

En Europa, la sequía es considerado como un peligro natural que ha golpeado duramente en las últimas décadas. Haro Monteagudo et al. (2015) menciona que probablemente va a ser más frecuentes y más graves debido a la mayor probabilidad de inviernos y veranos cálidos del norte del Mediterráneo.

Actualmente gran parte del oeste de los Estados Unidos, incluyendo California, Nevada, Nuevo México y Arizona, y a lo largo de las llanuras del sur de Texas y Oklahoma, experimentan una sequía muy grave, según el Monitor de Sequía de Estados Unidos. Investigadores de la National Aeronautics and Space Administration (NASA), afirman que el suroeste se enfrentarán durante la segunda mitad del siglo XXI a la peor sequía persistente vista en los tiempos antiguos o modernos como consecuencia del calentamiento global.

A nivel local, (Soto P. & Del Caltillo P., 2013, p. 1) representado a la ANA, mencionan que "... entre el año 2000 y 2010 se reportaron 163 eventos de sequías, siendo mayor en la vertiente del Pacífico con 157 eventos, seguidos por la vertiente del Titicaca con 25 eventos y la vertiente del Atlántico con 11 eventos. Dichas sequías han ocurrido con distinta frecuencia, siendo mayor el número las de frecuencia anual y las que ocurren entre 3 y 9 años con 85 eventos y 70 eventos. Asimismo, la sequía del año 2010, fue de las más severas para el oriente peruano. En total el país se afectaron 66,724 familias y 332,087 has".

Por otro lado, en situaciones críticas como los eventos extremos, es decir, las sequías y los períodos de avenidas, la ausencia y la escasa información adiciona un desafío más a las entidades encargadas de mitigar, controlar y preservar los daños y perjuicios, produciendo una disminución de la eficiencia de las medidas tomadas.



1.3. Formulación del problema

1.3.1. Problema principal

- ¿Cuál es el desempeño del modelo *RNA-NARX* en la predicción de sequía agrícola regional, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro?

1.3.2. Problemas secundarios

- ¿Cuál es el índice de sequía más representativo en función al crecimiento y desarrollo de la vegetación dentro de la cuenca del Mantaro?
- ¿Cuál es la relación que existe entre los índices SPI estimados por la técnica numérica Máxima Verosimilitud y la técnica de Inteligencia Artificial Algoritmos Genéticos?
- ¿Cómo es el comportamiento de la sequía regional en la cuenca del Mantaro y cómo se distribuyen las precipitaciones para diferentes períodos de retorno?

1.4. Justificación e importancia de la investigación

El fenómeno de las sequías no se puede evitar, sin embargo, se puede mitigar sus efectos negativos ambientales y socioeconómicos, con una adecuada planificación y gestión a través de un análisis congruente como es el caso de la predicción con metodologías robustas conforme se plantea en el presente trabajo de investigación.

Velasco, Ochoa y Gutiérrez (2005, p. 51), Menciona: "*el investigador científico, ve en la ocurrencia de la sequía la oportunidad de estudiar sus causas, efectos y particularidades, desde una perspectiva objetiva y rigurosamente científica e imparcial, con la aplicación de metodologías fundamentales y consistentes. Con ello pretende probar teorías y aportar explicaciones plausibles del fenómeno y elementos de juicio, que apoyen las decisiones a tomar tanto por un usuario como por administradores del recurso. Sus métodos ayudan al mejoramiento de la planeación y, desde un enfoque imparcial y objetivo buscar y encontrar soluciones aceptables y viables ante situaciones de severidad y riesgo creciente.*"

El mismo Velasco et al. (2005, p. 38) dice "*La vulnerabilidad a la sequía está en relación inversa al grado de desarrollo social y económico de las áreas afectadas: en tanto que para los países más desarrollados, la sequía rara vez representa una amenaza severa, porque disponen de los medios económicos, estructurales y no estructurales para afrontarla, en los de menor desarrollo, una sequía es sinónimo de hambre, desastre y más pobreza.*"

Schwabe, Albiac, Conor, Hassan y Meza Gonzales (2011), investigaron y analizaron la región suroeste más árida dentro de los Estados Unidos (Arizona, California, Colorado, Nevada, Nuevo México, Utah y Wyoming), los mismos que indicaban que la



región sufriría un descenso medio en la precipitación y manifestaron que era probable que experimente una mayor escasez de agua y una mayor susceptibilidad a la sequía. Efectivamente el año en curso California y la región oeste experimenta una sequía grave.

Por lo tanto, una investigación sobre los eventos futuros (predicción) de sequías es fundamental para encontrar soluciones sostenibles para la problemática de manejo del recurso hídrico y una asesoría adecuada en el riesgo de sequía. La sequía a diferencia de otros desastres, evoluciona más lentamente a lo largo del tiempo, sus impactos generalmente se expanden a lo largo de un periodo prolongado de tiempo más allá de la duración del fenómeno. Estas características hacen posible una mitigación más efectiva de sus efectos más adversos, si se tiene un monitoreo de alerta oportuno de una sequía.

Las estimaciones de las precipitaciones por satélite pueden ser utilizados como una fuente alternativa de información, siendo calibrados y validados en áreas específicas, debido a su naturaleza indirecta de las mediciones de la radiación (Masud Cheema & G., 2012, 8). Asimismo, la implementación de las técnicas de IA contribuyen como una herramienta para el cálculos de los indicadores de sequías.

1.5. Objetivos

1.5.1. Objetivos Generales

- Determinar el desempeño del modelo *RNA-NARX* en la predicción de sequía agrícola regional, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro.

1.5.2. Objetivos Específicos

- Determinar el índice de sequía más representativo en función al crecimiento y desarrollo de la vegetación dentro de la cuenca del Mantaro.
- Determinar la relación que existe entre los índices SPI estimados por la técnica numérica Máxima Verosimilitud y la técnica de Inteligencia Artificial Algoritmos Genéticos.
- Analizar el comportamiento de las sequías en la cuenca del Mantaro mediante la técnica de L-Momentos y realizar el mapeo del periodo de retorno de eventos de sequía.

1.6. Hipótesis

- El modelo *RNA-NARX* permite la predicción de la sequía agrícola, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales



como entradas externas, en la cuenca del Mantaro.

1.7. Variables e indicadores

Las variables y los indicadores, se resumen en la tabla 1.1.

Tabla 1.1: Variables e Indicadores del estudio

No	Variable	Indicador
1	Sequías (<i>Dependiente</i>)	Magnitud. Severidad. Intensidad.
2	Índice de Precipitación Estandarizada SPI (<i>Independiente</i>)	Precipitaciones mensuales.
3	Índice de Precipitación Evapotranspiración Estandarizada SPEI (<i>Independiente</i>)	Precipitaciones mensuales. Temperaturas máximas y mínimas mensuales.
4	Índice de Condición de Vegetación VCI (<i>Independiente</i>)	Parámetro de localización espacial - Latitud.
5	Índices Climáticos Globales ICG (<i>Independiente</i>)	Índices de Vegetación de Diferencia Normalizada NDVI. MEI, SOI, ONI, BEST, C, E, PDO, PNA, NIÑO 1+2, NIÑO 3, NIÑO4 y NIÑO 3.4

Fuente: elaboración propia

1.8. Tipo y nivel de investigación

1.8.1. Tipo de Investigación

El presente trabajo de investigación es de tipo *Aplicada no Experimental*, porque busca contribuir a la ampliación del conocimiento científico en la predicción de las sequías.

1.8.2. Nivel de Investigación

De la formulación del problema *¿Cuál es el desempeño del modelo RNA-NARX en la predicción de sequía agrícola regional, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro?*, el nivel de investigación corresponde a *EXPLICATIVO*. El mismo que se encarga de buscar el porqué de los hechos mediante el establecimiento de relaciones



causa-efecto. En este sentido, los estudios explicativos pueden ocuparse tanto de la determinación de las causas (investigación postfacto), como de los efectos (investigación experimental).

1.9. Período de análisis

Corresponde al período de operación y entrega de resultados de los satélites que registra las precipitaciones espaciales de PISCO y PRODUCTO MODIS 13A3. El primero se utiliza para calcular los índices de sequías SPI y SPEI y viene en operación desde enero de 1981 hasta la actualidad. El PRODUCTO MODIS 13A3, se usa para obtener los NDVI e índice VCI, cuyo registro está a disposición desde marzo de 2000 hasta la actualidad.

Por lo tanto, para la presente investigación, se toma como el periodo de análisis desde enero 1981 hasta diciembre de 2015.

1.10. Fuentes de información

1.10.1. Fuentes primarias

También denominada fuente directa. Bajo la solicitud presentada al SENAMHI, se obtuvieron las precipitaciones mensuales de las estaciones pluviométricas de la cuenca Mantaro. De manera alternativa se obtuvieron los datos de las precipitaciones mensuales del proyecto Cachi-Ayacucho y Electrooperú.

1.10.2. Fuentes secundarias

Corresponde a la adquisición de las variables climáticas por sensoramiento remoto como son: las precipitaciones mensuales, temperaturas máximas, mínimas mensuales, NDVI, ICG y las variables espaciales como son: la latitud longitud y elevación de la zona de estudio.

PISCO : Información puesta a disposición en formato raster en la dirección <http://peruclima.pe/> por SENAMHI. La resolución espacial es de $5km^2$ con registros temporales desde 1981 hasta la actualidad.

PRODUCTO MODIS 13A3 : Para obtener los datos NDVI mensuales con resolución espacial de $1km^2$, disponibles en la dirección <https://lpdaac.usgs.gov>

Climate Prediction Center (CPC)-NOAA : Proporcionan información de las variables globales importantes en la variabilidad climatológica como el Oceanic Niño Index (ONI), Pacific Decadal Oscillation (PDO), Southern Oscillation Index (SOI), Atlantic Multidecadal Oscillation (AMO), etc, disponibles en la dirección <http://www.cpc.ncep.noaa.gov>



ASTER GDEM : Modelo de Elevación Digital GLobal del satélite ASTER. Los datos se encuentran en formato Geotiff con resolución espacial de 30 metros y hace referencia al geoide WGS84/EGM96. Para la cuenca del Mantaro se considera los cuadrángulos S11W076, S11W077, S12W075, S12W076, S12W077, S13W074, S13W075, S13W076, S13W077, S14W074, S14W075, S14W076. La dirección electrónica para la descarga es <http://gdem.ersdac.jspacesystems.or.jp/search.jsp>

1.11. Instrumentos utilizados

Para llegar al objetivo de la tesis,

Las aplicaciones utilizadas son

RStudio : Código libre y entorno de desarrollo integrado (IDE) para **R**. Lenguaje de programación para el cálculo de los parámetros óptimos mediante la técnica de los AG, RNA, tratamiento de imágenes satelitales, resultados estadístico, gráficos, entre otros.

Python : Lenguaje de programación para la ejecución del programa Drought Artificial Neural Network and Genetic Algorithm (DANNGA). Apoyo con las distintas librerías para el tratamiento de datos de las precipitaciones satelitales y cálculos del SPI.

QGIS : Sistema de Información Geográfica (SIG) de código libre para plataformas GNU/Linux, Unix, Mac OS, Microsoft Windows y Android. Para la generación de mapas y resultados raster y vector.

ICI-RAFT : Programa desarrollado por el Centro Internacional de Gestión Integrada de recursos Hídricos para el análisis regional de frecuencias.

REFRAN-CV : Software escrito en los lenguajes de programación PHP y R, también para el análisis regional de frecuencias con ciertas características adicionales a ICI-RAFT.

LATEX : Sistema de composición de textos, orientado a la creación de documentos escritos que presenten una alta calidad tipográfica. Para la generación y presentación del informe final.

1.12. Técnicas de recolección

En principio, se realizó los procedimientos administrativos para la adquisición de los datos de las estaciones que monitorea SENAMHI, asimismo la adquisición directa de PISCO, para la variable precipitaciones y temperaturas a nivel mensual.

A través del satélite *terra*, del producto MODIS13A3, se obtuvieron los NDVI para la cuenca del Mantaro, a fin de obtener los VCI y determinar las sequías agrícolas.

Por último, la adquisición de las variables ICG, en tiempo real se obtuvieron desde las direcciones electrónicas del Centro de Predicción Climática de los Estados Unidos.

Capítulo II: Marco teórico y marco conceptual, desarrollo de la teoría y variable

2.1. Precipitación por percepción remota

En la actualidad, existen dos fuentes de estimaciones de precipitación: estaciones pluviométricas y tecnología de teledetección (Shaofeng, Wenbin, Aifeng & Tingting, 2011), el primero corresponde a las estaciones instaladas sobre la superficie terrestre con ubicación puntual y fija, automática o mecánica para el registro y lectura de las precipitaciones y el segundo corresponde a la teledetección siendo éste multipropósito para el registro de precipitaciones, vegetación, topografía y climatología.

Un sistema de teledetección, se encuentra compuesto esencialmente por 2 elementos: la plataforma que corresponde al satélite y el sensor remoto, sin restarle importancia a los demás elementos como el objeto, la fuente de energía, el sistema de transmisión y la función del usuario (Caicedo Carrascal, 2008).

La ventaja de los equipos telemétricos, es que acorta el tiempo de procesamiento de los datos de precipitación (Ponce, 1989). Sin embargo, los pluviómetros son generalmente considerados como instrumentos de mayor precisión y calidad durante la recepción de datos que los equipos radares (Hong & Gourley, 2015, p. 29).

2.1.1. Precipitación CHIRPS

CHIRPS, es un algoritmo para proporcionar información climatológico satelital en el rango latitudinal 50 Sur-50 Norte, denominándose cuasi-global. La resolución de las celdas es de 0.05 grados, registrados desde 1981 hasta la actualidad (precipitaciones). Estos datos están disponibles en varios formatos (NetCDF, TIFF, BIL, PNG) y se encuentran en subdirectorios independientes en las direcciones CHIRPS homepage <http://chg.geog.ucsb.edu/data/chirps>, CHIRPS FAQ http://chg-wiki.geog.ucsb.edu/wiki/CHIRPS_FAQ y CHIRPS paper <http://pubs.usgs.gov/ds/832/>

Dos productos CHIRPS son producidas operacionalmente: una versión preliminar rápida, y una versión final posterior. El producto CHIRPS preliminar está disponible, para todo el dominio, utilizando solamente sistemas de telecomunicaciones y datos de

Conagua (Méjico). El producto CHIRPS final, aprovecha muchas estaciones meteorológicas para la corrección del anterior y proporciona el resultado final en la tercera semana del mes siguiente.

Pete Peterson (2014) menciona que desde la versión 1.8 se ha añadido nuevas estaciones en todo el mundo. En particular, 400 en el Perú, 1200 a través de África de Sharon Nicholson, 11000 en Brasil, más de 1000 en México (Conagua), 76 de África del Sur, 500 en Rusia, 250 en Asia Central y 400 a través de América Central. Funk et al. (2015, p. 10), muestra la validación realizada para el Perú, mediante 403 estaciones proporcionadas por la ANA.

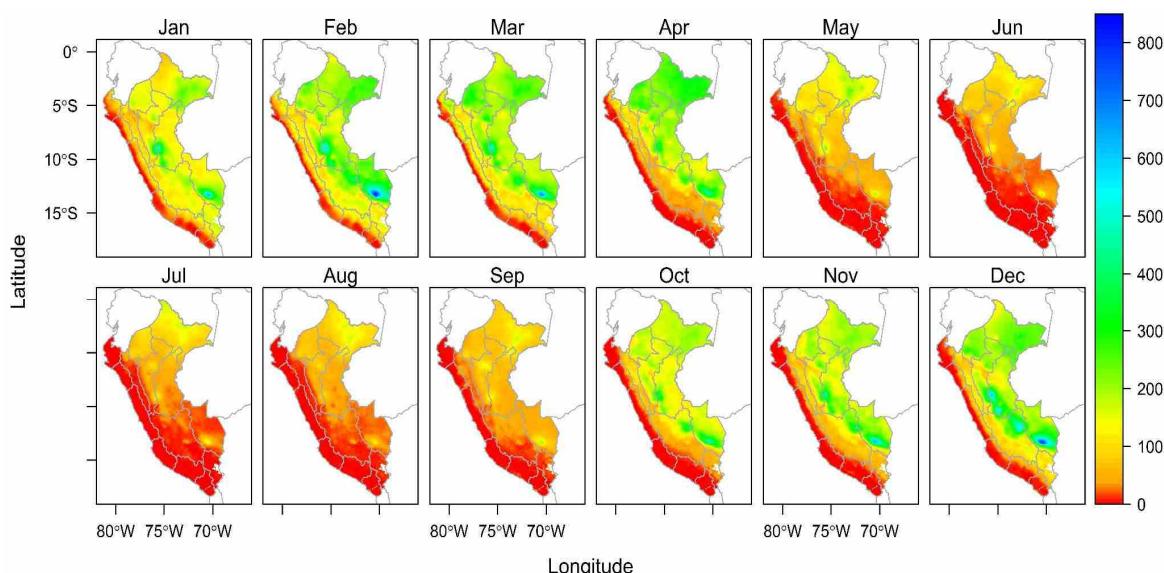
2.1.2. Precipitación PISCO

PISCO es acrónimo de **P**eruvian **I**nterpolation **D**ata of the **S**ENAMHI's **C**limatological and **H**idrological **O**bservations, es producto de la investigación realizada por SENAMHI.

Para la obtención de las precipitaciones mensuales con corrección (P_c), se emplea la ecuación 2.1, el mismo que consiste en restar la precipitación CHIRPS (P_{CHIRPS}) y los valores de interpolación (IP) de la diferencia de P_{CHIRPS} y precipitaciones de las distintas estaciones registradas por SENAMHI ($P_{SENAMHI}$) según el esquema 2.2.

$$P_c = P_{CHIRPS} - IP(P_{CHIRPS} - P_{SENAMHI}) \quad (2.1)$$

Figura 2.1: Climatología mensual de datos grillados - precipitación PISCO 1981-2013



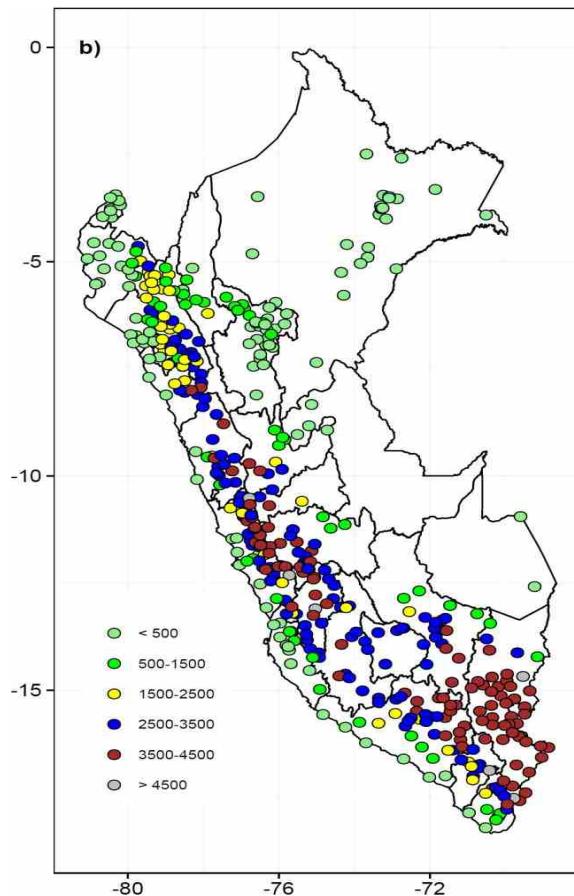
Fuente: Adaptado desde Seminario: Estudios e investigaciones Hidrológicas SENAMHI, enero de 2015 - Waldo Lavado

Los resultados de las variables climatológicas como la evaporación, precipitaciones mensuales, temperatura mínima, temperatura máxima y los índices de precipi-

tación estandarizada son puestas a disposición en formatos raster en la dirección <http://peruclima.pe/>, desde el año 1981 hasta la actualidad.

La fuente de información PISCO, por las operaciones expuestas anteriormente, se encuentran calibradas y validadas dentro del territorio peruano.

Figura 2.2: Estaciones disponibles para la interpolación PISCO



Fuente: Adaptado desde Seminario: Estudios e investigaciones Hidrológicas SENAMHI, enero de 2015 - Waldo Lavado

2.2. Definición de sequía

Existen diferentes concepciones de sequía, debido a la gran diversidad y puntos de vista del estudio de sequías (Yevjevich, 1967). A diferencia de las avenidas, las sequías no tienen una definición universal (Nagaranjan, 2009).

Wilhite y Glantz (1985), categoriza la definición de las sequías en *conceptual* y *operacional*, coincidiendo en muchas literaturas en la definición conceptual como "*deficiencia de precipitaciones durante un periodo de tiempo relativamente prolongado*" (Marcos, 2001). Mientras la definición operacional trata de identificar el inicio, la gravedad, y la terminación de los períodos de sequía. Por lo tanto, ésta última definición, se puede utilizar para analizar la frecuencia de las sequías, la gravedad y la duración de un



período de retorno dado (Mishra & Singh, 2010).

Tabla 2.2: Definiciones de sequías comunes

Organización)/Autor	Año	Concepto de sequía
The World Meteorological Organization	1986	"La sequía significa una deficiencia prolongada, extendida en la precipitación".
The UN Convention to Combat Drought and Desertification	1994	"La sequía significa el fenómeno que se produce naturalmente cuando las lluvias han sido considerablemente inferiores a los niveles normales registrados, causando graves desequilibrios hidrológicos que afectan negativamente a la tierra y a los sistemas de producción de recursos".
The Food and Agriculture Organization	1983	"La sequía es el porcentaje de años, cuando los cultivos se pierden por la falta de humedad".
The encyclopedia of climate and weather	1996	"La sequía es un período prolongado, estacional, anual o varios años de la escasez de lluvias en relación con la estadística multianual para una región".
Gumbel	1963	"La sequía es el valor anual más pequeña del caudal diario".
Palmer	1965	"Una desviación significativa de las condiciones hidrológicas normales de un área".
Linseley et al	1959	"La sequía es un período sostenido de tiempo sin lluvias importantes".

Fuente: adaptado desde Mishra y Singh (2010, p. 206)

De acuerdo a las definiciones, Wilhite y Glantz (1985) clasifica las sequías en 4

Sequía Meteorológica Corresponde a la escasez de la **precipitación**, relacionado directamente con los índices globales del océano y la atmósfera. Por lo tanto, las sequías meteorológicas están relacionadas a los datos climáticos, y en la mayoría de los casos, presentan información específica para cada región particular que varía en función de las características del clima regional (Marcos, 2001). Asimismo, es considerado como el inicio a las sequías agrícolas, hidrológicas y socio económicas respectivamente. De acuerdo a la OMM, Svoboda, Hayes y Wood (2012), se puede estudiar con el SPI de 1 ó 2 meses.



Sequía Agrícola Se presenta debido a la escasez de agua en el suelo y es susceptible al normal desarrollo de los cultivos. La deficiencia de la humedad del suelo impide el desarrollo de una determinada planta o cultivo en sus diferentes fases de crecimiento. Por lo general, las sequías agrícolas, son secuelas de las sequías meteorológicas a corto plazo presentando una reducción de los productos agrícolas y pérdidas de plantas. De acuerdo a la OMM et al. (2012), se puede estudiar con el SPI de 1 y 6 meses.

Sequía Hidrológica Corresponde a la escasez del caudal o volumen de agua superficial y sub superficial y refleja disminución de los niveles de escorrentía (ríos), volúmenes de embalses, lagunas, en la ausencia de humedad del suelo y en la disminución de los niveles del agua subterránea, siendo secuencia de las sequías agrícolas a mediano o largo plazo. De acuerdo a la OMM et al. (2012), se puede estudiar con el SPI entre unos 6 y 24 meses o más.

Sequía Socio Económica Está vinculada a la sequías agrícola e hidrológica y se produce cuando se presenta la escasez del agua hasta provocar impactos negativos económicos de una población. Es el resultado de las sequías mencionadas.

Característica de la sequía Yevjevich (1967), identificó tres parámetros de sequía importantes: la duración (L), la severidad (D) y la intensidad (I) a partir de un nivel de truncamiento o umbral (X_0) definido.

En la figura 2.3, se muestran los tres parámetros:

Duración L : Se conoce también como la longitud del run (-), expresado en meses o años consecutivos.

Severidad D : Indica la acumulación de la deficiencia de una sequía según el umbral X_0 definido. Es decir, la suma de los déficit durante la longitud o duración.

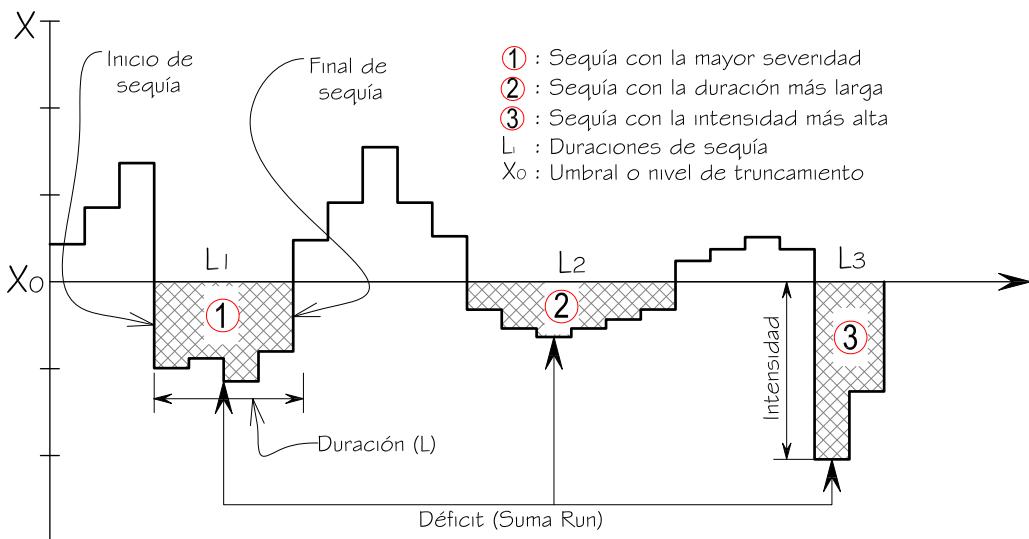
Intensidad I : Es la razón entre la severidad y la duración.

2.3. Análisis Regional de Frecuencias L-Momentos para Sequías

El Análisis Regional de Frecuencias-L Momentos (ARF-LM), está basado en la metodología propuesta por Hosking y Wallis (1997), quienes justifican el método en la estimación de las magnitudes (cuantiles) de los eventos extremos (avenidas y sequías) correspondientes a diferentes períodos de retorno.

Muchos autores como Acuña, Felipe, Ordoñez y Arboleda (2011), N. B. Guttman, Hosking y Wallis (1993), Núñez et al. (2011), Wallis, Schaefer, Barker y Taylor (2007), Yang et al. (2010), han reconocido ventajas del ARF-LM para la caracterización de sequías.

Figura 2.3: Características de la sequía para un nivel de truncamiento X_0



Fuente: Adaptado desde Mishra y Singh (2010, p. 209)

Núñez et al. (2011, p. 353), cita a los trabajos recientes del método ARF-LM para preparar los mapas de sequías: Estados Unidos, Werick (1995); México, Hallack-Alegria y Watkin (2007); Turquía, Yurekli y Anli (2008); Alemania, Demuth y Külls (1997); en Nueva Zelanda, Pearson (1995), entre otros.

2.3.1. Definición de L-Momentos

Los L-Momentos, son combinaciones lineales de momentos ponderados de probabilidad, que miden la localización, escala y forma de la distribución de probabilidades. Hosking y Wallis (1997), mencionan que esta técnica, tiene la capacidad de caracterizar un mayor rango de funciones de distribución, considerando mayor robustez en presencia de outlier y menor sensibilidad a la asimetría y entre otros, por lo tanto, concluyen que al utilizar éste método, se obtienen estimaciones más cercanas a los valores observados.

Los momentos ponderados de probabilidad son

$$\beta_r = \frac{1}{N} \sum_{j=r+1}^N (x_j) \frac{(j-1)(j-2)\dots(j-r)}{(N-1)(N-2)\dots(N-r)} \quad (2.2)$$

los cuatro primeros son

$$\beta_0 = \frac{1}{N} \sum_{j=1}^N x_j \quad (2.3)$$

$$\beta_1 = \frac{1}{N} \sum_{j=2}^N x_j \frac{j-1}{N-1} \quad (2.4)$$



$$\beta_2 = \frac{1}{N} \sum_{j=3}^N x_j \frac{(j-1)(j-2)}{(N-1)(N-2)} \quad (2.5)$$

$$\beta_3 = \frac{1}{N} \sum_{j=4}^N x_j \frac{(j-1)(j-2)(j-3)}{(N-1)(N-2)(N-3)} \quad (2.6)$$

Análogamente, los cuatro primeros momentos lineales están definidos como:

$$\begin{aligned} l_1 &= \beta_0 \\ l_2 &= 2\beta_1 - \beta_0 \\ l_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ l_4 &= 6\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \quad (2.7)$$

Por otro lado, están los ratios o denominados coeficientes de momentos lineales, que sirven para hacer comparaciones con diferentes estaciones y definen los parámetros de escala y forma de una distribución de probabilidad.

$$\tau_r = \frac{l_r}{l_2} \quad (2.8)$$

siendo los más importantes

$$\begin{aligned} L - CV &= \tau = \frac{l_2}{l_1} \\ L - Skewness &= \tau_3 = \frac{l_3}{l_2} \\ L - Kurtosis &= \tau_4 = \frac{l_4}{l_2} \end{aligned} \quad (2.9)$$

El primer L-Momento τ , es la medida del coeficiente de variación ($L - CV$). τ_3 es la medida del sesgo ($L - Skewness$) y τ_4 es la medida de la curtosis ($L - Kurtosis$). De esta manera los L-Momentos, se utilizan para estimar los parámetros de ajuste de una distribución de probabilidad.

Logística Generalizada (GLO)

- Definición: Tiene 3 parámetros: ξ (localización), α (escala) y k (forma)

$$f(x) = \frac{\alpha^{-1} e^{-(1-k)y}}{(1 + e^{-y})^2}; \quad y = \begin{cases} \frac{-1}{k} \log \left[1 - \frac{k(x-\xi)}{\alpha} \right]; & k \neq 0 \\ \frac{x-\xi}{\alpha}; & k = 0 \end{cases} \quad (2.10)$$



$$F(x) = \frac{1}{1 + e^{-y}} \quad (2.11)$$

$$x(F) = \begin{cases} \xi + \frac{\alpha [1 - (\frac{1-F}{F})^k]}{k}; & k \neq 0 \\ \xi - \alpha \log(\frac{1-F}{F}); & k = 0 \end{cases} \quad (2.12)$$

- L-Momentos: vienen definidos para $-1 < k < 1$

$$l_1 = \xi + \alpha \left[\frac{1}{k} - \frac{\pi}{\sin(k\pi)} \right]; \quad l_2 = \frac{\alpha k \pi}{\sin(k\pi)}; \quad \tau_3 = -k; \quad \tau_4 = \frac{1 + 5k^2}{6} \quad (2.13)$$

- Parámetros:

$$k = -\tau_3; \quad \alpha = \frac{l_2 \sin(k\pi)}{k\pi}; \quad \xi = l_1 - \alpha \left[\frac{1}{k} - \frac{\pi}{\sin(k\pi)} \right] \quad (2.14)$$

Valores Extremos Generalizado (GVE)

- Definición: Tiene 3 parámetros: ξ (localización), α (escala) y k (forma)

$$f(x) = \alpha^{-1} e^{-(1-k)y - e^{-y}}; \quad y = \begin{cases} \frac{-1}{k} \log \left[1 - \frac{k(x-\xi)}{\alpha} \right]; & k \neq 0 \\ \frac{x-\xi}{\alpha}; & k = 0 \end{cases} \quad (2.15)$$

$$F(x) = e^{-e^{-y}} \quad (2.16)$$

$$x(F) = \begin{cases} \xi + \frac{\alpha [1 - (-\log F)^k]}{k}; & k \neq 0 \\ \xi - \alpha \log(-\log F); & k = 0 \end{cases} \quad (2.17)$$

La GEV, se divide en tres tipos:

Tipo I : $F(x) = \exp(e^{-x})$, si $-\infty < x < \infty$

Tipo II : $F(x) = \exp(x^{-\delta})$, si $0 \leq x < \infty$

Tipo III : $F(x) = \exp(-|x|^{-\delta})$, si $-\infty < x \leq 0$

- L-Momentos: vienen definidos para $k > -1$

$$\begin{aligned} l_1 &= \xi + \alpha \frac{1 - \Gamma(1+k)}{k} \\ l_2 &= \frac{\alpha(1 - 2^{-k})\Gamma(1+k)}{k} \\ \tau_3 &= \frac{2(1 - 3^{-k})}{1 - 2^{-k}} - 3 \\ \tau_4 &= \frac{5(1 - 4^{-k}) - 10(1 - 3^{-k}) + 6(1 - 2^{-k})}{1 - 2^{-k}} \end{aligned} \quad (2.18)$$



- Parámetros:

$$k \approx 7.8590C + 2.9554C^2; \quad C = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3}$$
$$\alpha = \frac{l_2 k}{(1 - 2^{-k})\Gamma(1 + k)}; \quad \xi = l_1 - \alpha \frac{1 - \Gamma(1 + k)}{k} \quad (2.19)$$

En la sección apéndice Hosking y Wallis (1997, pp. 191-209), muestra 11 funciones de distribuciones de probabilidad, cuyas ecuaciones, algoritmos y parámetros están demostradas para cada uno de ellos.

2.3.2. Procedimiento de Análisis Regional de sequías L-Momentos.

Para la aplicación del ARF-LM del presente trabajo de investigación, se utiliza la *Guía metodológica para la aplicación del Análisis Regional de Frecuencia de Sequías basado en L-momentos y resultados de aplicación en América Latina* de Nuñez y Verbist (2010), publicada a través del Programa Hidrológico Internacional (PHI) de la Oficina Regional de Ciencia para América Latina y el Caribe de la Organización de las naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y siendo una extensión a la metodología propuesta por Hosking y Wallis (1997). Considera los siguientes pasos:

Análisis Exploratorio de datos. Previo a los análisis, es importante conocer la estadística del conjunto de datos para la evaluación de las sequías. Los datos para la presente tesis corresponden a las precipitaciones mensuales espaciadas en grillas de 0.05 grados.

En este procedimiento, se verifican la existencia de datos anómalos (outliers) sometiéndose a diferentes test estadísticos como *Mann Kendall* (Estacionariedad), *Spearman* (Coeficiente de correlación) *Durbin-Watson* o *Randomness* (Prueba de independencia), entre otros, para asegurar la adecuada homogeneidad en el ARF-LM.

Identificación de regiones homogéneas. A través de algunas variables de agrupamiento tales como latitud, longitud, altitud, precipitaciones medias anuales, temperaturas mínimos, máximos y promedios, se realizada la etapa de regionalización o también denominado clustering regional, usando metodologías como SOM, K-means, Ward, entre otros, para representar un grupo de estaciones como regiones con características comunes y simplificar grandes cantidades de información para el ARF-LM.

La identificación de regiones homogéneas, es la etapa principal del ARF-LM, ya que según esté procedimiento se establecerá el número de regiones de análisis, por lo tanto, para conseguir tal *homogeneidad* (similitud bajo las características externas de cada estación), se consideran la evaluación del test de discordancia y homogeneidad.

- Prueba de discordancia: Hosking y Wallis (1997), desarrollaron una medida de discordancia D_i que evalúa el grado en que los L-Momentos de una estación (grilla



o celda), difieren de los L-Momentos regionales. Las causas son múltiples

$$\mathbf{u}_i = \begin{pmatrix} \tau^i \\ \tau_3^i \\ \tau_4^i \end{pmatrix} \quad (2.20)$$

donde \mathbf{u}_i es un vector columna que contiene los coeficientes L-Momentos τ , τ_3 y τ_4 de cada estación i .

El valor promedio de la región o grupo es:

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i \quad (2.21)$$

Por lo tanto, la medida de la discordancia D_i para la estación i es definido por:

$$D_i = \frac{N}{3} (\mathbf{u}_i - \bar{u})^T \mathbf{A}^{-1} (\mathbf{u}_i - \bar{u}); \quad \mathbf{A} = \sum_{i=1}^N (\mathbf{u}_i - \bar{u})(\mathbf{u}_i - \bar{u})^T \quad (2.22)$$

Tabla 2.3: Valores críticos de discordancia D_i

No de estaciones	D_i	No de estaciones	D_i
5	1.333	11	2.632
6	1.648	12	2.757
7	1.917	13	2.869
8	1.140	14	2.971
9	2.329	15	3.000
10	2.491	-	-

Fuente: adaptado desde Nuñez y Verbist (2010, p. 26)

de la tabla 2.3, se considera estaciones discordantes cuando $D_i > 3$ (mayores a 15 estaciones).

- Prueba de Homogeneidad: Para la aceptación de una región *homogénea*, se aplica el estadístico H_1 propuesto por Hosking y Wallis (1997) y ampliada por Wallis et al. (2007). Este estadístico, observa la variabilidad de las estaciones dentro de un grupo o región. Los promedios regionales de L-CV, L-skewness y L-kurtosis, representados por τ^R , τ_3^R , τ_4^R respectivamente son calculados como:

$$\tau^R = \frac{\sum_{i=1}^N n_i \tau^i}{\sum_{i=1}^N n_i}; \quad \tau_3^R = \frac{\sum_{i=1}^N n_i \tau_3^i}{\sum_{i=1}^N n_i}; \quad \tau_4^R = \frac{\sum_{i=1}^N n_i \tau_4^i}{\sum_{i=1}^N n_i} \quad (2.23)$$

Existen tres medidas de dispersión para medir la heterogeneidad, H_1 H_2 y H_3



que están en función de V_1 , V_2 y V_3 relacionados al coeficiente de variación, sesgo y curtosis respectivamente.

$$\begin{aligned} V_1 &= \sqrt{\frac{\sum_{i=1}^N n_i (\tau^i - \tau^R)^2}{\sum_{i=1}^N n_i}} \\ V_2 &= \frac{\sum_{i=1}^N n_i \sqrt{(\tau^i - \tau^R)^2 + (\tau_3^i - \tau_3^R)^2}}{\sum_{i=1}^N n_i} \\ V_3 &= \frac{\sum_{i=1}^N n_i \sqrt{(\tau_3^i - \tau_3^R)^2 + (\tau_4^i - \tau_4^R)^2}}{\sum_{i=1}^N n_i} \end{aligned} \quad (2.24)$$

Estas relaciones se ajustan a la distribución Kappa (Hosking & Wallis, 1997) y se realiza un número de simulaciones de una región con N estaciones. Para cada simulación se calcula V y se determina además la media μ_ν y la desviación estándar σ_ν .

$$\begin{aligned} H_1 &= \frac{V_1 - \mu_\nu}{\sigma_\nu} \\ H_2 &= \frac{V_2 - \mu_{\nu 2}}{\sigma_{\nu 2}} \\ H_3 &= \frac{V_3 - \mu_{\nu 3}}{\sigma_{\nu 3}} \end{aligned} \quad (2.25)$$

El criterio de homogeneidad, será a través del valor H_1 , el cual será utilizada según Wallis et al. (2007), de la tabla 2.4, quienes consideran como región homogénea para $H_1 < 3$.

Tabla 2.4: Valores críticos de Homogeneidad H_i

Hosking y Wallis (1997)	Wallis, Schaefer, Barker y Taylor (2007)	Condición
$H_1 < 1$	$H_1 < 2$	Región Homogénea.
$1 \leq H_1 < 2$	$2 \leq H_1 < 3$	Región posiblemente Heterogénea.
$H_1 \geq 2$	$H_1 \geq 3$	Región Heterogénea.

Fuente: adaptado desde Nuñez y Verbist (2010, p. 25)

Identificar la función de distribución de probabilidad. Realizada la clasificación regional o la identificación de grupos homogéneos, es posible determinar las funciones de distribuciones de probabilidad para cada región homogénea. Hosking y Wallis (1997), consideran entre los más importantes a cinco distribuciones como: Logística Generalizada (GLO), Valores Extremos Generalizado (GEV), Pareto Generalizado (GPA), Normal Generalizado (GNO) y Pearson tipo III (PE3).



La aplicación de las distribuciones mencionadas y extendidas están codificadas en los paquetes *lmom*¹ y *lmomRFA*² del lenguaje de programación R, cuyo autor es J.R.M. Hosking.

La medida de bondad de ajuste para una distribución asignada, se expresa como:

$$Z^{DIS} = \frac{\tau_4^R - \tau_4^{DIS}}{\sigma_4} \quad (2.26)$$

donde: τ_4^R es L-kurtosis promedio regional, τ_4^{DIS} L-Kurtosis de la distribución adaptada, σ_4 es la desviación estándar promedio regional, y $|Z^{DIST}|$ es el estadístico de ajuste.

El ajuste es adecuado, si el estadístico $|Z^{DIST}| \leq 1.64$ para un grado de significación estadístico del 90 %.

Determinar los cuantiles. Para determinar los cuantiles de una estación, previa la identificación de la distribución de probabilidades, se usa la ecuación:

$$Q^{(i)}(F) = l_1^{(i)} q(F) \quad (2.27)$$

donde: $l_1^{(i)}$, es la media de los datos observados en la estación i , $q(F)$ es la Curva de Crecimiento Regional (CCR) que varía de 0 a 1 y $Q^{(i)}$, es el cuantil de la estación i . Para el estudio de sequías, Núñez et al. (2011), Nuñez y Verbist (2010) aplicaron el 40 % de la precipitación normal, es decir, con 60 % de déficit.

Mapeo del periodo de retorno de eventos de sequía Para cada estación, pixel, o grilla, se establece una probabilidad de ocurrencia o periodo de retorno ($Tr = \frac{1}{p}$ evento extremo izquierdo) que está asociada a una función de distribución de probabilidades (Nuñez & Verbist, 2010).

Wallis et al. (2007, p. 424), recomienda tres ecuaciones para estimar la relación que existe entre los L-Momentos y el índice de avenida, considerado comúnmente la Precipitación Media Anual Regional (PMAR).

$$\text{L-Momentos Ratios} = Ae^{-B(MAPR)} + C \quad (2.28)$$

$$\text{L-Momentos Ratios} = A + B \ln(MAPR) \quad (2.29)$$

$$\text{L-Momentos Ratios} = A(MAPR)^2 + B(MAPR) + C \quad (2.30)$$

donde A , B y C , son coeficientes que son determinados por técnicas de optimización como mínimos cuadrados, evaluando el error de la raíz cuadrática media o cualquier otro

¹Dirección electrónica: <https://cran.r-project.org/web/packages/lmom/index.html>

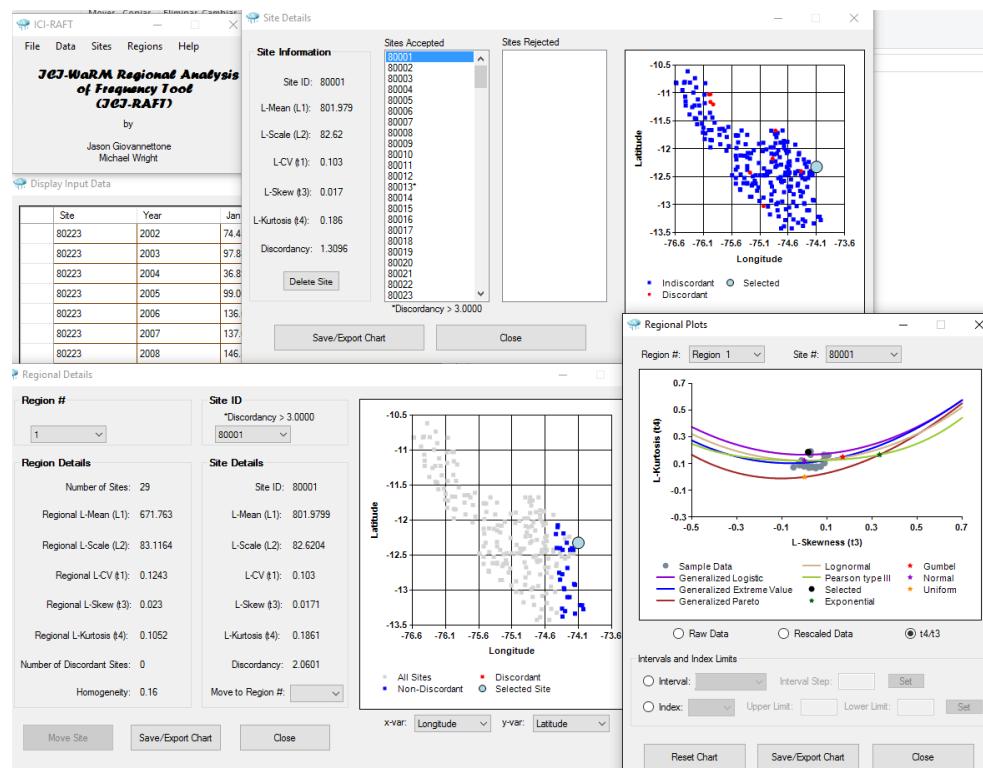
²Dirección electrónica: <https://cran.r-project.org/web/packages/lmomRFA/index.html>

estadístico. y los L-Momentos, son los momentos lineales correspondientes al coeficiente de variación (L-CV), coeficientes de asimetría (L-Skewness) y coeficientes de Kurtosis (L-Kurtosis).

2.3.3. Software L-Momentos

ICI-RAFT : Programa desarrollado por el Centro Internacional de Gestión Integrada de Recursos Hídricos (ICIWaRM) para calcular la *frecuencia-intensidad* de un evento extremo (inundaciones o sequía), haciendo uso del ARF mediante el método L-Momentos, propuestos por Hosking y Wallis (1997). El programa ICI-RAFT está escrito en el lenguaje de programación Visual Basic 2010 y tomada del código fuente de Hosking y Wallis (1997) escrita en Fortran.

Figura 2.4: Software IC-RAFT para análisis regional de frecuencias



Fuente: Software IC-RAFT

Es posible descargar de manera gratuita desde la dirección: <http://www.gwadi.org/tools/ici-raft> y sigue los pasos:

- Generación de los datos de acuerdo al formato proporcionado. Los datos de las precipitaciones son mensuales y con sus respectivos códigos de estación. Por otro lado, las características para considerar la regionalización son: Latitud, Longitud, Altitud, temperaturas , precipitación media anual, entre otras variables externas.
- Se carga la base de datos de precipitaciones mensuales y sus respectivas características.



- Preparación de datos, que consiste en filtrar los datos tales como: inicio del mes de análisis, duración en meses, el número mínimo de registros, las unidades de cada variable consideradas como características externas.
- Análisis de discordancia de las estaciones. Los resultados son los test de discordancia respecto a toda la cuenca en estudio. Se recomienda la observación y verificación de las estaciones con discordancia mayores a 3.
- Haciendo uso de las características de las estaciones, se realiza la regionalización, asignando un factor peso a cada variable según su influencia particular.
- Verificación del test de homogeneidad según Hosking y Wallis (1997). H_1 mayores a 3, **NO SE ACEPTE** la región como homogénea, sino como heterogénea, por lo tanto, reasignar los pesos o mover estaciones entre unas y otras para obtener el test H_1 menores a 3.
- Obtención del análisis regional de frecuencias para cada región homogénea y evaluación de los cuantiles para diferentes períodos de retorno o viceversa.

REFRAN-CV : Es un software escrito en los lenguajes de programación PHP y R, que hace uso del método L-Momentos para estimar las diferentes funciones de distribución de probabilidades de las precipitaciones.

Figura 2.5: Software REFRAN-CV para análisis regional de frecuencias



Fuente: Software REFRAN-CV



El software es de libre distribución desde la dirección electrónica <http://www.aquaknow.net/es/tag/tags/refran-cv> y está compuesto por seis módulos:

- Importación de datos y pre procesamiento. Verifica la correcta estructura de la base de datos para continuar con el módulo siguiente.
- Definición de regiones homogéneas. Agrupa en regiones, según las características de las variables climáticas.
- Análisis regional de frecuencias. En cada región se define la función de probabilidad.
- Parámetros de interpolación. De acuerdo a una ecuación regional, se definen los parámetros entre los L-Momentos y la variable principal de estudio (precipitación media anual).
- Generación de Mapas de los L-Momentos.
- Cartografía. Corresponden a la generación de mapas respecto a los períodos de retorno asignados.

2.4. Índices de sequía

Los índices de sequía vienen a ser mediciones cuantitativas que caracterizan los niveles de sequía, percibiendo datos de uno o más variables tales como las precipitaciones, temperatura o evapotranspiración (Zargar, Sadiq, Naser & Khan, 2011).

En la actualidad existen una gran cantidad de indicadores de sequías. Zargar et al. (2011, p. 333), indican que existen más de 100 índices de sequía ya sea para los tipos meteorológicos, agrícolas e hidrológicos, los mismos que describen 74 índices operacionales en el artículo “A review of drought indices”.

Por otro lado, Vicente Serrano et al. (2012, p. 687) realizaron un estudio para evaluar diferentes índices de sequía (SPI, SPEI y Índice de Severidad de Sequía de Palmer (PDSI)) para la cuantificación de caudales, humedad del suelo y crecimiento de vegetación. Encontraron "que la correlación más alta se produce entre 70-90 % de los casos con el SPI y SPEI en función de la variable y la estación del año, mientras que los índices de PDSI representan menos del 15 %". También manifiestan que existen diferencias pequeñas entre el SPI y SPEI (10 % superior para el SPEI).

En la presente tesis, se consideran los índices VCI, SPI y SPEI conforme se define a continuación.

2.4.1. Índice de Condición de Vegetación (VCI)

El VCI es definido por Kogan y Sullivan (1993), y es aplicado para la evaluación de sequías donde se manifiestan los cambios del proceso de *vegetación*, a través de la escala 0 a 1, siendo 0 la condición extremadamente malo y el valor de 1 condiciones óptimas.



$$VCI_i = \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (2.31)$$

donde el VCI es el índice de la condición de vegetación y el NDVI es el índice normalizado de vegetación, $NDVI_{min}$, $NDVI_{max}$ son los mínimos y máximos calculados por los pixeles correspondientes de cada mes de análisis. El NDVI se calcula mediante la ecuación siguiente:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2.32)$$

donde NIR , es la reflectancia espectral del canal infrarrojo cercano y RED es la reflectancia en el canal rojo del visible. El NDVI es un indicador para conocer la cantidad, calidad y el desarrollo de áreas de vegetación en función a la medición mediante senso-ramientos remotos espaciales. Los valores de éste índice son adimensionales, donde nos indican el vigor de la clorofila de la vegetación: valores cercanos a cero indican vegetación poco vigorosa o simplemente ausencia de vegetación y por el contrario, valores cercanos a 1 vegetación muy vigorosa.

Rojas Acuña et al. (2008), analizaron el NDVI de todo el Perú en pixeles de $8 \times 8 \text{ km}^2$, encontrando valores desde 0.0 hasta 0.7. En la costa el promedio es bastante bajo, siendo cercano a 0. En la sierra desde 0.4 (zona norte) hasta 0.1 (sur) y en la selva desde 0.5 hasta 0.7.

Los resultados de la validación del VCI han demostrado tener una excelente capacidad de detectar sequías y la medición en el tiempo de su aparición, duración, intensidad, y el impacto sobre la vegetación (Kogan, 1995).

Tabla 2.5: Clasificación de sequías según los VCI

VCI	Clasificación
0.90 a 1.00	Humedad Extrema
0.80 a 0.90	Humedad Severa
0.60 a 0.80	Humedad Moderada
0.50 a 0.60	Humedad Leve
0.40 a 0.50	No Sequía
0.30-a 0.40	Sequía Leve
0.20 a 0.30	Sequía Moderada
0.10 a 0.20	Sequía Severa
<0.1	Sequía Extrema

Fuente: Elaboración propia

Quiring y Ganesh (2010, p. 330), menciona que una de las principales ventajas del VCI es que, debido a su producto satelital, puede proporcionar datos en tiempo real con una resolución espacial relativamente alta. Además, el VCI utiliza una metodología completamente independiente para el monitoreo de la sequía, mientras que todos los



otros índices dependen, en cierta medida, en los datos meteorológicos basado en una estación o procedimientos estadístico.

Muchos autores como Du et al. (2013), Quiring y Ganesh (2010), Vicente-Serrano (2006), entre otros, han correlacionado el índice VCI con otros índices, tomando como índice patrón al VCI. Por lo tanto, en la presente investigación se correlacionan el VCI, y los índices SPI y SPEI, para determinar la escala temporal y su relación con la vegetación en la unidad de estudio.

2.4.2. Índice de Precipitación Estandarizada (SPI)

El índice SPI fue propuesto por McKee et al 1993 y se ha popularizado debido a su planteamiento teórico y por usar sólo el registro mensual de precipitaciones preferentemente con más de 30 años de registro histórico (Campos, 2014).

La Organización Meteorológica Mundial (OMM), menciona que a lo largo de los años se han debatido mucho sobre qué índices se sequía deberían utilizarse en determinados climas, al mismo que se han elaborado muchas definiciones e índices de sequía teniendo eso presente, La universidad de Nebraska - Lincoln (EEUU) organizó del 8 al 11 de diciembre de 2009 el curso internacional sobre índices y sistema de alerta temprana de sequías. Es así que la OMM et al. (2012), desarrollan una guía recomendando que para describir las características de las sequías meteorológicas, y mencionan que "todos los Servicios Meteorológicos e Hidrológicos Nacionales (SMHN) utilizaran el Índice Normalizado de Precipitación o comúnmente conocido como SPI por sus siglas en inglés, que es uno de los más utilizados en más de 70 países".

En el SPI original la distribución de frecuencias de la precipitación es descrita usando la función de probabilidad Gamma 2 parámetros (Farahmand & AghaKouchak, 2015).

En la mayoría de los casos, la mejor distribución que se ajusta a los datos de precipitación observacional es la distribución Gamma (Rossi, Vega & Brunella., 2007). La función de densidad de probabilidad para la distribución Gamma está dada por la expresión:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad \text{para } x > 0 \quad (2.33)$$

donde x , es la variable o precipitación mensual, α y β son los parámetros de forma y escala respectivamente de la distribución Gamma, los mismos que se pueden estimar usando el método de máxima verosimilitud o técnicas de inteligencia artificial como los AG (Hassanzadeh et al., 2012; Xin She, Amir Hossein, Siamak & Amir Hossein, 2012). $\Gamma(\alpha)$, es la función Gamma definida por la integral:

$$\Gamma(\alpha) = \lim_{n \rightarrow \infty} \prod_{v=0}^{n-1} \frac{n! n^{y-1}}{y+v} \equiv \int_0^\infty y^{\alpha-1} e^{-y} dy \quad (2.34)$$

En general, la función Gamma es evaluado numéricamente tomando el parámetro α . Con el fin de modelar los datos observados con una función de densidad Gamma dis-



tribuido, es necesario estimar los dos parámetros. Edwars y McKee (1997) usan la aproximación siguiente

$$\alpha = \frac{1 + \sqrt{1 + 4A/3}}{4A} \quad (2.35)$$

$$\beta = \frac{\bar{x}}{\alpha} \quad (2.36)$$

Siendo A una variable auxiliar

$$A = \ln(\bar{x}) - \frac{\sum_{i=1}^{nn} \ln(x_i)}{nn} \quad (2.37)$$

donde nn , es el número de datos no nulos de la serie x_i , cuya media aritmética es \bar{x}

Luego de estimar los parámetros α y β , la función de densidad de probabilidad $g(x)$ 2.33, es integrado con respecto a x y se obtiene la función de distribución acumulada $G(x)$

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-\frac{x}{\beta}} dx \quad (2.38)$$

Haciendo $t = x/\beta$, la ecuación anterior se convierte en una función de distribución acumulada incompleta

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (2.39)$$

Como se aprecia, $G(x)$ no es válida para precipitaciones igual a cero ($x_i = 0$), es así que se toma en cuenta la probabilidad de tener valores nulos q y la probabilidad de no tenerlos $p = 1 - q$, de tal manera se obtenga la función de distribución acumulada completa $H(x)$ y se evalúa la probabilidad de no excedencia de un valor x . Su expresión es:

$$H(x) = q + (1 - q)G(x) \quad (2.40)$$

Edwars y McKee (1997), presentan una solución aproximada, utilizando la técnica de conversión desarrollado en Abramowitz y Stegun (1965) que convierte la probabilidad acumulada en una variable estándar Z . Entonces, el índice SPI se define como:

$$Z = SPI = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{para } 0 < H(x) \leq 0.5 \quad (2.41)$$

$$Z = SPI = + \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad \text{para } 0.5 < H(x) \leq 1 \quad (2.42)$$



donde

$$t = \sqrt{\ln \left[\frac{1}{(H(x))^2} \right]} \quad \text{para } 0 < H(x) \leq 0.5 \quad (2.43)$$

y

$$t = \sqrt{\ln \left[\frac{1}{(1 - H(x))^2} \right]} \quad \text{para } 0.5 < H(x) \leq 1 \quad (2.44)$$

y por último las constantes son:

$$\begin{aligned} c_0 &= 2.515517 & c_1 &= 0.802853 & c_2 &= 0.010328 \\ d_0 &= 1.432788 & d_1 &= 0.189269 & d_2 &= 0.001308 \end{aligned}$$

Thomas B. McKee et al. (1993), implementaron la clasificación de las sequías según los resultados de los SPI, de acuerdo a la tabla 2.6.

Tabla 2.6: Clasificación de sequías según los SPI

SPIs	Clasificación
2.0 y más	Extremadamente húmedo
1.5 a 1.99	Muy húmedo
1.0 a 1.49	Moderadamente húmedo
-0.99 a 0.99	Normal o aproximadamente normal
-1.0 y -1.49	Moderadamente seco
1.5 a -1.99	Severamente seco
-2.0 a más	Extremadamente seco

Fuente: Adaptado de OMM, Svoboda, Hayes y Wood (2012, p. 5)

2.4.3. Índice de Precipitación Evapotranspiración Estandarizada (SPEI)

El índice SPEI, propuestos por Vicente-Serrano, Beguería y López-Moreno (2010), es un índice de sequía multiescalar basado en la diferencia entre la precipitación mensual y la Evapotranspiración Potencial (PET).

Para un cierto mes, el balance hídrico conocido como la diferencia D entre la precipitación P y la PET se pueden expresar según la ecuación 2.45 (Vicente-Serrano et al., 2010)

$$D = P - PET \quad (2.45)$$

donde D es un indicador cuantitativo de la medida del agua en déficit o exceso y se puede agregar en diferentes escalas de tiempo (Hao & Singh, 2015).



Para determinar la PET, Liu, Wang, Shao, Jia y Li (2016, p. 285), menciona que "varios métodos han sido propuestos, usando parámetros meteorológicos. Entre ellos el método de Penman-Moteith (PM), siendo adoptado ampliamente como el procedimiento estándar para el cálculo del PET por muchas asociaciones, tales como la Comisión Internacional de Riegos y Drenajes (ICID), Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) y la Sociedad Americana de Ingenieros Civiles (ASCE)".

Por simplicidad de cálculo, se puede usar el método de Thornthwaite, el cual sólo requiere de las temperaturas medias mensuales como dato (Hao & Singh, 2015). En otro caso, el método de Hargreaves (temperaturas máximas y mínimas mensuales como dato).

La distribución de probabilidades para el valor D , fue ajustada a la función de distribución *log-logística* de tres parámetros.

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x - \gamma}{\alpha} \right)^{\beta-1} \left[1 + \left(\frac{x - \gamma}{\alpha} \right)^{\beta} \right]^{-2} \quad (2.46)$$

$$F(x) = \left[1 + \left(\frac{\alpha}{x + \gamma} \right)^{\beta} \right]^{-1} \quad (2.47)$$

donde α , β y γ son los parámetros de escala, forma y origen respectivamente, $f(x)$, función de densidad de probabilidad y $F(x)$ función de probabilidad acumulada. Vicente-Serrano et al. (2010), usó el procedimiento de L-Momentos para obtener los tres parámetros, cuyas relaciones son:

L-Momentos de la serie D :

$$\begin{aligned} l_1 &= w_0 \\ l_2 &= w_0 - 2w_1 \\ l_3 &= w_0 - 6w_1 + 6w_2 \\ l_4 &= w_0 - 12w_1 + 30w_2 - 20w_3 \end{aligned} \quad (2.48)$$

Los parámetros:

$$\beta = \frac{2w_1 - w_0}{6w_1 - w_0 - 6w_2} \quad (2.49)$$

$$\alpha = \frac{(w_0 - 2w_1)\beta}{\Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 - \frac{1}{\beta}\right)} \quad (2.50)$$

$$\gamma = w_0 - \alpha\Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 - \frac{1}{\beta}\right) \quad (2.51)$$

donde w_s son los momentos ponderados de probabilidad de orden s , $s = 0, 1, 2$



$$w_s = \frac{1}{N} \sum_{i=1}^N (1 - F_i)^s D_i \quad (2.52)$$

y F_i es el estimador de frecuencias calculados como

$$F_i = \frac{i - 0.35}{N} \quad (2.53)$$

donde i es el rango de observaciones en orden creciente y N es la cantidad o longitud de datos.

Considerando la aproximación clásica de Abramowitz y Stegun (1965) (ecuaciones 2.41, 2.42, 2.43 y 2.44), se calcula el SPEI. El criterio de evaluación es similar al índice SPI.

El índice SPEI, presenta ventajas respecto a otros índices, ya que éste es multiescalar e incluye la variable temperatura y su aplicación es adecuada para el monitoreo de la sequía relacionado con el cambio climático (Hao & Singh, 2015).

2.4.4. Índices Climáticos Alternos

Son considerados como indicadores alternos para una descripción general de las sequías. Lobo (2006), describe algunos procedimientos (guía) para evaluar la climatología de América Latina y Caribe.

Índice de Aridez (IAR) Indica el déficit o exceso de la disponibilidad hídrica en una temporada de análisis, teniendo la relación siguiente.

$$IAR = \frac{P_m}{PET} \quad (2.54)$$

donde P_m es la precipitación media del periodo de análisis y PET la evapotranspiración potencial calculada por el método de Thornthwaite, Hargreaves o Penman Monteith.

Tabla 2.7: Clasificación del índice IAR

Zonas	IAR
Hiperáridas	<0.05
Áridas	0.05 a <0.20
Semiárida	0.20 a <0.50
Subhúmedas secas	0.50 a <0.65
Subhúmedas húmedas	0.65 a 1
Húmedas	>1

Fuente: Adaptado de Lobo (2006, p. 3)



Índice de Fournier Modificado (IFM) El IFM, está en relación a la carga de sedimentos en los ríos. Lobo (2006), cita a Arnoudus (1980), y menciona "...modificó el índice IF, considerando las cantidades de lluvia de todos los meses del año: índice de Fournier Modificado (IFM) y no sólo la del mes más húmedo".

$$IFM = \frac{\sum_{i=1}^{12} p_i^2}{P} \quad (2.55)$$

donde p_i es la precipitación mensual y P es la precipitación anual. Este valor, se debe calcular para cada año (anual) y luego evaluar el promedio de los anuales en cada estación.

Tabla 2.8: Clasificación del índice IFM

Descripción	IFM
Muy bajo	<60
Bajo	60-90
Moderado	90-120
Alto	120-160
Muy alto	>160

Fuente: Adaptado de Lobo (2006, p. 6)

Índice de Estacionalidad (IE) Define los aspectos temporales de la precipitación y evalúa, compara la concentración de las precipitaciones entre estaciones (Lobo, 2006)

$$IE = 100 \frac{\sum_{i=1}^{12} p_i^2}{P^2} \quad (2.56)$$

Tabla 2.9: Clasificación del índice IE

Descripción	IE(%)
Uniforme	8.3-10
Moderadamente estacional	10-15
Estacional	16-20
Fuertemente estacional	21-50
Irregular	50-100

Fuente: Adaptado de Lobo (2006, p. 7)

2.5. Redes Neuronales Artificiales

Las RNA son modelos computacionales diseñados para simular el funcionamiento del cerebro humano y la forma cómo éste procesa información a través del **aprendizaje**



que se da en la etapa de entrenamientos entre los patrones de entrada (recepción de datos) y salida (objetivo del aprendizaje) o propiamente en la vecindad de los patrones. El sistema de cálculo más complejo es el cerebro humano, superando a cualquier ordenador existentes, por su propia naturaleza mejorada en tiempo, en eficacia a cualquier ordenador.

El ordenador siempre estará por debajo del cerebro humano, con algunas ventajas como por ejemplo en la realización de tareas repetitivas, resolución de ecuaciones matemáticas complejas, entre otros. La computación neuronal, es simular un cerebro humano dentro de un ordenador según la experiencia de los ejemplos aprendidos.

Muchos problemas de la actualidad no pueden ser tratados mediante los algoritmos y métodos básicos, por lo que se recurren a metodologías nuevas como son Lógica Difusa, las RNA, las Redes de Petri, los diagnósticos a fallos, Algoritmos Genéticos, un conjunto de metodologías que ayudan a cooperar en la solución de problemas (Basogain Olabe, 2000).

El desarrollo de las RNA inició en los años 50 con McCulloch y Pitts³, inspirado por el deseo de entender el cerebro humano y emular su funcionamiento con la naturaleza y el desarrollo como técnica ha experimentado un renacimiento sólo en las décadas pasadas debido al esfuerzo de Hopfield en la redes neuronales iterativos autoasociados. Dentro de las últimas décadas, las RNA ha experimentado un enorme resurgimiento debido al desarrollo de algoritmos sofisticados y la evolución emergente de las herramientas computacionales.

La habilidad para identificar una relación desde patrones dados, hacen posible que las RNA resuelvan problemas complejos tales como reconocimiento de patrones, modelamientos no lineales, clasificación asociación y control.

Desde los principios de los años 90, las RNA en la Hidrología fueron usados de manera exitosa tales como modelamiento de precipitación escorrentía, pronóstico de caudales, modelamiento de agua subterránea, calidad del agua, manejo y gestión de recursos de agua, pronóstico de precipitaciones, series de tiempo hidrológico y operaciones de embalse (ASCE, 2000a).

2.5.1. Características de las Redes Neuronales Artificiales

Las RNA tienen su origen en las redes biológicas de los seres vivos, cuyas características son:

Aprender : Adquirir conocimiento a partir de la experiencia (datos).

Generalizar : Una vez realizada el entrenamiento, se le pueden presentar a la red datos distintos a los usados durante el aprendizaje. La respuesta obtenida dependerá del parecido de los datos con los ejemplos de entrenamiento.

Abstraer : Las redes neuronales artificiales son capaces de extraer o abstraer las características esenciales de las entradas aprendidas, de esta manera pueden procesar correctamente datos incompletos o distorsionados (presencia de ruidos).

³McCulloch neurofisiólogo y Walter Pitts matemático.

Distribución de Memoria : El conocimiento acumulado por la red se halla distribuido en numerosas conexiones, esto tiene como consecuencia la tolerancia a fallos: una red neuronal es capaz de seguir funcionando adecuadamente a pesar de sufrir lesiones con destrucción de neuronas o sus conexiones, ya que la información se halla distribuida por toda la red, sin embargo en un programa tradicional un pequeño fallo en cualquier punto puede invalidarlo todo y dar un resultado absurdo o no dar ningún resultado.

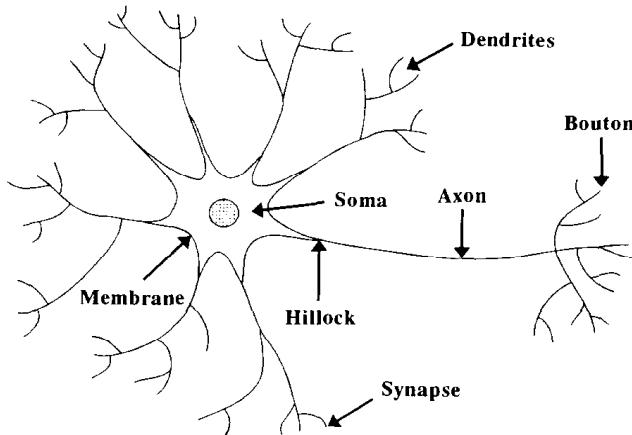
La aplicabilidad de las RNA tiene que darse en aquellos campos donde el cerebro humano tiene una respuesta adecuada y su implementación puede ser en la forma de hardware o software (E. Y. Li, 1994).

Las RNA son considerados como una herramienta de caja negra que tiene una gran capacidad en el modelamiento y predicción (ASCE, 2000a; Guoqiang, Patuwo. & Y, 1998; J. Abrahart, M. See & P. Solomatine, 2008; Lek & Guégan, 1999).

2.5.2. Estructura Básica de una Red Neuronal

El elemento básico de un modelo biológico es la célula nerviosa o neurona.

Figura 2.6: Estructura física de una neurona biológica



Fuente: Adaptado desde E. Y. Li (1994)

Modelo Biológico La neurona es el elemento básico del modelo biológico. En 1906, el Dr. Santiago Ramón y Cajal⁴ obtuvo Premio Nobel de Medicina en 1906 por descubrir los mecanismos que gobiernan la morfología y los conectivos de las células nerviosas, una nueva y revolucionaria teoría empezó a ser llamada la *doctrina de la Neurona* (Basogain Olabe, 2000).

Santiago y Cajal, demostró que las neuronas son células que tienen varias entradas y una sola salida, realizando experimentos en ratones, donde con tinta de pluma señalaban

⁴Médico Español, especializado en histología y anatómo-patología microscópica

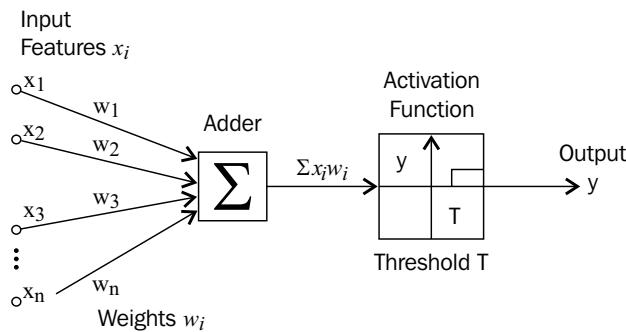


varias entradas y conseguían un canal de salida, lo que intuyó que las Neuronas tiene varias entradas y una sola salida.

Modelo Artificial Usando la analogía biológica, se puede construir un modelo de una neurona computacional y este modelo es conocido como el modelo de *McCulloch-Pitts* de una neurona (Forte, 2015).

McCulloch y Pitts (1943), propusieron este modelo de una red neuronal como una máquina computacional, publicado en el Boletín de Matemática Biofísica.

Figura 2.7: Estructura física de una neurona artificial



Fuente: Adaptado desde Forte (2015)

Tabla 2.10: Comparación entre Redes Neuronales Biológicas y Artificiales

RNB	RNA
Dendritas	Ingresos: $X_0X_1\dots X_n$
Sinapsis	Pesos: W_{ij}
Núcleo	Sumatoria y Función de transferencia
Axón	Y_i

Fuente: Elaboración propia

2.5.3. Función de activación

Llamado también función de transferencia. Determina la relación entre el ingreso y salidas de un nodo y la red. En general, la función de activación introduce un grado de no linealidad que es evaluado por muchas aplicaciones de RNA. Las funciones de transferencia, permiten mayor flexibilidad al sistema y por ende mayor capacidad en el funcionamiento de las RNA.

Las funciones más comunes son:

Función escalón (hardlim) : Esta función crea neuronas que clasifican las entradas en dos categorías diferentes, característica que le permite ser empleada en la red

tipo Perceptrón.

$$a = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases} \quad (2.57)$$

Función lineal (*purelin*) : La salida de una función de transferencia lineal es igual a su entrada.

$$a = n \quad (2.58)$$

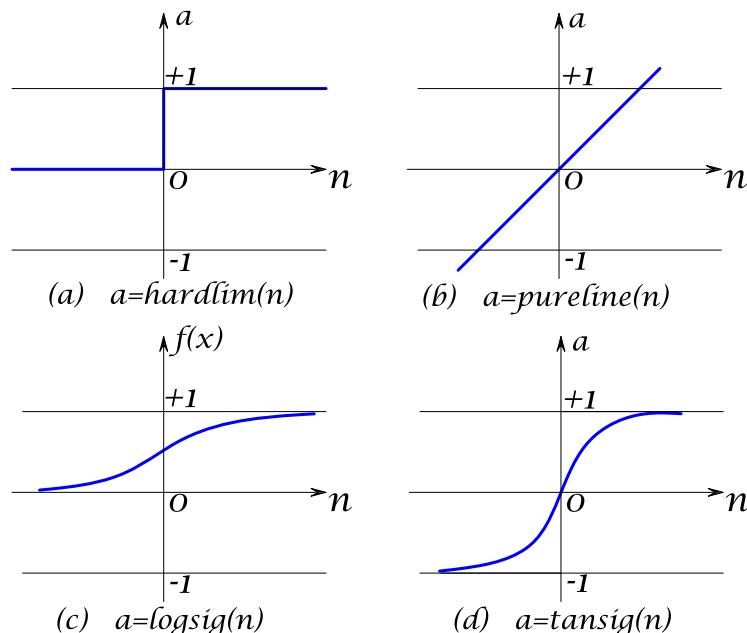
Función logarítmica sigmoidea (*logsig*) : Esta función toma los valores de entrada, los cuales pueden oscilar entre más y menos infinito, y restringe la salida a valores entre cero y uno, de acuerdo a la expresión:

$$a = \frac{1}{1 + e^{-n}} \quad (2.59)$$

Función tangente hiperbólica (*tansig*) : Es una de las funciones más usadas en las RNA por su flexibilidad, el amplio rango de resultados que ofrece (rango de salida entre -1 y 1).

$$a = \frac{2}{1 + e^{-2n}} - 1 \quad (2.60)$$

Figura 2.8: Tipos de función de transferencia (a) hard limit, (b) lineal, (c) log sigmoidal y (d) tangente sigmoidal



Fuente: Elaboración propia

2.5.4. Entrenamiento de las Redes Neuronales Artificiales

Se denomina entrenamiento de las RNA al proceso de configuración de una Red Neuronal para que las entradas produzcan las salidas deseadas a través del fortalecimiento

de las conexiones. Consiste en el ajuste de pesos w_{ij} según la regla de aprendizaje. Gradualmente los valores w_{ij} van convergiendo hacia valores de la salida deseada o consistente.

Usando una técnica de búsqueda, el proceso de aprendizaje se esfuerza para obtener un conjunto de pesos que corresponden a un mínimo global (Basheer & Hajmeer, 2000). Los tipos son supervisados y no supervisados.

Método Supervisado Existe un maestro que presenta unos patrones de entrada a la red, compara la salida de la red con la salida deseada y ajusta los pesos de la red de forma que disminuye el error. Biológicamente es poco probable que exista este tipo de aprendizaje.

Existe un maestro que se encarga de comparar la salida deseada con una salida real que da las redes neuronales, siendo imprescindible en los sistemas supervisados disponer de esta salida deseada (target).

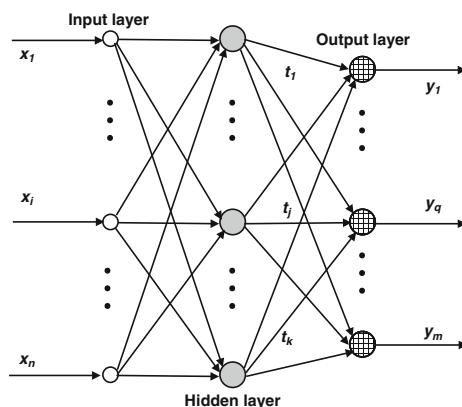
Método No Supervisado No existe un maestro, no hace falta el vector correspondiente deseado, sino que una RNA va a generar una salida y_i y el mismo en función de un criterio. Modificará los pesos w_{ij} para la nueva salida y_i .

2.5.5. El perceptrón Multicapa

El perceptrón es la forma más simple de una red neuronal usado para la clasificación de patrones linealmente separables (Haykin, 1999). Las redes multicapa o perceptrón multicapa surgen a raíz de las limitaciones del perceptrón simple.

Las Redes Neuronales Artificiales Multicapa (RNAM) es entrenada en forma supervisada mediante el algoritmo de retropropagación, compuesta por dos o más capas de neuronas con función de activación sigmoidal logarítmica, sigmoidal tangente hiperbólica o lineal.

Figura 2.9: Estructura RNA Multicapa



Fuente: Adaptado desde Krasnopolksy (2013)



Está formado por el vector de entrada uno o más capas ocultas y la capa de salida; las capas ocultas pueden tener cualquier número de neuronas, mientras que el tamaño del vector de entrada y el número de neuronas de la capa de salida se seleccionan de acuerdo a las entradas y salidas de la redes respectivamente.

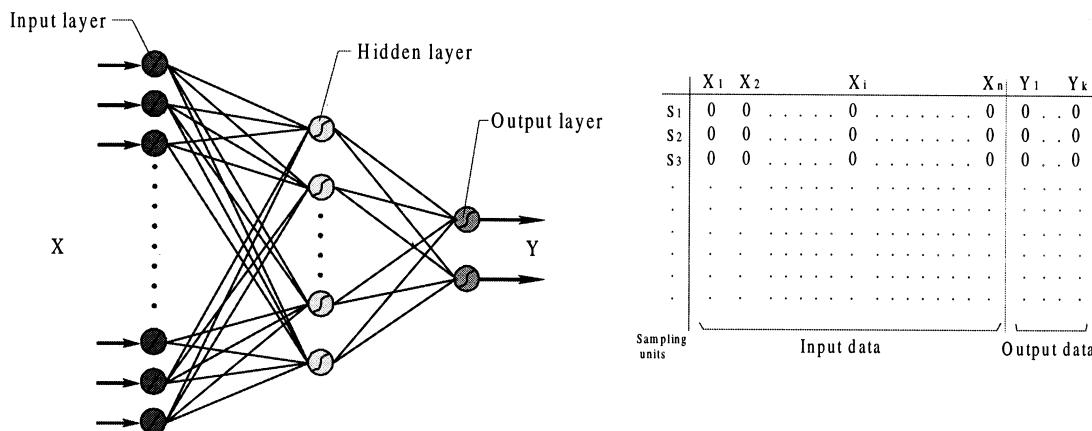
Todas las neuronas de cada capa deben tener la misma función de activación, pudiendo ser esta función diferente a la de otras capas.

2.5.6. Arquitectura de la Red

Viene a ser la forma geométrica de la conexión de las neuronas que forman la red, siendo definidas por el número de capas, neuronas por capas y la forma de conexión.

Una RNA es típicamente compuesto por capas de nudos. Según Guoqiang et al. (1998, p. 42) en un perceptrón multicapa se necesita determinar las siguientes variables: el número de nudos de ingreso, número de capas ocultas con sus respectivos nudos ocultos y finalmente el número de nudos de salida.

Figura 2.10: Arquitectura RNA Multicapa



Fuente: Adaptado desde Lek y Guégan (1999)

El número de nudos de ingreso, corresponde al número de variables en el vector de ingreso.

Las capas y nudos ocultos, juegan un rol muy importante para la resolución de problemas no lineales. Si éstos no existieran, sería un perceptrón simple, capaz de resolver sólo problemas de carácter lineal, equivalentes a los modelos estadísticos de predicción lineal. Muchos autores sólo usan una capa oculta para propósitos de predicción, sin embargo éste modelo puede requerir muchos nudos ocultos. Dos capas ocultas pueden proporcionar buen desempeño para algunos tipos de problemas (Guoqiang et al., 1998). Por lo tanto, la determinación de éstos parámetros es a base de ensayos *prueba-error*. Guoqiang et al. (1998) cita a varios autores, los mismos que determinan la relación para el número de nudos ocultos, conforme se muestra en la 2.11.

Finalmente, el número de nudos de salida, está directamente relacionada con el problema en estudio.



Tabla 2.11: Técnicas para establecer número de nudos ocultos

Autor(es)	Relación
Lippmann (1987), Hecht-Nielsen (1990)	$2n + 1$
Wong (1991)	$2n$
Tang y Fishwick (1993)	n
Kang (1991)	$n/2$

Donde n es el número de nudos de ingreso

Fuente: Elaboración propia

Retropropagación La red Retropropagación (Backpropagation) es un tipo de red con aprendizaje supervisado. Su funcionamiento consiste en el aprendizaje de un conjunto predefinido de pares entradas-salidas, empleando un ciclo de propagación-adaptación de dos fases: una vez que se ha aplicado un patrón a la entrada de la red como estímulo, éste se propaga desde la primera capa a través de las capas intermedias u ocultas de la red, hasta generar una salida.

La señal de salida se compara con la salida deseada y se obtiene un error para cada una de las salidas. El error se propaga hacia atrás, partiendo de la capa de salida, pasa por las capas ocultas hasta llegar a la capa de entrada. Las neuronas de las capas ocultas y de entrada, sólo reciben una fracción del total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite capa por capa, hasta que todas las neuronas de la red hayan recibido un error que describa su contribución relativa al error total. Basándose en el error percibido, se actualizan los pesos de conexión de cada neurona, para hacer que la red converja hacia un estado que permita clasificar correctamente todos los patrones de entrenamiento.

La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan así mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del conjunto de entradas (Pelaéz Chávez, 2012).

2.5.7. Modelo Self Organizing Map

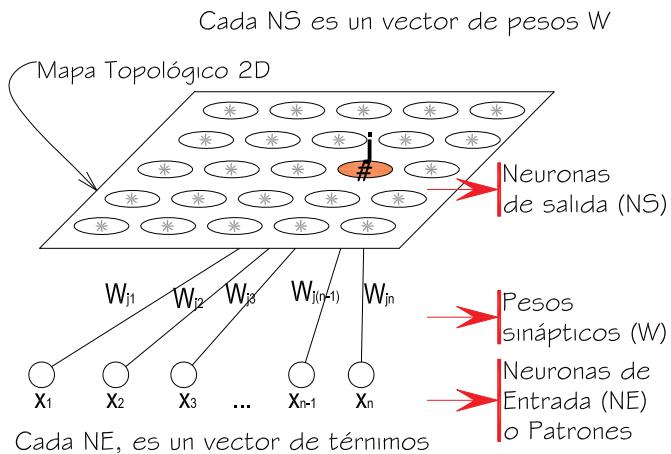
El modelo SOM, también conocido como mapas auto organizados de Kohonen, se caracterizan por no tener un maestro o supervisor para el aprendizaje. Son redes que consta de 2 capas, el primero denominado neuronas de entrada (NE) y el segundo neuronas de salida (NS) y no existen capas ocultas, por lo que un conjunto de conexiones siempre son hacia adelante y cada neurona tiene asociada a un peso W entre la entrada y salida.

Son redes competitivas especialmente indicadas para trabajar en problemas de clasificaciones de alta dimensionalidad (Gutiérrez, Cano, Cofiño & Sordo, 2004; Teuvo Kohonen, 2013). Según T. Kohonen (1990), su aplicación es multidisciplinario tales como: reconocimiento de patrones estadísticos, control de procesos industriales, siste-

mas automáticos de digitales, telecomunicación adaptativa, compresión de imágenes, clasificación radar de hielo marino, problemas de optimización, entre otros.

Según Basogain Olabe (2000, p. 35), "la red de Kohonen presenta ventajas sobre técnicas clásicas de reconocimiento de patrones, porque además de utilizar la arquitectura paralela, provee una representación gráfica de las relaciones entre los patrones".

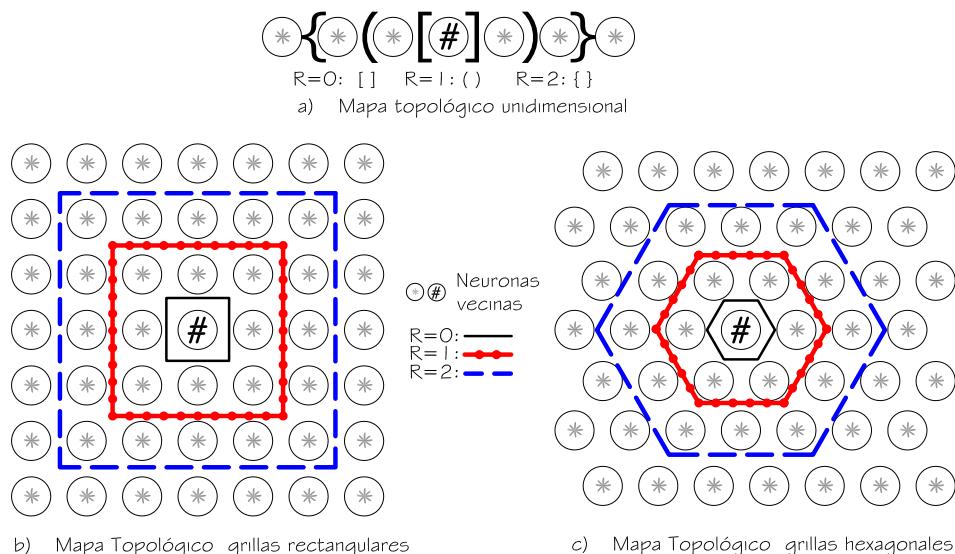
Figura 2.11: Estructura del modelo Self Organizing Map



Fuente: Elaboración propia

La figura 2.11, muestra que cada neurona de entrada está conectada a cada neurona del mapa bidimensional mediante los pesos sinápticos W . Cada neurona procesa la similitud entre el vector de entrada y su propio vector de pesos. En cada ciclo o paso, lo que se pretende es que la neurona vencedora modifique sus pesos hasta el más parecido del patrón de entrada.

Figura 2.12: Estructura de vecindades



Fuente: Elaboración propia



La característica más común es la de clasificar o extraer similitudes desde un patrón de entrada, normalmente con muchos valores para proyectar a un mapa topológico 2D de dimensión pequeña (generalmente) y que guarden relación con el grupo de datos. Por lo tanto, el objetivo principal de la red SOM, es transformar el grupo de datos (volúmenes de información multidimensionales) de entrada en un mapa discreto 1D ó 2D (clustering).

En la figura 2.12, se puede observar los mapas topológicos unidimensionales (1D) y bidimensionales (2D) con geometrías rectangulares y hexagonales que usualmente están conectadas unas con otras, también se observa vecindades de distintos tamaños ($R=0$, 1 y 2).

Durante el entrenamiento, el número de neuronas y la arquitectura son constantes. Hagan, Demuth, Beale y De Jesús (1996), Haykin (1999), T. Kohonen (2001), coinciden en el siguiente algoritmo:

Paso 1 Inicializar los pesos W_{ij} . Vienen a ser valores aleatorios a los vectores de referencia.

Paso 2 Mientras la condición de parada sea falsa, hacer los pasos 3-9

Paso 3 Para cada vector de ingreso \mathbf{x} hacer los pasos 4-5

Paso 4 Para cada paso j , calcular:

$$D(j) = \sum_i (W_{ij} - x_i)^2 \quad (2.61)$$

Paso 5 Encontrar el índice \mathbf{J} , tal que $D(\mathbf{J})$ sea mínimo.

Paso 6 Para todos los pasos j , actualizar los pesos sinápticos usando la relación

$$W_{ij}(t+1) = W_{ij(t)} + \eta[x_i - W_{ij}(t)] \quad (2.62)$$

Paso 7 Actualizar la taza de aprendizaje

Paso 8 Reducir el radio de vecindad topológico en el tiempo específico

Paso 9 Prueba de condición de parada

Donde η viene a ser la tasa de aprendizaje y t es la coordenada del tiempo discreto.

2.5.8. Modelo No Lineal Auto regresivas con entradas Externas (NARX)

El modelo NARX, es una red neuronal recurrente para modelamientos no lineales y específicamente en series de tiempo. Tiene un lazo de realimentación desde sus neuronas de salida hasta la capa de entrada, conformado por variables causales considerados como valores del pasado. Hagan et al. (1996, 22:10), menciona que el modelo NARX es popular y adecuado para problemas de control y modelamiento dinámico.



Hagan et al. (1996), Haykin (1999), definen la ecuación para el modelo NARX

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-n_y), u(t-1), u(t-2), \dots, u(t-n_u)) \quad (2.63)$$

donde y son las neuronas de salida. El siguiente valor de la señal de salida dependiente $y(t)$ es regresivo a los valores previos de señales y valores anteriores de una señal de entrada independiente, conocida también como variables exógenas $u(t)$.

Esta red se diferencia de las estáticas (backpropagation) por contener elementos de retro-alimentación con retardos. Por ello, la denominación de red dinámica, donde la salida, no depende únicamente de las entradas actuales a la red, sino también de las entradas, salidas y estados de la red actual y anterior. NARX está basado en el Modelo ARX Lineal que se utiliza comúnmente en problemas de series de tiempo.

El proceso de entrenamiento es como sigue:

Paso 1 Recibe la serie de tiempo de entrenamiento con sus respectivas entradas externas.

Paso 2 Finalizada la etapa del entrenamiento, se cierra el ciclo, lo cual conecta las salidas con la Línea de retardo (TDL) respectivo.

Paso 3 Para iniciar la predicción, requiere los últimos valores del pasado con las respectivas entradas externas.

Paso 4 Cada iteración recibe una nueva entrada externa y retorna la predicción del un paso a la vez.

2.6. Algoritmos Genéticos

Son algoritmos de búsqueda basados en la mecánica de la selección natural y genética natural propuesta por Holland Jhon (1989) (E. Goldberg, 1989). Son algoritmos matemáticos altamente paralelo que transforma un conjunto de objetos matemáticos con respecto al tiempo, empleando operaciones modeladas de acuerdo al principio darwiniano de reproducción y supervivencia del más apto. En otros términos, son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización (búsqueda del **mejor**) y basados en el proceso genético de los organismos vivos.

Los AG han sido desarrollados por John Holland, colegas y sus estudiantes en la Universidad de Michigan. Holland, impartía un curso titulado *Teoría de sistemas adaptativos*. Dentro de este curso, se crearon las ideas que más tarde se convertirían en los AG. Por tanto, cuando Holland se enfrentó a los algoritmos genéticos, los objetivos de su investigación fueron dos: el primero, imitar los procesos adaptativos de los sistemas naturales, y el segundo, diseñar sistemas artificiales que retengan los mecanismos importantes de los sistemas naturales (E. Goldberg, 1989).

Los AG, utilizan una analogía directa del fenómeno de evolución en la naturaleza. Trabajan con una población de individuos, cada uno representando una posible solución a un problema dado. Los más aptos se reproducen con otros individuos de la población, produciendo descendientes con características de ambos padres. Los padres son



seleccionados aleatoriamente, pero sus probabilidades de selección están en relación a la evaluación de los cromosomas.

Aunque los AG son fáciles de describir y de programar, su comportamiento puede ser complejo y existen muchas preguntas abiertas acerca de cómo funcionan y para qué tipo de problemas sus aplicaciones son más adecuados (Mitchell, 1996, p. 21).

Mujahid y Mathew (2014), presenta problemas reales para evaluar la aplicación y ejecución de los AG, encontrando los valores de los parámetros que proporciona la solución más óptima (máxima generación, la población óptima, tasa de cruce óptimo, tasa de mutación óptima).

2.6.1. Bases Biológicas

En la naturaleza, los individuos de una población compiten constantemente con otros por recursos como alimentación, agua, refugio, entre otros. Los individuos que tienen más éxito en la lucha por los recursos tienen mayor posibilidad de sobrevivir y generalmente una descendencia mayor y viceversa para los individuos peor adaptados.

La combinación de características buenas de diferentes ancestros pueden originar en ocasiones, que la descendencia esté incluso mejor adaptada al medio que los padres. De esta manera, las especies evolucionan adaptándose más y más al entorno a medida que transcurren las generaciones (Beasley, Bull & Martin, 1993).

Tabla 2.12: Similitud con Sistemas biológicos

Sistema Biológico	Algoritmo Genético
Los hijos tienen un código genético independiente de sus padres.	Los hijos reciben un código independiente de sus padres a través del operador mutador.
Los padres son gradualmente reemplazados por sus hijos.	Los padres mueren inmediatamente una vez que nacen los hijos.
La población es cada vez más apta y se adapta al ambiente con el paso del tiempo.	La población de soluciones es cada vez mejor para resolver el problema.

Fuente: Elaboración propia

2.6.2. Terminologías y Operadores de AG

2.6.2.1. Codificación

La codificación en AG, es un proceso de representación de los genes de los individuos. Estos procesos pueden ser conformados usando bits, números, letras, vectores, listas o cualquier otro objeto. La codificación principalmente depende de la naturaleza de la solución del problema. Por ejemplo, uno puede codificar directamente con números enteros ó reales.



Figura 2.13: Codificación Comunes

Codificación Binaria						
Cromosoma 1	0	1	1	0	1	0
Cromosoma 2	1	1	0	1	0	0
Codificación Octal						
Cromosoma 1	3	5	2	0	7	1
Cromosoma 2	1	4	0	3	5	2
Codificación Hexadecimal						
Cromosoma 1	9AE7					
Cromosoma 2	3CBF					

Fuente: Elaboración propia

Codificación Binaria: La codificación más común es la cadena binaria, el mismo que es representado como la figura 2.13. Cada cromosoma codifica una cadena binaria (bit). Cada bit en la cadena puede representar alguna característica de la solución. Por consiguiente, cada cadena de bit es una solución, pero no específicamente la mejor solución. Otra posibilidad es que la cadena entera o cromosoma puede representar un número. La codificación binaria está compuesta generalmente de ceros y unos y la longitud de la cadena depende en la importancia.

Codificación Octal: Esta codificación usa cadenas que contienen números reales entre $[0 - 7]$

Codificación Hexadecimal: Esta codificación usa cadenas que contienen números hexadecimales (sistema de numeración de base 16) $[0 - 9, A - F]$.

2.6.2.2. Operador de Selección

Son los encargados de escoger qué individuos va a disponer de oportunidades de reproducción y cuáles no.

Se otorga un mayor número de oportunidades de reproducción a los individuos más aptos, por lo tanto, la selección de un individuo estará relacionada con su valor de ajuste. No se debe sin embargo eliminar por completo las opciones de reproducción de los individuos menos aptos, pues en pocas generaciones la población se volvería homogénea.

Una opción bastante común consiste en seleccionar el primero de los individuos participantes en el cruce mediante métodos existentes y el segundo de manera aleatoria.

2.6.2.3. Operador de Cruce

Selecciona dos padres al azar y corta sus ristras de cromosomas en una posición escogida al azar en ambos padres. Crea los hijos intercambiando las sub ristras finales



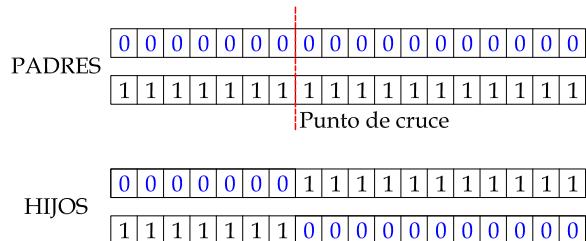
llamado comúnmente colas o produciéndose dos nuevos cromosomas completos. Ambos descendientes heredan genes de cada uno de los padres.

Habitualmente el operador de cruce no se aplica a todos los pares de individuos que han sido seleccionados para emparejarse, sino que se aplica de manera *aleatoria*, normalmente con una probabilidad comprendida entre $[0.50 - 1.0]$ ó $[0.6 - 0.9]$. En el caso en que el operador de cruce no se aplique, la descendencia se obtiene simplemente duplicando los padres.

De Jong (1975) en su tesis doctoral, precidido por John Holland, investigó el comportamiento del operador de cruce basado en múltiples puntos, concluyendo que el cruce basado en dos puntos, representaba una mejora, mientras que añadir más puntos de cruce no beneficiaba el comportamiento del algoritmo.

La ventaja de tener más de un punto de cruce radica en que el espacio de búsqueda puede ser explorado más fácilmente, siendo la principal desventaja el hecho de aumentar la probabilidad de ruptura de buenos esquemas.

Figura 2.14: Operación Cruce



Fuente: Elaboración propia

2.6.2.4. Operador de Mutación

Evita que solo se considere un subconjunto de las posibles soluciones. Una mutación, en términos evolutivos, es un suceso bastante poco común, siendo estos contundentes en la evolución y constituyen a la diversidad genética de la especie.

En los AG, primero se establece la frecuencia de mutación, se examina cada gen de cada hijo. Si su número generado aleatoriamente está por debajo de esa probabilidad, se cambia el gen (es decir de 0 a 1 ó de 1 a 0), si no, se deja como está.

La probabilidad de que un gen de la población sea mutado es el parámetro conocido como probabilidad de mutación P_m . Por cada uno de los gens se genera un número de aleatorio k y si la probabilidad de mutación del gen i $k_i < P_m$, entonces el gen i será mutado, de lo contrario se mantiene intacto.

La probabilidad de mutación es muy baja, generalmente menor a 1 %. Esto se debe sobre todo a que los individuos suelen tener un ajuste menor después de mutados. Sin embargo, se realizan mutaciones para garantizar que ningún punto del espacio de búsqueda tenga una probabilidad nula de ser examinado, es más, su importancia resalta en asegurar la convergencia de los AGs.



Si se trabaja con codificaciones binarias, consiste simplemente en negar un bit. También es posible realizar la mutación intercambiando los valores de los dos alelos del cromosoma. Con otro tipo de codificación no binaria se incrementan o decrementan a un gen por un valor aleatorio próximo a 1.

La búsqueda del valor óptimo para la probabilidad de mutación, es una cuestión que ha sido motivo de varios trabajos. Así De Jong (1975) recomienda la utilización de una probabilidad de la mutación:

$$P_m = \frac{1}{l} \quad (2.64)$$

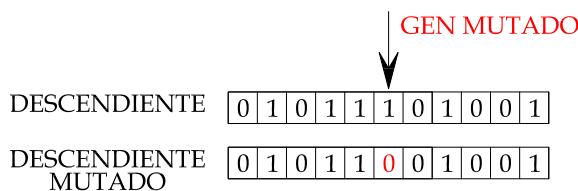
siendo l la longitud del string.

Schaffer et al. (1989) utilizan resultados experimentales para estimar la tasa óptima proporcional a

$$P_m = \frac{1}{\lambda^{0.9318} l^{0.4535}} \quad (2.65)$$

donde, λ denota el número de individuos en la población.

Figura 2.15: Operación Mutación



Fuente: Elaboración propia

2.6.3. Diagrama de Flujo Algoritmos Genéticos

Mitchell (1995, p. 5), propone un diagrama similar a la figura 2.16 para la solución de problemas bien definidos:

Paso 01 Generación aleatoria de una población de N cromosomas y L bits.

Paso 02 Evaluar la aptitud de cada cromosoma en la población.

Paso 03 Previa evaluacion, realizar la operación de la selección con un par de cromosomas de los padres de la población actual.

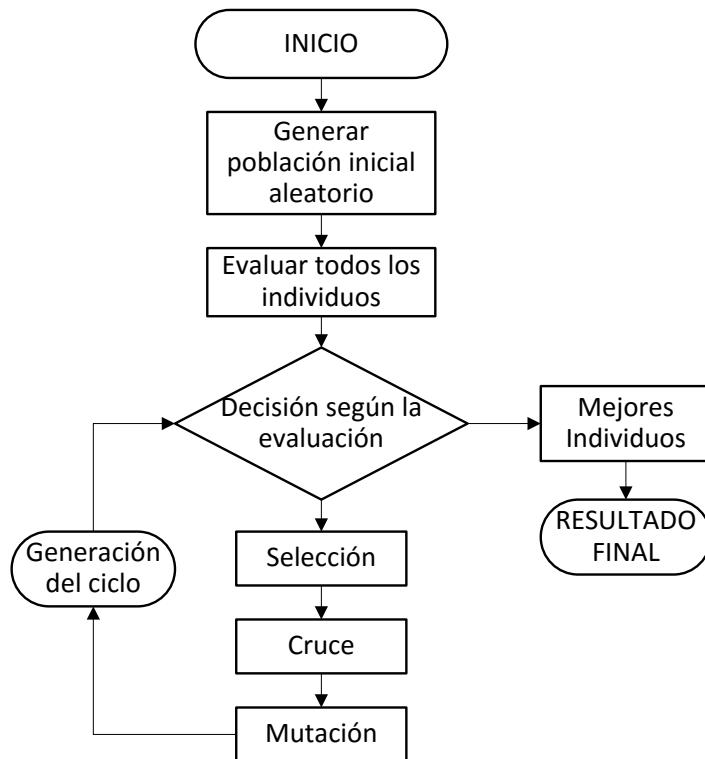
Paso 04 Mediante la probabilidad del cruce, se realiza esta operación en un punto elegido de manera aleatoria para formar dos descendencias (hijos). Y si no existe cruce, se forman dos descendencias que son copias fieles de sus respectivos padres.

Paso 05 Mutar las dos descendencias en cada gen con una probabilidad designada y mostrar los cromosomas resultantes de la nueva población.

Paso 06 Reemplazar la población actual con la nueva población.

Paso 07 Repetir el ciclo.

Figura 2.16: Algoritmo genético diagrama de flujo



Fuente: Elaboración propia

Cada iteración de este proceso se llama una *generación*. Por lo tanto, los AG son secuencias iterativas que recomienda Mitchell (1995) desde 50 hasta 500 o más generaciones, denominándose un *run*.

2.6.4. Algoritmos Genéticos Computacional

Actualmente existen numerosas aplicaciones escritas en distintos lenguajes de programación como Fortran, C, C++, Java, R, Python, Perl, Matlab entre otros, sin embargo muchos de ellos son bajo licencia pública o denominados programas comerciales.

Python es un lenguaje de programación de propósito general muy fácil de aprender, con una sintaxis característica que hace que los programas escritos en él sean muy legibles, ampliamente utilizado por empresas como Google o la NASA y, además, libre. Fue creado en 1991 por Guido van Rossum, un programador holandés (Cano, 2012). En bioinformática, Python o Perl es ampliamente utilizado (Manoj, Mohammad, Naveen & Deepti, 2010).

Actualmente, existen librerías en Python para los AG, tal como indica el cuadro 2.13

Al igual que Python, existen múltiples paquetes o librerías escritas en el lenguaje R, por ejemplo la librería **AG**, el mismo que es ampliamente utilizado para la búsqueda del mejor individuo con salidas tabulares y gráficas de los resultados. La librería **GA**, está alojada en el repositorio <https://cran.r-project.org>.



Tabla 2.13: Librerías Python para la solución de los AG

Librería	Característica	Dirección electrónica
Pyevolve	Está bien documentado pero no incluye multiobjetivo.	http://pyevolve.sourceforge.net/
Pygene	No incluye multiobjetivo	https://github.com/blaa/PyGene
deap	Incluye multiobjetivo	https://github.com/deap/deap
inspyred	Incluye multiobjetivo	https://pypi.python.org/pypi/inspyred
Pybrain	Incluye multiobjetivo	http://pybrain.org/

Fuente: Elaboración propia

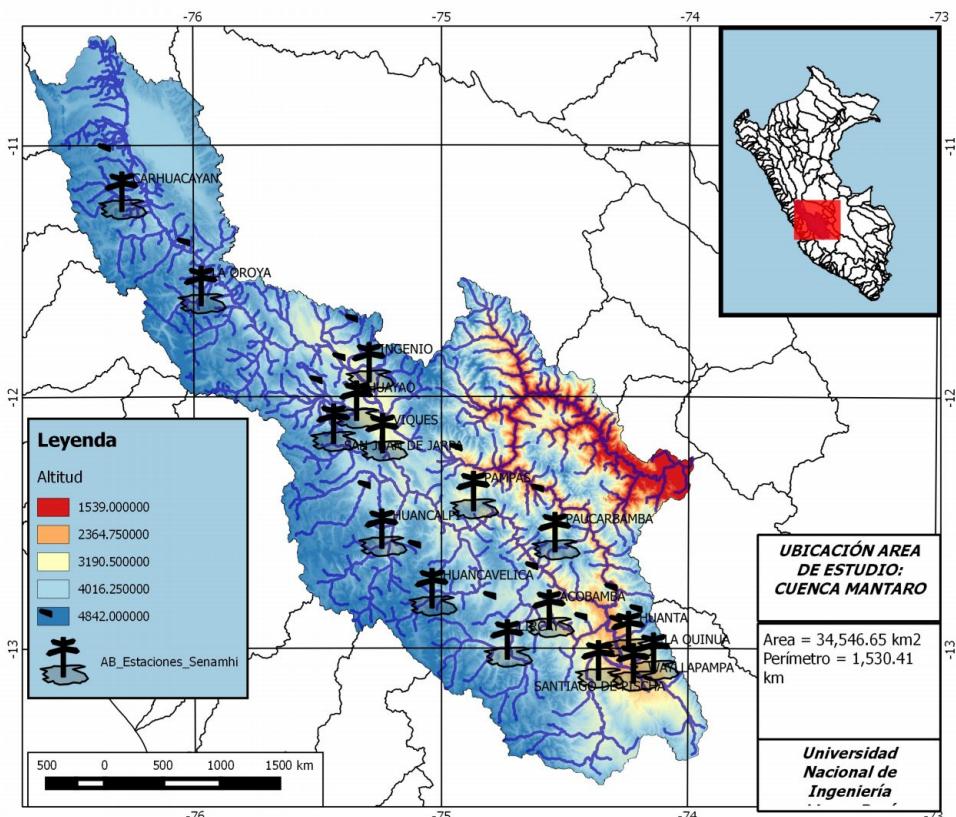
La velocidad de cálculo es relativa según el tamaño de la población o tipo de lenguaje de programación (interpretado o compilado). Respecto al Tamaño de la población es crucial en los AG, el mismo que determina el tamaño de la memoria, velocidad de convergencia y la velocidad de búsqueda (El-Mihoub, Hopgood, Nolle & Battersby, 2006).

Capítulo III: Desarrollo de la Investigación

3.1. Identificación de la zona de estudio

En principio, según la definición de la muestra o sub grupo de la población (Hernández Sampieri, Hernández Collado & Baptista Lucio, 2014, p. 171), corresponde a un muestreo **No Probabilístico** o dirigido según el alcance, tipo y nivel exploratorio de la investigación.

Figura 3.17: Delimitación de la cuenca del Mantaro y ubicación de estaciones pluviométricas



Fuente: Elaboración propia software QGIS



La zona de estudio, corresponde a la cuenca del Mantaro distribuidos en 1145 celadas consideradas como estaciones sintéticas. Se ubica en la región hidrográfica del río Amazonas y viene a ser una de las cuencas pertenecientes a la vertiente del Atlántico. Geográficamente se ubica en el centro del Perú entre los paralelos $10^{\circ}34'30''$ y $13^{\circ}35'30''$ de latitud sur, y entre los meridianos $73^{\circ}55'00''$ y $76^{\circ}40'30''$ de longitud oeste. Representa el 2.8 % del territorio peruano.

La muestra de estudio, es identificada a través del índice de aridez expuesta en el capítulo 4.1, que corresponde a la región más árida, es decir, la zona con mayor recurrencia a las sequías y pertenece a la región sur de la cuenca del Mantaro.

Las estaciones meteorológicas activas (estado: Funcionando) dentro de la zona de estudio son 16 (tabla 3.14), monitoreadas por SENAMHI.

Tabla 3.14: Ubicación de las estaciones meteorológicas

Zona	Estación	Latitud	Longitud	Altitud (msnm)
Norte	Carhuayacan	$11^{\circ} 12' 1''$	$76^{\circ} 17' 1''$	4150
Norte	La Oroya	$11^{\circ} 34' 34''$	$75^{\circ} 57' 56''$	4007
Centro	Ingenio	$11^{\circ} 52' 51''$	$75^{\circ} 17' 16''$	3422
Centro	San Juan de Jarpa	$12^{\circ} 07' 30''$	$75^{\circ} 25' 55''$	3650
Centro	Viques	$12^{\circ} 09' 45''$	$75^{\circ} 14' 03''$	3186
Centro	Huayao	$12^{\circ} 02' 01''$	$75^{\circ} 20' 20''$	3360
Centro	Huancalpi	$12^{\circ} 32' 32''$	$75^{\circ} 14' 14''$	4450
Sur	Huancavelica	$12^{\circ} 46' 45''$	$75^{\circ} 02' 02''$	3770
Sur	Lircay	$12^{\circ} 58' 57''$	$74^{\circ} 43' 43''$	3513
Sur	Acobamba	$12^{\circ} 51' 51''$	$74^{\circ} 33' 32''$	3356
Sur	Pampas	$12^{\circ} 23' 35''$	$74^{\circ} 51' 58''$	3240
Sur	Paucarbamba	$12^{\circ} 33' 20.4''$	$74^{\circ} 32' 9.4''$	3370
Sur	Huanta	$12^{\circ} 56' 55''$	$74^{\circ} 14' 13''$	2682
Sur	La Quinua	$13^{\circ} 02' 02''$	$74^{\circ} 08' 07''$	3316
Sur	Wayllapampa	$13^{\circ} 04' 36''$	$74^{\circ} 13' 01''$	2470
Sur	Santiago de Pischa	$13^{\circ} 03' 55''$	$74^{\circ} 21' 31''$	2990

Fuente: Elaboración Propia

3.2. Metodología

Para lograr el objetivo de la presente investigación, se plantea procedimientos de secuencia como se muestran en las figuras 3.18 y 3.19. La figura 3.18 contiene 6 etapas, que consiste en:

Primera etapa : correspondiente a la adquisición de datos por percepción remota de las Variables Climáticos Regionales (VCR), tales como las precipitaciones y temperaturas máximas y mínimas mensuales dentro de la unidad de estudio.

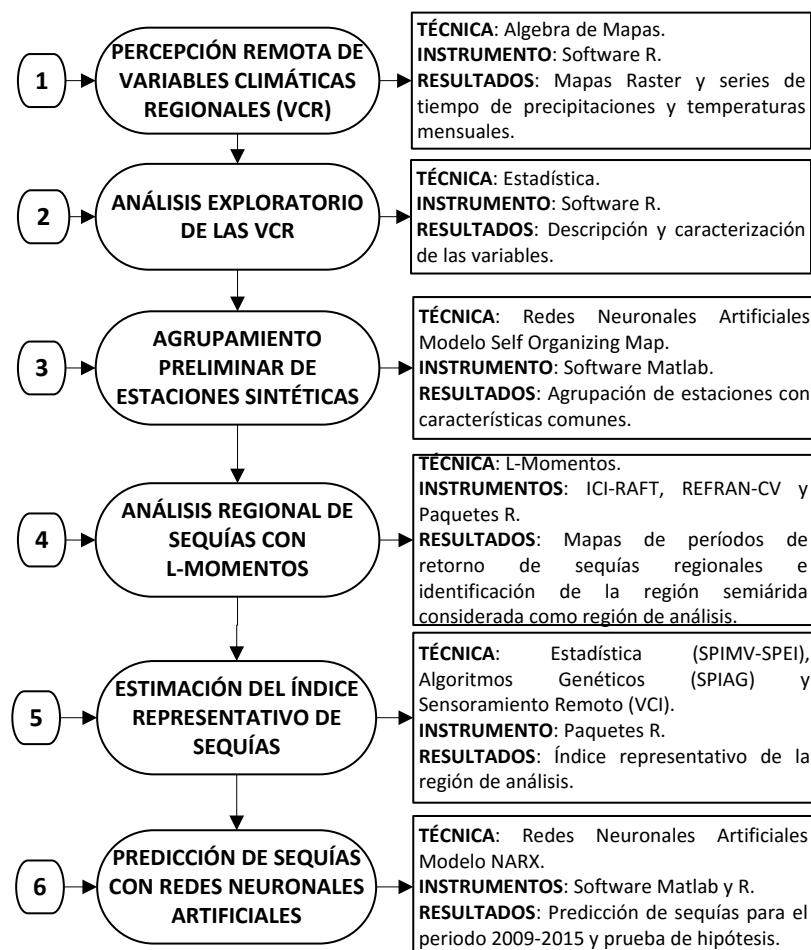
Para la adquisición de los registros de precipitaciones mensuales, se recurrió a los datos PISCO, el mismo que interviene con la metodología de interpolación geo-

estadística entre la variable precipitación mensual y la covariable datos grillados de CHIRPS, obteniendo modelos de semivariograma residual entre las precipitaciones CHIRPS y precipitaciones puntuales registrados por SENAMHI. Siendo éste una de las alternativas de corrección a los datos adquiridos por percepción remota, para la obtención de los mapas de precipitación mensual en formatos raster, de resolución 0.05 grados latitud y longitud (equivalente a 5km^2). Las series de tiempo de las variables precipitaciones y temperaturas registran desde 1981 hasta 2015 (35 años de registro).

Segunda etapa : corresponde al análisis exploratorio de las VCR, incluidos la altitud e índices externos, que permiten conocer las diferentes variables para fines de agrupamiento y análisis regional de sequías en la cuenca del Mantaro. Mediante el análisis estadístico descriptivo, se evalúa los datos anuales en cada celda o grilla a través de histogramas, diagrama de cajas y mapas raster.

De manera alternativa, en esta etapa se aprecian áreas o regiones húmedas y secas o áridas a través de los índices externos que están en función de las precipitaciones y la evapotranspiración potencial estimada por el método de Hargreaves.

Figura 3.18: Esquema metodológico: cálculo de los SPIs óptimos a partir de precipitaciones por sensoramiento remoto



Fuente: Elaboración propia



Tercera etapa : consiste en agrupar de manera preliminar las estaciones sintéticas, consideradas 1145 inicialmente. La finalidad de ésta etapa es reducir el conjunto de datos **discordantes**, que no permiten realizar de manera eficiente la etapa de regionalización. El modelo RNA-SOM es una técnica robusta para llevar a cabo este proceso, que consiste en agrupar las estaciones sintéticas con características similares a partir de las variables regionales como la Precipitación Media Anual (PMA), temperaturas, elevación, latitud y longitud de cada celda considerada como estación sintética.

Los resultados finales vienen a ser los prototipos, conocidos también como los pesos de las conexiones entre las neuronas de entrada y salida. De esta manera se han obtenido resultados con mejoras respecto a los estadísticos Discordancia y Homogeneidad del análisis regional que corresponden el siguiente paso.

Cuarta etapa : A través de la técnica de L-Momentos, propuestas por Hosking y Wallis (1997) y mas tarde metodologías propuestas por Nuñez y Verbist (2010), Wallis et al. (2007), se realizan el análisis regional de sequías que consisten en la identificación de regiones homogéneas, identificación de la función de distribución de probabilidades, determinación de los cuantiles y finalmente el mapeo de las precipitaciones para distintos períodos de retorno.

El objetivo de esta etapa es, describir a nivel estadístico la distribución de precipitaciones en la cuenca del Mantaro, a través de los mapas raster que se elaboran para diferentes períodos de retorno. Asimismo, se corroboró la existencia de una región semiárida, es decir, región con presencia de precipitaciones inferiores a lo normal. A éste último, se le asigna como **región de análisis**.

Quinta etapa : En la *región de análisis*, se estiman los índices SPIMV, SPIAG y SPEI para las escalas temporales 1, 3, 6, 9, y 12 meses a fin de identificar el mejor índice denominado **índice representativo** con su correspondiente escala temporal frente a la respuesta de la vegetación de la cuenca del Mantaro.

El índice de sequía patrón para esta evaluación es el VCI de Kogan y Sullivan (1993), cuya resolución espacial es 1km x 1km y registro temporal desde marzo 2000 hasta diciembre 2015. Los estadísticos de evaluación son la correlación de Pearson y el diagrama de Taylor. Los SPIMV propuestas por (Thomas B. McKee et al., 1993), son soluciones numéricas de la distribución de probabilidades Gamma, mientras los SPIAG, son estimados a través de los AG, basado en la teoría de la evolución de Darwin, para calcular los parámetros α y β , de la distribución Gamma (Hassanzadeh et al., 2012).

Con población igual 1000 individuos, 100 ciclos de generación, cruzamiento y mutación igual a 0.8 % y 0.1 % respectivamente, se consiguen los parámetros óptimos para estimar los SPIs. Finalmente el índice SPEI de Vicente-Serrano et al. (2010), incluye el parámetro de la Evapotranspiración Potencial para estimar cuantitativamente los índices de sequía.

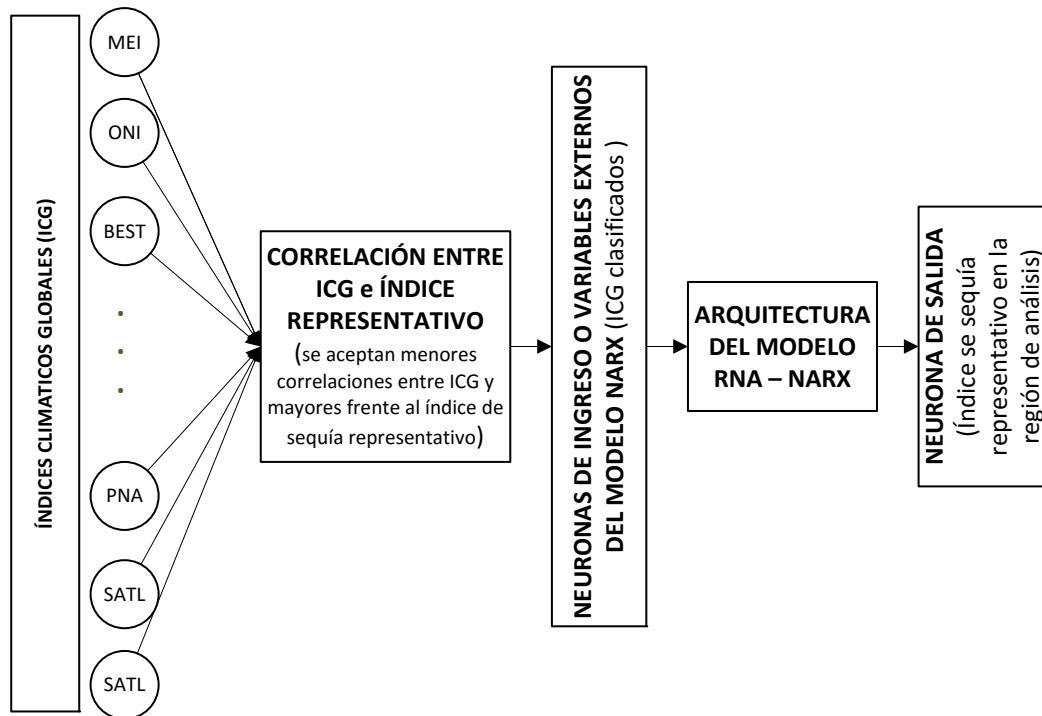
Sexta etapa : En esta última etapa, la serie temporal de la *región de análisis* y el **índice representativo** se divide en dos grupos. El primero es el 80 % del registro total (1981-2008) y se usa para la construcción del modelo predictivo mediante las RNA-NARX y el segundo grupo, 20 % restante (2009-2015) para la validación

y prueba de hipótesis del modelo y la investigación planteada.

La función de entrenamiento es mediante *Levenberg-Marquardt*, derivado en la técnica de optimización del método de Newton y función de transferencia *Tangente sigmoidal*. Para el modelo NARX, las neuronas de entrada son las variables externas consideradas a los (ICG), previa evaluación de la correlación cruzada entre sí y el *índice representativo* considerado como neurona de salida. Los coeficientes menos significativos entre los ICG y mas significativos entre el *índice representativo* se evalúa para evitar la redundancia de neuronas de entrada y relación significativa respectivamente.

La arquitectura de la red se establece a través del procedimiento *ensayo-error* progresivo, evaluando el desempeño en cada caso. La figura 3.19, muestra el procedimiento durante la construcción del modelo de predicción y finaliza con la obtención de los pesos de conexión entre las capas y neuronas respectivas, para un test estadístico (Pearson) significativo, tanto en el entrenamiento, validación y prueba del modelo.

Figura 3.19: Esquema metodológico: predicción de sequía



Fuente: Elaboración propia

Para validar el modelo, se realiza la prueba de la hipótesis, que consiste en la comparación de dos grupos: los índices simulados con RNA-NARX y los índices obtenidos que corresponde al 20 % del registro mencionado anteriormente.



3.2.1. Optimización de los parámetros función Gamma para estimar los SPIs utilizando AG

Como fuentes principales se toman a Xin She et al. (2012) del capítulo 6 "Ant Colony Optimization for Estimating Parameters of Flood Frequency Distributions", adaptado para esta investigación de Colonias de Hormigas a Algoritmos Genéticos, y el paquete Standardized Precipitation Index using Genetic Algorithm Method (SPIGA) publicado por el autor de la tesis en el repositorio CRAN-R, dirección electrónica <https://cran.r-project.org/web/packages/SPIGA/index.html>.

Para calcular los parámetros de la distribución estadística Gamma según la ecuación 2.38, usando los AGs, se establecen las funciones objetivos:

$$\text{Minimizar} \left(\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{P})^2} \right) \quad (3.66)$$

$$\text{Minimizar} \left(\sum_{i=1}^N (O_i - P_i)^2 \right) \quad (3.67)$$

Los valores observados, vendrían a ser las probabilidades empíricas que se evalúan mediante las ecuaciones de la tabla 3.15. Si la elección es Weibull, la ecuación sería:

$$O_i = F(X) = \frac{i}{N+1} \quad (3.68)$$

Donde N es el total de muestras.

Para determinar el rendimiento y la calidad sobre los resultados se toma al Coeficiente de Determinación (CD) y el coeficiente de Nash-Sutcliffe (1970) también conocido como Coeficiente de Eficiencia (CE).

$$CD = \frac{\sum_{i=1}^N [(O_i - \bar{O})(P_i - \bar{P})]}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \quad ; -1 \leq CD \leq 1 \quad (3.69)$$

$$CE = 1 - \frac{\sum_{i=1}^N [(O_i - P_i)^2]}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad ; -\infty < CE \leq 1 \quad (3.70)$$

donde. O es el valor Observado y P es valor calculado o predecido.

Tabla 3.15: Distribuciones de probabilidades empíricas

Método	año	Probabilidad	Período de retorno
California	1923	$\frac{m}{N}$	$\frac{N}{m}$
Hazen	1930	$\frac{m-0.5}{N}$	$\frac{N}{m-0.5}$
Weibull	1939	$\frac{m}{N+1}$	$\frac{N+1}{m}$

(continúa en la página siguiente)



Tabla 3.15: Continuación de la tabla

Método	año	Probabilidad	Período de retorno
Gregodavyev	1955	$\frac{m-0.3}{N+0.4}$	$\frac{N+0.4}{m-0.3}$
Blom	1958	$\frac{m-0.375}{N+0.25}$	$\frac{N+0.25}{m-0.375}$
Turkey	1962	$\frac{3m-1}{3N+1}$	$\frac{3N+1}{3m-1}$
Griengorten	1963	$\frac{m-0.44}{N+0.12}$	$\frac{N+0.12}{m-0.44}$

Fuente: Adaptado desde Agbede O.A. y Abiona O (2012)

3.2.2. Información Climatológica Global

En la actualidad existen diversos índices entre atmosféricos y climáticos que se ubican en distintas regiones del planeta cumpliendo el rol de monitoreo. Estos datos, se pueden descargar de manera gratuita, desde la dirección electrónica <http://www.esrl.noaa.gov/psd/data/climateindices/list/> y están expresados en series de tiempo.

La información climática global, es implementado como argumentos de ingreso al modelo RNA-NARX, previa evaluación de cada variable, expuesta en la sección de metodología.

Índice Multivariado ENSO (MEI) Es considerado índice mixto que involucra variables atmosféricas como aceáticas en su composición. Es el promedio de seis variables sobre el Pacífico tropical: presión atmosférica a nivel del mar, componente zonal (este-oeste) y meridional (norte - sur) del viento en superficie, temperatura de la superficie del mar, temperatura del aire en superficie y cantidad total de nubosidad. Los valores positivos, indican la fase cálida y por ende presencia de El Niño. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/mei.data>.

Índice de El Niño Oceánico (ONI) Es una serie de tiempo de las anomalías de la temperatura superficial marina calculadas a partir de mediciones in situ efectuadas en la región El Niño 3.4 ($5^{\circ}\text{N}-5^{\circ}\text{S}$, $120-170^{\circ}\text{O}$). Los episodios fríos y cálidos son definidos cuando la anomalía se encuentra por arriba o por debajo de un valor umbral ($+0.5^{\circ}\text{C}$) durante, al menos, cinco periodos consecutivos. Se definen como episodios El Niño (+) o de La Niña (-) (Ver Tabla 3.16). La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/oni.data>.

Tabla 3.16: Clasificación de los eventos según el valor del ONI - NOAA en la región del Pacífico tropical 3.4

Categoría	Índice Oceánico El Niño (ONI)
Cálido fuerte	$ONI \geq +1.40$
Cálido moderado	$+0.90 \leq ONI < +1.40$
Cálido débil	$+0.50 \leq ONI < +0.90$
Neutro	$-0.50 \leq ONI < +0.50$
Frío débil	$-0.90 \leq ONI < -0.50$
(continúa en la página siguiente)	



Tabla 3.16: Continuación de la tabla

Categoría	Índice Oceánico El Niño (ONI)
Frío moderado	$-1.50 \leq ONI < -0.90$
Frío fuerte	$ONI \leq -1.50$

Fuente: NOAA

Oscilación Decadal del Pacífico (PDO) El Índice de la Oscilación Decadal del Pacífico (PDO por sus siglas en inglés), resume las variaciones mensuales de la temperatura superficial marina en el Océano Pacífico al norte de los 20°N (Fig. 14). La periodicidad de este evento es de 15 a 20 años. Los cambios de temperatura en esa región tienen un gran impacto en el clima de Norteamérica porque alteran el patrón de vientos, particularmente la ubicación de la corriente de chorro y de las tormentas. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/pdo.data>.

Índice de la Oscilación del Sur (SOI) Este valor es un indicador de la diferencia de presión entre Tahití y Darwin, que se presenta debido al efecto de subir y bajar que se presenta en el Pacífico Ecuatorial. Con frecuencia la magnitud de esta oscilación está relacionada con el desarrollo e intensidad de los eventos de El Niño y La Niña. Los valores negativos de las diferencias de presión sin normalizar menores a -8 indican los episodios El Niño, en tanto que los valores positivos mayores a 8 son típicos de La Niña; sin embargo, cuando los datos son normalizados los valores límite son -0.5 y 0.5 (ver Tabla 3.17). La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/soi.data>.

Tabla 3.17: Clasificación de los eventos según el valor del SOI - Western Regional Climate Center NOAA

Clave	Intensidad del evento	SOI	Clasificación del evento
I EN	Fuertemente Negativo	≤ -1.00	El Niño Intenso
M EN	Moderadamente Negativo	≤ -0.50	El Niño Moderado
N	Normal	-0.50 y $+0.50$	Normal
M LN	Moderadamente Positivo	$\geq +0.50$	La Niña Moderada
I LN	Fuertemente Positivo	$\geq +1.00$	La Niña Intensa

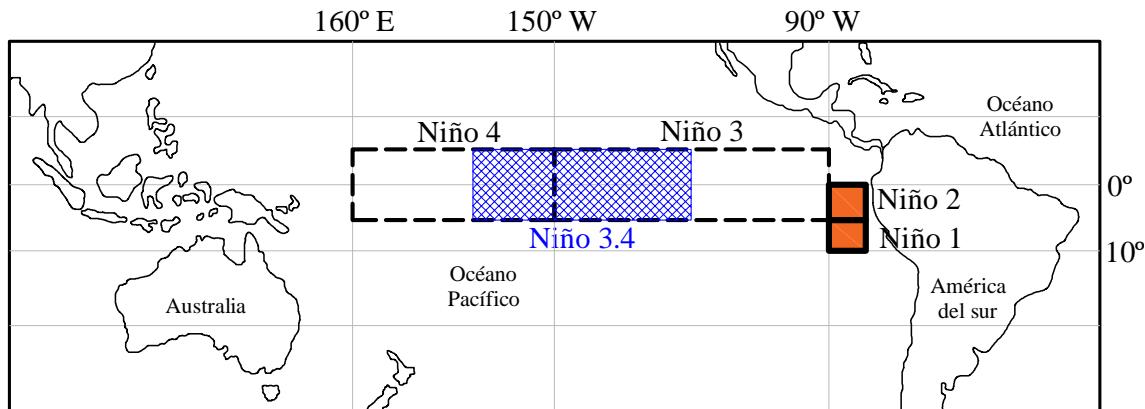
Fuente: Obtenido desde <http://www.cdc.noaa.gov/>

Oscilación Multianual del Atlántico (AMO) Es un modo de variabilidad que ocurre en el Océano Atlántico Norte y que tiene su principal expresión en la Temperatura de la Superficie del Mar (TSM). Se asocia a las sequías en Norteamérica y Norte de Brasil. Fue identificado el año 2001 por Stanley B. Goldenberg y publicado en The Recent Increase in Atlantic Hurricane Activity: Causes and Implications. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/amon.us.long.data>.

Temperatura de la superficie del mar región Niño 1+2 Corresponde a la zona del Pacífico Tropical Extremo Oriental 0-10°S y 90°O-80°O. Esta región incluye las costas de Perú y las Islas Galápagos, representa un indicador de los cambios inducidos por

El Niño en los patrones de variabilidad de la costa del Pacífico de América del Sur. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/nina1.data>.

Figura 3.20: Localización de Regiones EL NIÑO



Fuente: Elaboración propia

Temperatura de la superficie del mar región Niño 3 Corresponde a la variación de la TSM en la zona del Pacífico Tropical Oriental entre: 5°N-5°S y 150°O-90°O. Alguna vez, se consideró como el principal enfoque para el monitoreo y predicción del EL NIÑO, sin embargo años después se determinaron que el NIÑO 3.4 y ONI proporcionan mejor enfoque para la definición de los eventos de EL NIÑO y LA NIÑA. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/nina3.data>.

Temperatura de la superficie del mar región Niño 3.4 Corresponde a la zona del Pacífico Tropical Este entre: 5°N-5°S y 170°O-120°O. Si la TSM, excede en ± 0.40 °C por un periodo de 6 o más meses, se definen eventos de EL NIÑO o la NIÑA. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/nina34.data>.

Temperatura de la superficie del mar región Niño 4 Corresponde a la variación de la TSM en la zona del Pacífico Tropical Central entre: 5°N-5°S y 160°E-150°O. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/nina4.data>.

Índice C y E (Takahashi, Montecinos, Gobanova & Dewite, 2011), muestran dos índices oceánicos como el C y E, considerando ser **no correlacionales e independientes** y demuestran una buena representación para el ENSO. El índice C corresponde al Pacífico Central y el índice E al Pacífico Este. El índice C está asociada a las precipitaciones en los Andes.

$$C \approx 1.7\text{NIÑO}_4 - 0.1\text{NIÑO}_{1+2} \quad (3.71)$$

$$E \approx \text{NIÑO}_{1+2} - 0.5\text{NIÑO}_4 \quad (3.72)$$

La dirección de descarga es: <http://www.met.igp.gob.pe/datos/EC.txt>.

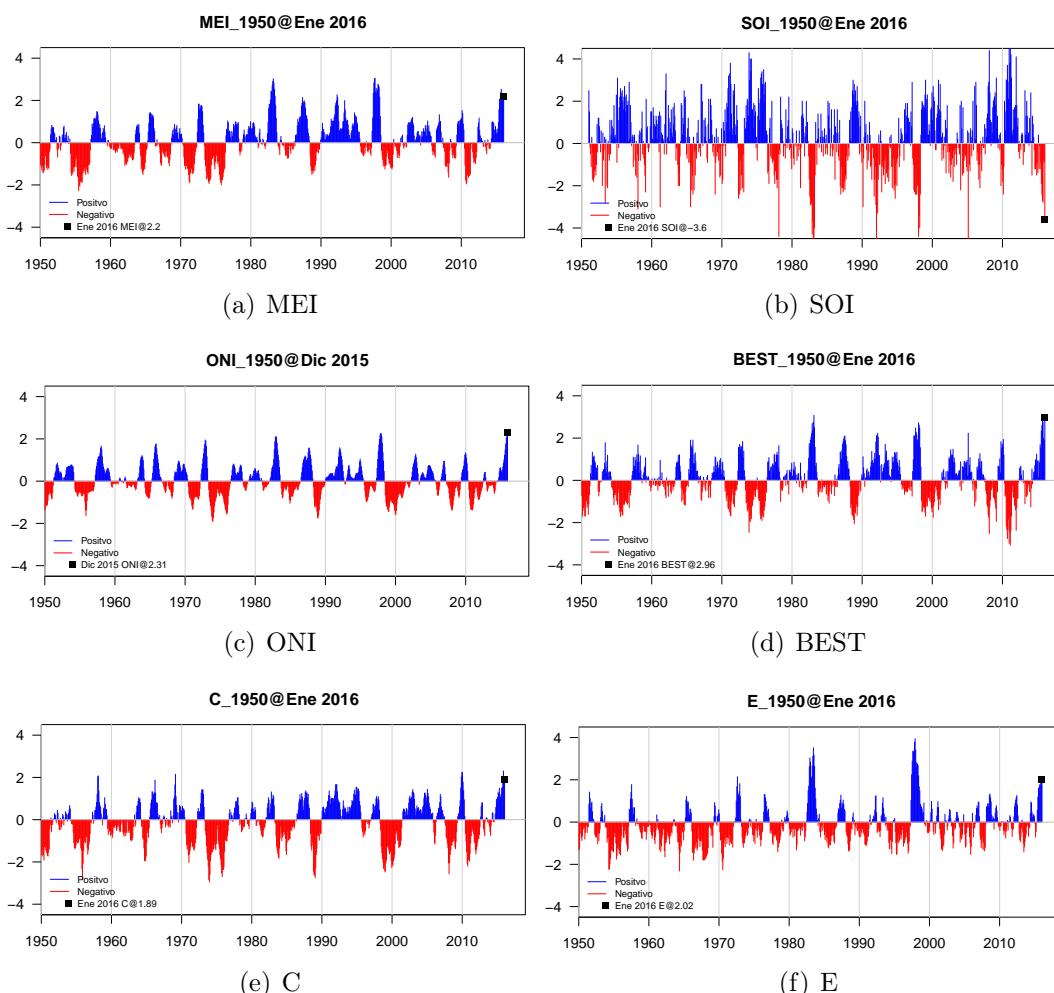
Índice BEST Índice bivariado ENSO. Calculado a partir de la combinación de una normalización de las series de tiempo SOI y TSM de Niño3.4. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/censo.data>.

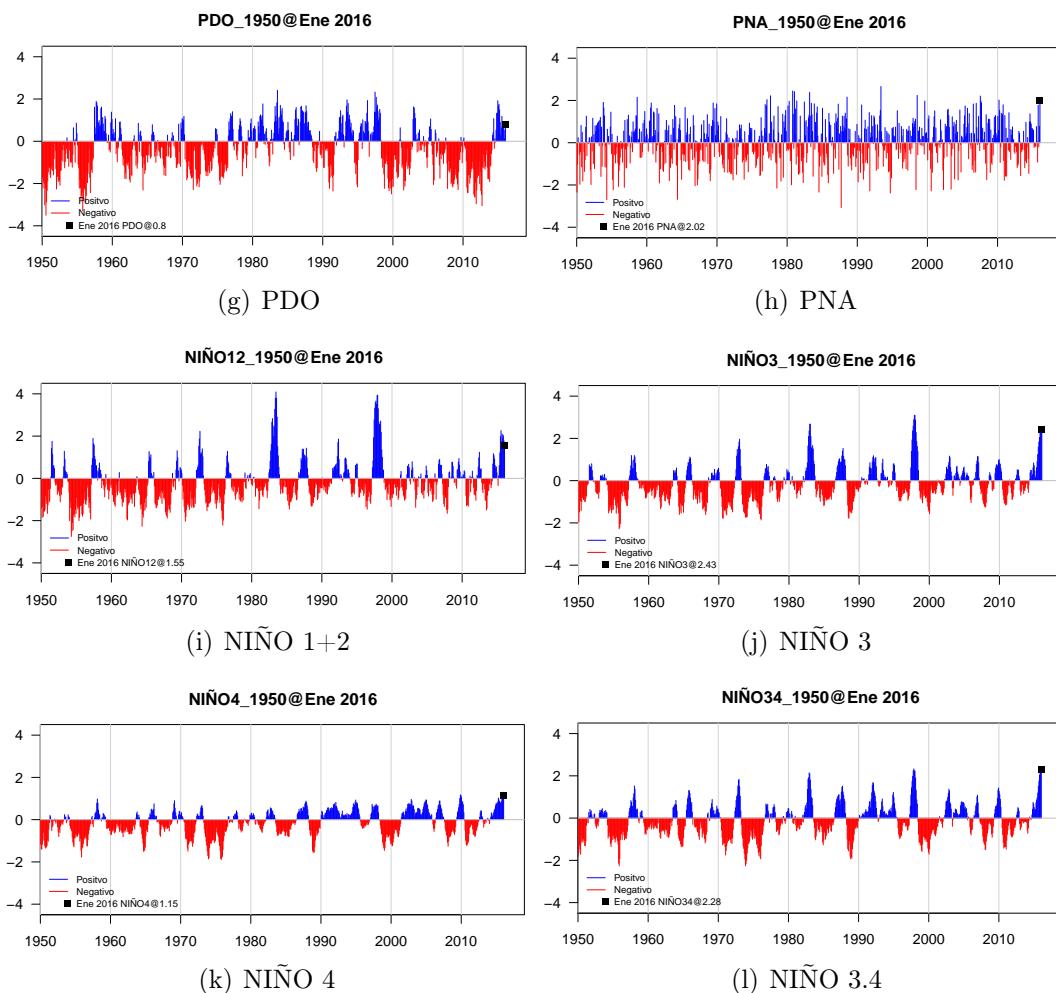
Índice NAO Corresponde al fenómeno climático u Oscilación en el Norte Atlántico. La dirección de descarga es: <http://www.esrl.noaa.gov/psd/data/correlation/nao.data>.

Índice PNA Índice del Pacífico de América del Norte. Es asociado a la Oscilación del EL NIÑO. La fase positiva de PNA, está vinculada a los episodios del calentamiento de Pacífico.

Existen más de 50 índices que presentan los centros mundiales especializados como la NOAA de los Estados Unidos, para analizar el sistema climático de todo el planeta.

Figura 3.21: Series de tiempo de los Índices Climáticos Globales





Fuente: Elaboración propia

3.2.3. Redes Neuronales Artificiales para predicción de sequías

3.2.3.1. Diseño de la arquitectura del modelo

La arquitectura de la red, incluye muchas neuronas interconectadas de capas de ingreso, capa(s) ocultas y salidas (Tayfur, 2012). Para el presente estudio, las neuronas de entrada conocidas también como variables externos del modelo NARX, representan los ICG según la clasificación estadística ya mencionada. En la capa oculta, en número de neuronas se establece mediante el proceso de *ensayo-error* de manera progresiva. En la capa salida, se tiene sólo una neurona denominada también como el índice representativo en la región de análisis.

3.2.3.2. Tipo de Red

El tipo de red es NARX, el cual consiste en dos actividades. El primero, es el paso *Forward* (adelante), propaga desde la capa de ingreso hacia las capas ocultas y capa salida. El segundo corresponde al paso *Backward* (atrás) con conexiones de retro-alimentación.



3.2.3.3. Función de Entrenamiento

La función de entrenamiento para el presente trabajo es **Levenberg-Marquardt**. Es una variación de la técnica de optimización del método de Newton documentado en *Neural Network Design* (Hagan et al., 1996) y *Neural Network Toolbox User Guide* (Beale, Hagan & Demuth, 2014).

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{A}_k^{-1} \mathbf{g}_k \quad (3.73)$$

$$\mathbf{A}_k \equiv \nabla^2 F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k}, \quad \mathbf{g}_k \equiv \nabla F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k}$$

Suponiendo que $F(\mathbf{x})$ es una función de suma de cuadrados.

$$F(\mathbf{x}) = \sum_{i=1}^N v_i^2(\mathbf{x}) = \mathbf{v}(\mathbf{x})^T \mathbf{v}(\mathbf{x}) \quad (3.74)$$

el j -ésimo elemento del gradiente es:

$$[\nabla F(\mathbf{x})]_j = \frac{\partial F(\mathbf{x})}{\partial x_j} = 2 \sum_{i=1}^N v_i(\mathbf{x}) \frac{\partial v_i(\mathbf{x})}{\partial x_j} \quad (3.75)$$

Expresando en forma matricial

$$\nabla F(\mathbf{x}) = 2\mathbf{J}^T(\mathbf{x})\mathbf{v}(\mathbf{x}) \quad (3.76)$$

donde

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial v_1(\mathbf{x})}{\partial x_1} & \frac{\partial v_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial v_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial v_2(\mathbf{x})}{\partial x_1} & \frac{\partial v_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial v_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_N(\mathbf{x})}{\partial x_1} & \frac{\partial v_N(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial v_N(\mathbf{x})}{\partial x_n} \end{bmatrix} \quad (3.77)$$

El elemento k, j de la matriz Hessiana se expresa como

$$[\nabla^2 F(\mathbf{x})]_{k,j} = \frac{\partial^2 F(\mathbf{x})}{\partial x_k \partial x_j} = 2 \sum_{i=1}^N \left[\frac{\partial v_i(\mathbf{x})}{\partial x_k} \frac{\partial v_i(\mathbf{x})}{\partial x_j} + v_i(\mathbf{x}) \frac{\partial^2 v_i(\mathbf{x})}{\partial x_k \partial x_j} \right] \quad (3.78)$$

Expresando en forma matricial

$$\nabla^2 F(\mathbf{x}) = 2\mathbf{J}^T(\mathbf{x})\mathbf{J}(\mathbf{x}) + 2\mathbf{S}(\mathbf{x}) \quad (3.79)$$

donde



$$\mathbf{S}(\mathbf{x}) = \sum_{i=1}^N v_i(\mathbf{x}) \nabla^2 v_i(\mathbf{x}) \quad (3.80)$$

Asumiendo que $\mathbf{S}(\mathbf{x})$ es pequeño, por lo tanto la matriz Hessiana es

$$\nabla^2 F(\mathbf{x}) \cong 2\mathbf{J}^T(\mathbf{x})\mathbf{J}(\mathbf{x}) \quad (3.81)$$

Considerando el método de Newton, se obtiene el algoritmo de **Levenberg Marquardt**

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) + \mu_k \mathbf{I}]^{-1} \mathbf{J}^T(\mathbf{x}_k) \mathbf{v}(\mathbf{x}_k) \quad (3.82)$$

Este algoritmo tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento está determinado por el Error Cuadrático Medio (MSE).

donde:

- \mathbf{x}_k : Vector que incluyen todos los pesos y bias de la red en la iteración k .
 \mathbf{A}_k : Notación matricial a la función $\nabla^2 F(\mathbf{x}_k)$
 \mathbf{g}_k : Notación matricial a la función $\nabla F(\mathbf{x}_k)$
 \mathbf{J} : Matriz Jacobiana

3.3. Evaluación de desempeño

Para la utilización y aceptación de un modelo, se dispone de índices estadísticos e índices de medidas del performance del conjunto de datos observados y calculados.

En la presente investigación, se tienen la patrones de observación y patrones de simulación o predicción, por ejemplo, los modelos de los Algoritmos Genéticos y técnicas numéricas como Máxima Verosimilitud para el cálculo de los parámetros óptimos de la función Gamma, al cual, para su evaluación se emplea el coeficiente de eficiencia (CE) para indicar el mejor modelo. Para la selección y evaluación de los índices de sequías como el VCI, SPI y SPEI, se emplean el coeficiente de Pearson (R), error cuadrático medio (MSE) y el coeficiente de eficiencia (CE). Por último, para la elección de los modelos en las RNAs (arquitectura de la red) se emplea el coeficiente de correlación de Pearson (R) tanto en el entrenamiento, validación, test y modelo global.

Coeficiente de correlación de Pearson (R) Indica el grado de asociación o relación que existe entre dos variables. El rango de R varía desde -1 a $+1$.

$$R = \left[\frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^N (P_i - \bar{P})^2}} \right]^{1/2} \quad (3.83)$$

Tabla 3.18: Escala correlación de Pearson



Valor	Significado
-1	Correlación negativa grande y perfecta
-0.9 a -0.99	Correlación negativa muy alta
-0.7 a -0.89	Correlación negativa alta
-0.4 a -0.69	Correlación negativa moderada
-0.2 a -0.39	Correlación negativa baja
-0.01 a -0.19	Correlación negativa muy baja
0	Correlación nula
0.01 a 0.19	Correlación positiva muy baja
0.2 a 0.39	Correlación positiva baja
0.4 a 0.69	Correlación positiva moderada
0.7 a 0.89	Correlación positiva alta
0.9 a 0.99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

Fuente: Orlando Suárez Mario (2012)

Error Medio Cuadrático (MSE) Consiste en la sumatoria de los errores cuadráticos entre los observados y simulados. El rango es desde 0 hasta $+\infty$, siendo los valores cercanos a cero ajustes perfectos.

$$MSE = \frac{\sum_{i=1}^N (O_i - P_i)^2}{N} \quad (3.84)$$

Coeficiente de eficiencia (CE) También conocido como el coeficiente de Nash Sutcliffe. El rango de CE es desde $-\infty$ (ajuste malo) hasta 1 (ajuste perfecto).

$$CE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (3.85)$$

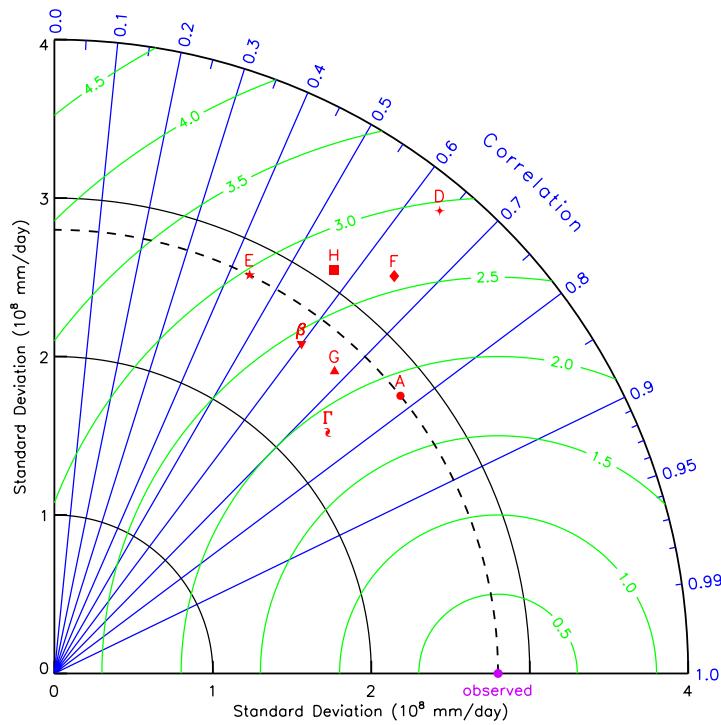
donde:

- N : Número de observaciones
 P_i : Valores simulados o pronosticados
 \bar{P} : Media de los valores simulados
 O_i : Valores observados
 \bar{O} : Media de los valores observados

Diagrama de Taylor Permite comparar varios modelos para determinar cuál de ellos es más predictivo. Recomiendan el uso para modelos climáticos por su simplicidad interpretativa y capacidad de agrupación de estadísticos.

El coeficiente de correlación se muestra en el arco extremo derecho del gráfico (rango de 0 a 1). Los arcos con líneas discontinuas muestran los valores de la desviación estándar con rango de 0 al valor máximo de desviación. El arco con líneas continuas representa la desviación estándar de los valores observados y se usa como referencia. Por último, los arcos contrarios a la desviación, representan el estadístico MSE que están en función de 0 al valor máximo de error cuadrático.

Figura 3.22: Diagrama de Taylor



Fuente: Adaptado desde

http://www-pcmdi.llnl.gov/about/staff/Taylor/CV/Taylor_diagram_examples.pdf

Capítulo IV: Análisis y resultado de la investigación

En el presente capítulo se muestran los resultados y el análisis de acuerdo a la metodología propuesta en el capítulo anterior. Inicia mediante el análisis exploratorio de las variables regionales locales como son las precipitaciones, temperaturas, altitudes y algunos índices externos que permiten caracterizar y describir la cuenca de estudio. Todos los resultados son tabulares y gráficos construidos mediante paquetes informáticos como R, Python, Matlab, REFRAN-CV, ICI-RAFT, entre otros.

Se realiza el análisis regional de frecuencias de las sequías y se determinan zonas húmedas y áridas en la cuenca de estudio, permitiendo la construcción de mapas con diferentes períodos de retorno para la evaluación de las sequías regionales. Este capítulo, finaliza con los resultados de los modelos predictivos de las sequías a través de las variables o índices climáticos globales.

4.1. Análisis exploratorio de las variables regionales

Dentro de la cuenca Mantaro, se evaluaron 7 variables regionales: la altitud (ALT), precipitación media anual (PMA), temperatura mínima (TMIN), temperatura máxima (TMAX), índice de aridez (IAR), índice de Fournier modificado (IFM) y el índice de estacionalidad (IE).

La información de las precipitaciones y temperaturas, han sido proporcionadas por PISCO-SENAMHI en formatos rasters extensión *.tif*. Cada celda es considerada como una estación sintética, por lo tanto, a lo largo de la cuenca se obtienen 1145 estaciones sintéticas.

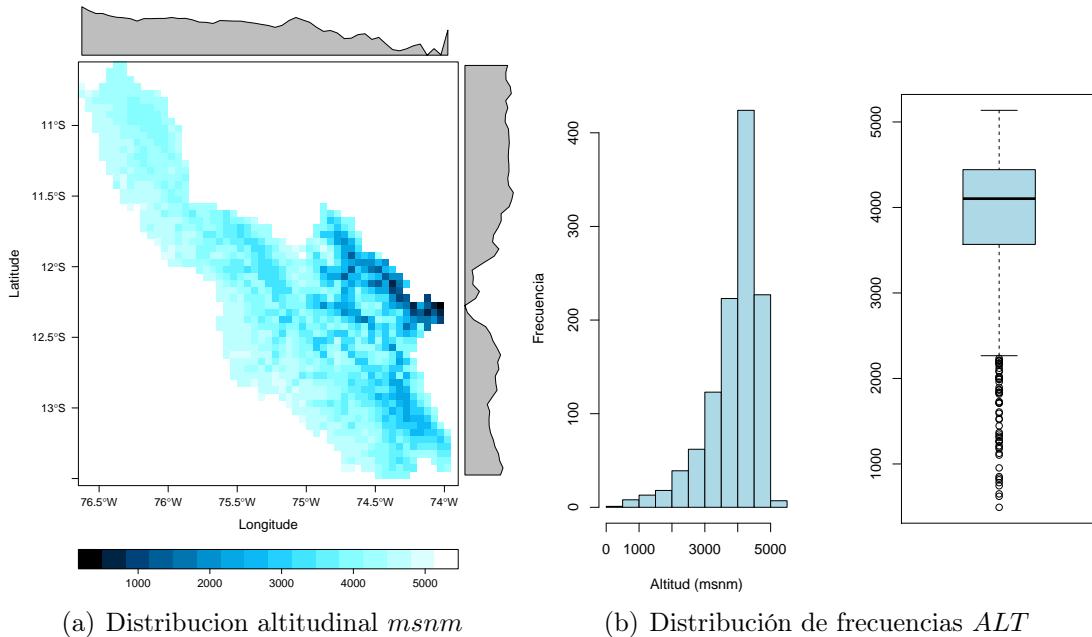
Tabla 4.19: Correlación de variables regionales

Variable	LON	LAT	ALT	PMA	TMIN	TMAX	IAR	IFM	IE
LON	1.00	-0.81	-0.55	0.02	0.23	0.42	-0.13	0.08	0.15
LAT	-0.81	1.00	0.16	-0.02	0.15	-0.07	0.04	-0.09	-0.19
ALT	-0.55	0.16	1.00	0.00	-0.71	-0.70	0.20	-0.03	-0.08
PMA	0.02	-0.02	0.00	1.00	-0.05	-0.04	0.94	0.93	-0.25
TMIN	0.23	0.15	-0.71	-0.05	1.00	0.92	-0.31	-0.07	-0.06
TMAX	0.42	-0.07	-0.70	-0.04	0.92	1.00	-0.35	-0.04	-0.01
IAR	-0.13	0.04	0.20	0.94	-0.31	-0.35	1.00	0.89	-0.21
IFM	0.08	-0.09	-0.03	0.93	-0.07	-0.04	0.89	1.00	0.12
IE	0.15	-0.19	-0.08	-0.25	-0.06	-0.01	-0.21	0.12	1.00

Fuente: Elaboración propia

Variable Altitud: La figura 4.23a muestra el raster de la variable altitud con sus respectivas variaciones de la magnitud (ver superior y extremo derecho de la figura 4.23a) en las direcciones longitud y latitud y la figura 4.23b, se aprecia el diagrama de frecuencias y diagrama de caja para el análisis estadístico. La altitud se concentra sobre los 4000 msnm, siendo 493 y 5136 msnm el mínimo y máximo respectivamente. La variación es descendente con la longitud y también descendente central con la latitud.

Figura 4.23: Distribución altitudinal cuenca del Mantaro



Fuente: Elaboración propia

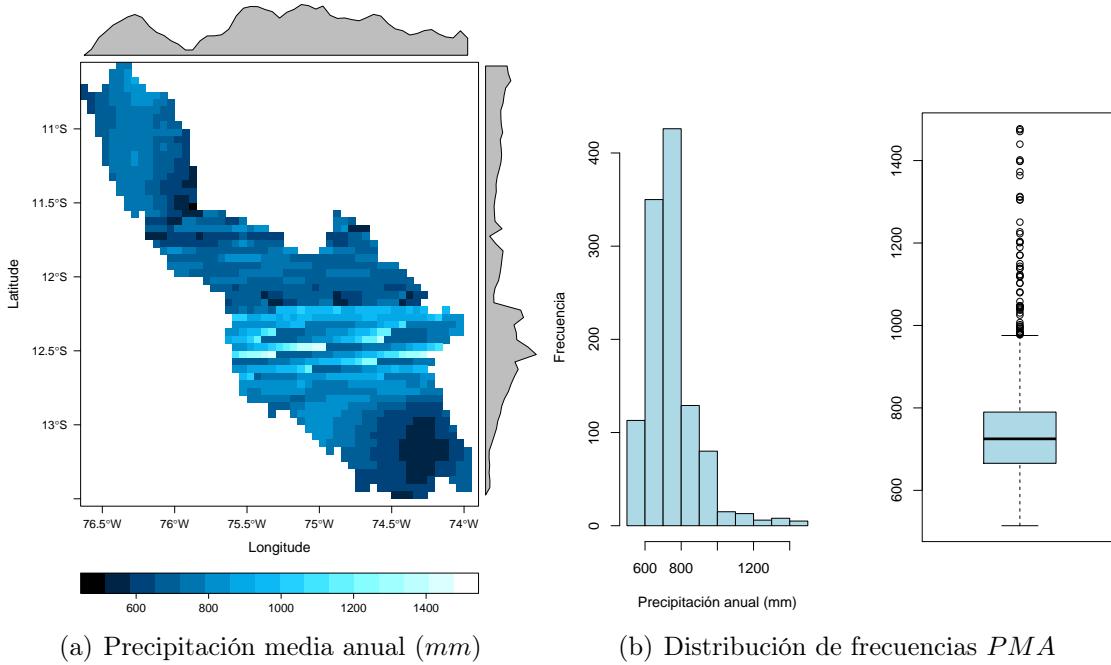
Variable Precipitación Media Anual: Se observa una precipitación media anual de 744.6 mm (4.24b), existen áreas con precipitaciones sobre los 1400 mm en la latitud central ligeramente desplazada hacia el sur. La PMA mínimo y máximo es de 514.2 y 1477 mm respectivamente.

También, se observa en la región sur precipitaciones menores a 600 mm, región que marca interés para el análisis se sequía. Según la tabla 4.19, la correlación con la variable altitud es nula en toda la región, de esta manera se confirma que sólo en regiones pequeñas existe correlación entre la PMA y la altitud.

Variable Temperatura Mínima: Esta variable varía entre -1.43 y 11.96°C (mínimo y máximo respectivamente) y se distribuye en promedio de 4.56°C. Esta variable presenta una correlación (Pearson) aceptable e indirecta de -0.71 con la variable altitud (tabla 4.19), también 0.23 con la longitud (directa). Por lo tanto, las temperaturas

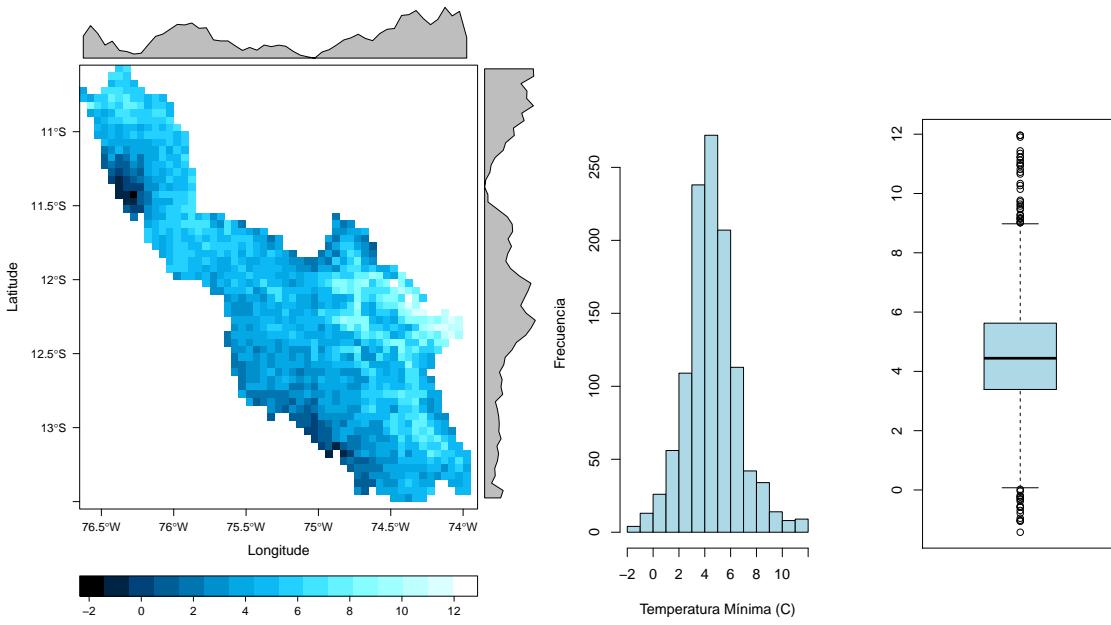
mínimas se distribuyen en las montañas alto-andinas de la cuenca del mantaro (figura 4.25). También es una variable importante para la selección o proceso de regionalización de la cuenca en estudio.

Figura 4.24: Distribución de Precipitaciones media anual



Fuente: Elaboración propia

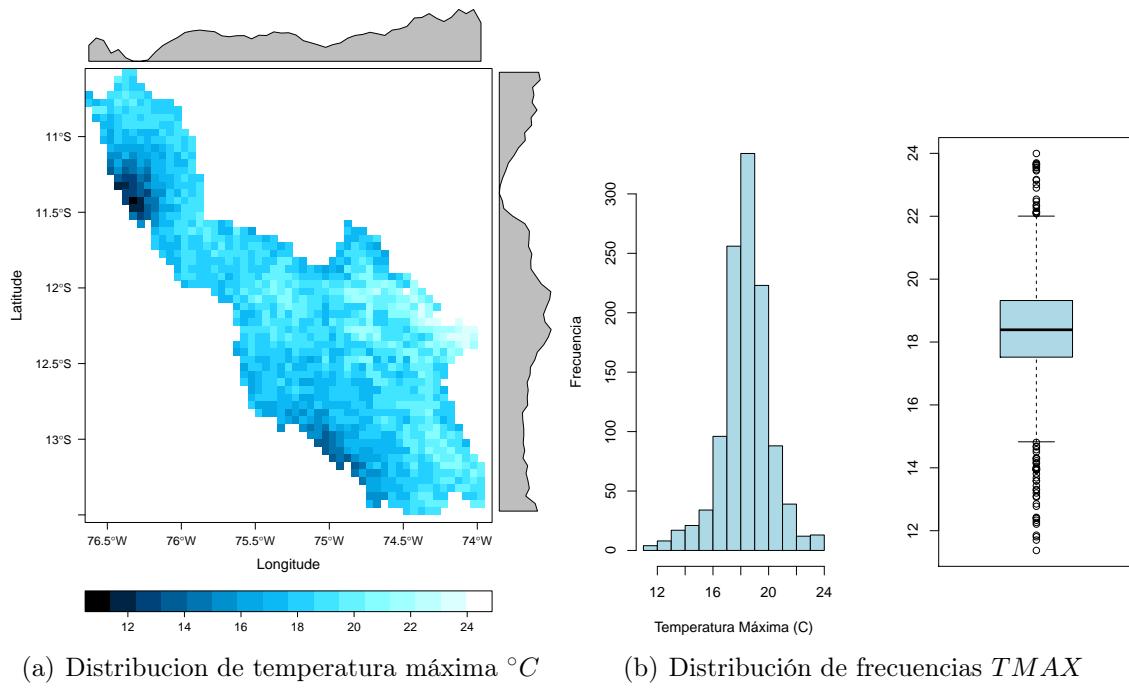
Figura 4.25: Distribución de Temperatura mínima



Fuente: Elaboración propia

Variable Temperatura Máximas: Varía entre 11.37 hasta 24°C mínimo y máximo respectivamente (figura 4.26b) y promedio de 18.35°C. La correlación con la variable altitud es directa (0.70). La distribución de esta variable es con tendencia a la longitud (figura 4.26a), en cuanto a la latitud es casi uniforme, excepto entre las latitudes 11.5° y 11°, donde se observa temperaturas menores a 12°C. Se concluye que las temperaturas máximas, se concentran en las zonas de menor altitud (región Este).

Figura 4.26: Distribución de Temperatura máxima



(a) Distribucion de temperatura máxima °C

(b) Distribución de frecuencias TMAX

Fuente: Elaboración propia

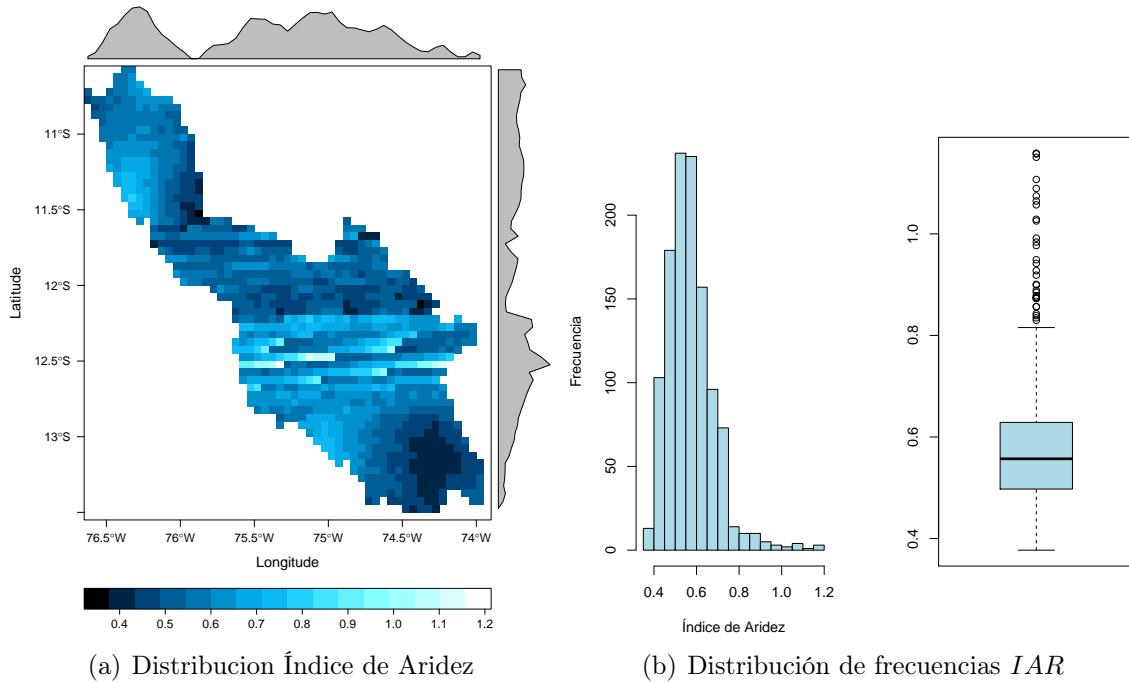
Variable Índice de Aridez: Presenta alta correlación directa con la variable PMA (0.94), y correlación indirecta (-0.34 y -0.35) con las temperaturas mínima y máxima respectivamente. Este índice ha sido calculado con la precipitación promedio y la evapotranspiración potencial (método Hargreaves). La distribución en la cuenca es de acuerdo a la figura 4.27: promedio igual a 0.57 y según la tabla 2.7, corresponde a la clasificación **Subhúmedas secas** y existe una zona bastante pronunciada a la clasificación **semiárida** (región sur, departamento de Ayacucho).

El índice IAR mínimo es igual a 0.38, por lo que se concluye que en la cuenca del Mantaro, no existen regiones áridas ni mucho menos hiperáridas (según la tabla 2.7). Asimismo, la figura 4.27b muestra los outliers (1.16 máximo), que indica la existencia de algunas regiones húmedas en la latitud 12.5° Sur.

Variable Índice de Fournier Modificado: El promedio de este índice es 107.8 (figura 4.28a), según a la tabla 2.8 corresponde a la clasificación de **moderado**, es decir, en toda la cuenca, la agresividad climática equivalente a la erosión hídrica es intermedia.

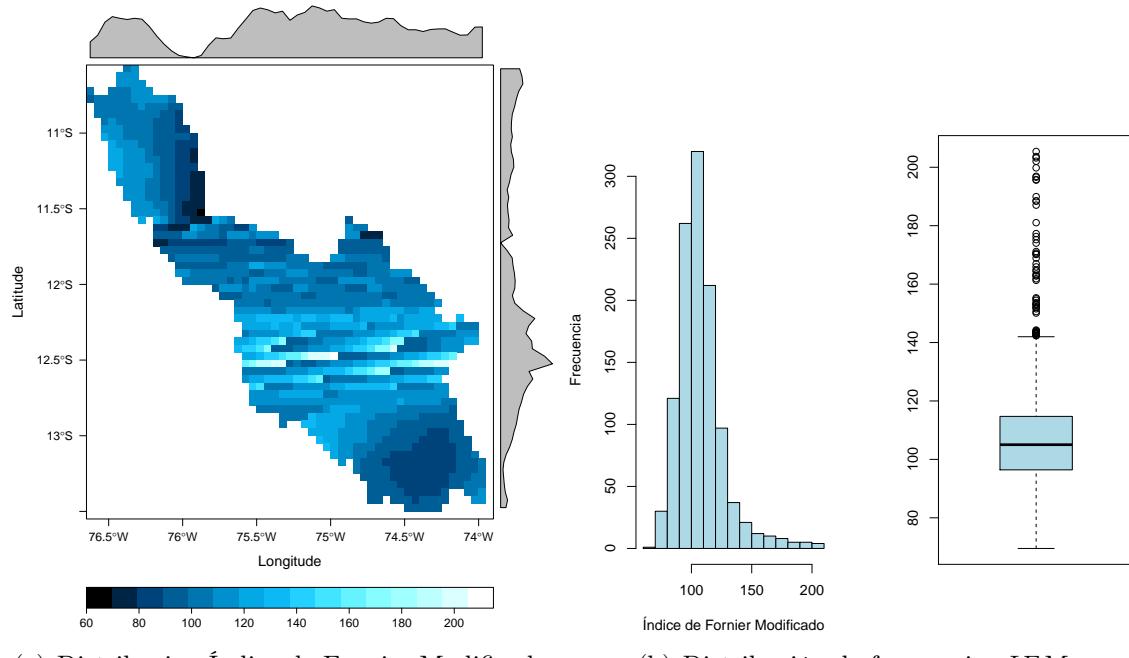
Por otro lado, los outliers de la figura 4.28b indican alta y muy alta agresividad (205.4 máximo), ubicada en la latitud 12.5° Sur.

Figura 4.27: Distribución de Índice de Aridez



Fuente: Elaboración propia

Figura 4.28: Distribución del Índice de Fournier Modificado

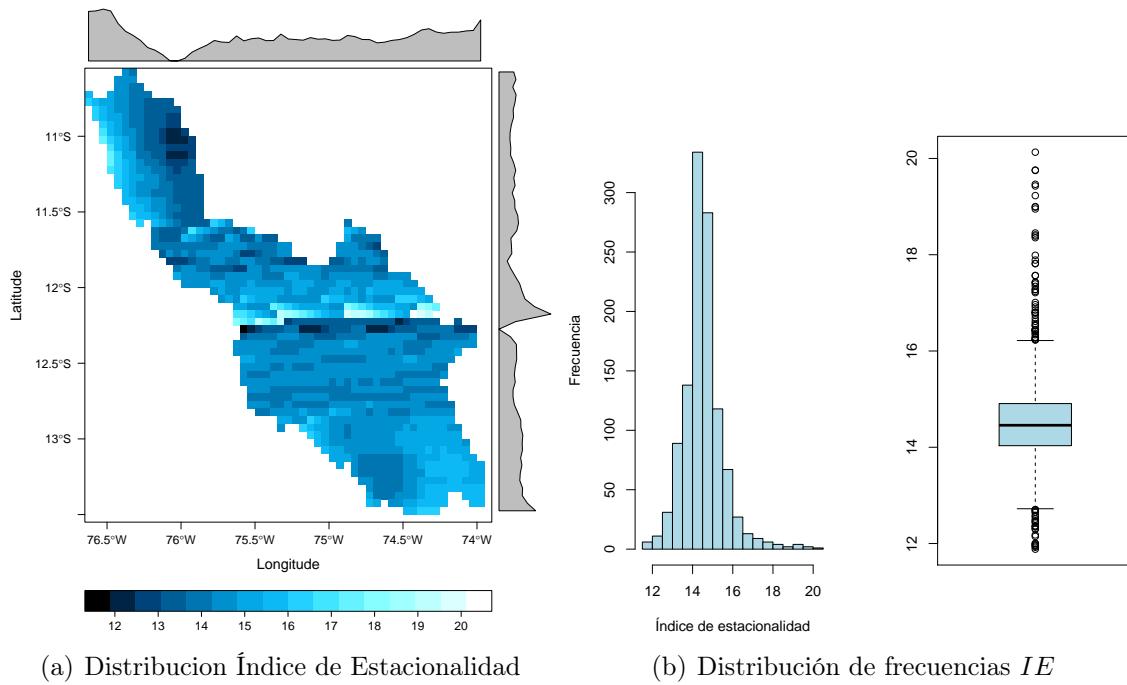


Fuente: Elaboración propia

Variable Índice de Estacionalidad: Este índice se distribuye en promedio igual a 14.53 y según la tabla 2.9 corresponde a **moderadamente estacional** (próximo a estacional), es decir, confirma que las precipitaciones en la cuenca del Mantaro se producen en los meses diciembre-abril.

Existe una región entre las latitudes 11.5°-11.0° sur con IE mayores a 17, cuya clasificación es **estacional**.

Figura 4.29: Distribución del Índice de Estacionalidad



Fuente: Elaboración propia

4.2. Clasificación Regional cuenca del Mantaro mediante SOM

La primera clasificación regional, es a través de la metodología SOM, con el propósito de reducir las regiones heterogéneas (grillas con valores atípicos, tendencias y no estacionarios), que no permitían en primera instancia, la homogenización regional mediante la técnica de L-Momentos. Para éste propósito, son considerados las siguientes variables regionales: longitud, latitud, altitud, precipitación media anual, temperatura mínima, y temperatura máxima. El índice de aridez, Fournier y estacionalidad, no son considerados por la correlación alta que tiene con las variables ya mencionadas (se evitó redundancia de variables clasificadorias).

También, este primer procedimiento es considerado para la reducción de los datos, los cuales son extensos en el espacio y relaciones estadísticas como las medidas de tendencia central, no permiten la clasificación o agrupación regional según las características o variables regionales expuestas. El modelo de inteligencia artificial auto organizado

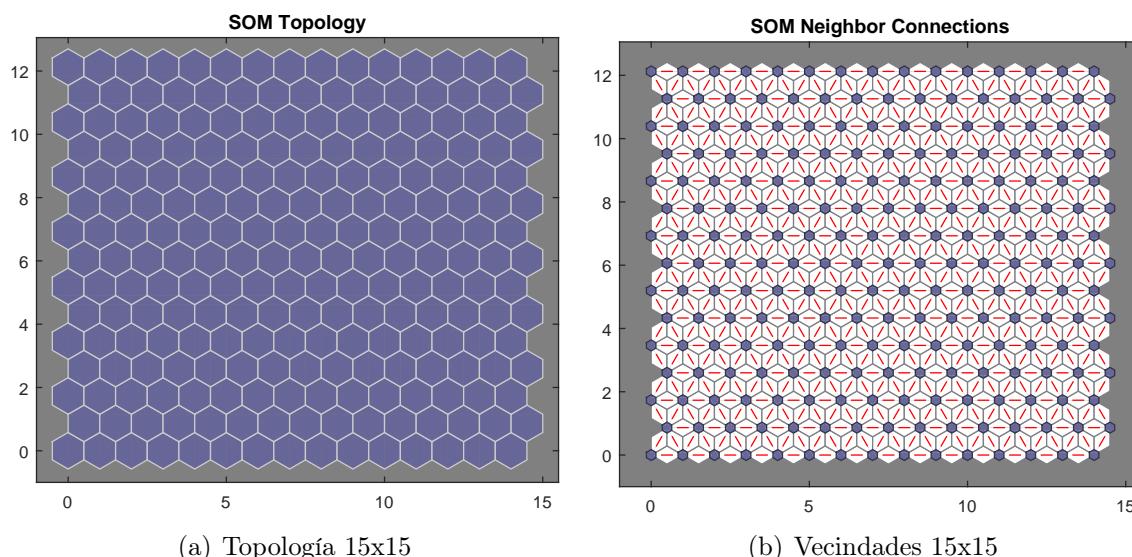
o también denominado la red neuronal de Kohonen, ha demostrado ser eficiente y robusto en la clasificación de patrones con grandes volúmenes de información. Utilizando el comando *selforgmap* del lenguaje de programación Matlab, se crea un mapa auto organizado y se definen los parámetros como dimensión (cantidad de neuronas de salida), tamaño inicial de la vecindad (por defecto, Matlab considera 3), topología de la red (por defecto es hexagonal), y función de distancia (por defecto es la función de distancia euclídea).

Para la clasificación SOM, se establecieron varios tipos de arquitectura, siendo el más adecuado la red de 225 NS (15x15), es decir, las neuronas clasificadorias están en equilibrio, ni muy saturadas tampoco nulas. En todos los casos la topología ha sido tipo hexagonal y la función de distancia euclídea. En entrenamiento para todo los casos, se han desarrollado en 50000 épocas ó ciclos, para observar el rendimiento en cada caso.

Las figuras 4.30a y b, muestran la topología de la red tipo rejilla hexagonal y las conexiones entre ellas respectivamente, es decir, que cada neurona central (sin considerar los bordes), tiene 6 neuronas vecinas con las cuales se conectan virtualmente. Las líneas horizontales y diagonales de la figura 4.30b, muestran las conexiones entre neuronas aledañas.

El modelo, se realizó con 50000 iteraciones (figura 4.31), obteniendo los resultados en formatos gráficos, y tabulares como los pesos ponderados de cada variable de ingresos según clasificación.

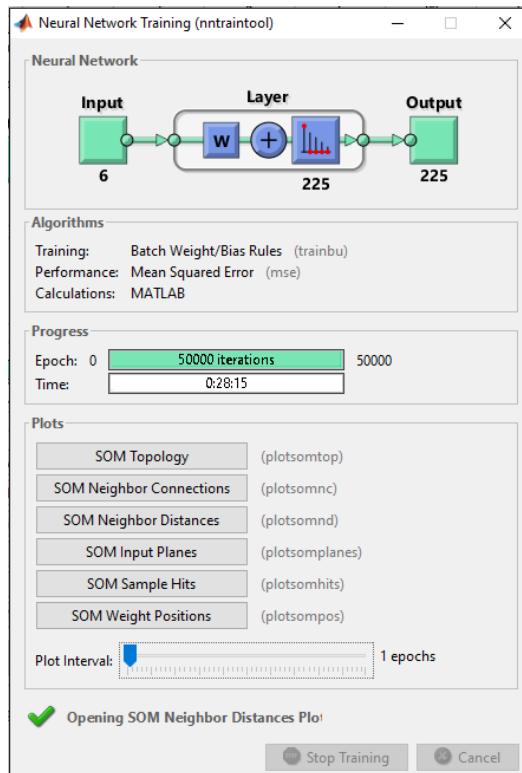
Figura 4.30: Topología y conexión de vecindades del modelo auto organizado



Fuente: Elaboración propia

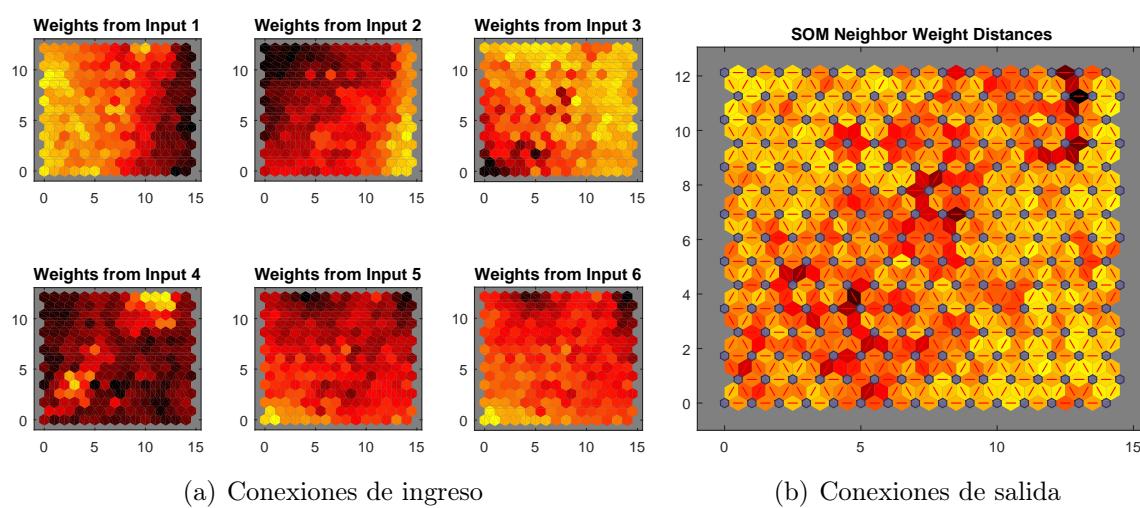
La figura 4.32, conocido como la representación U-Matrix (Unified distance Matrix), muestra la distancia entre neuronas adyacentes. La visualización es mediante los colores claros y oscuros. El color oscuro entre las neuronas, indica una distancia grande que representa un espacio importante entre los valores de los patrones de entrada y el

Figura 4.31: Ventana gráfica Matlab usando SOM



Fuente: Software Matlab

Figura 4.32: Distancias y pesos entre neuronas vecinas de los parámetros de ingreso y salida



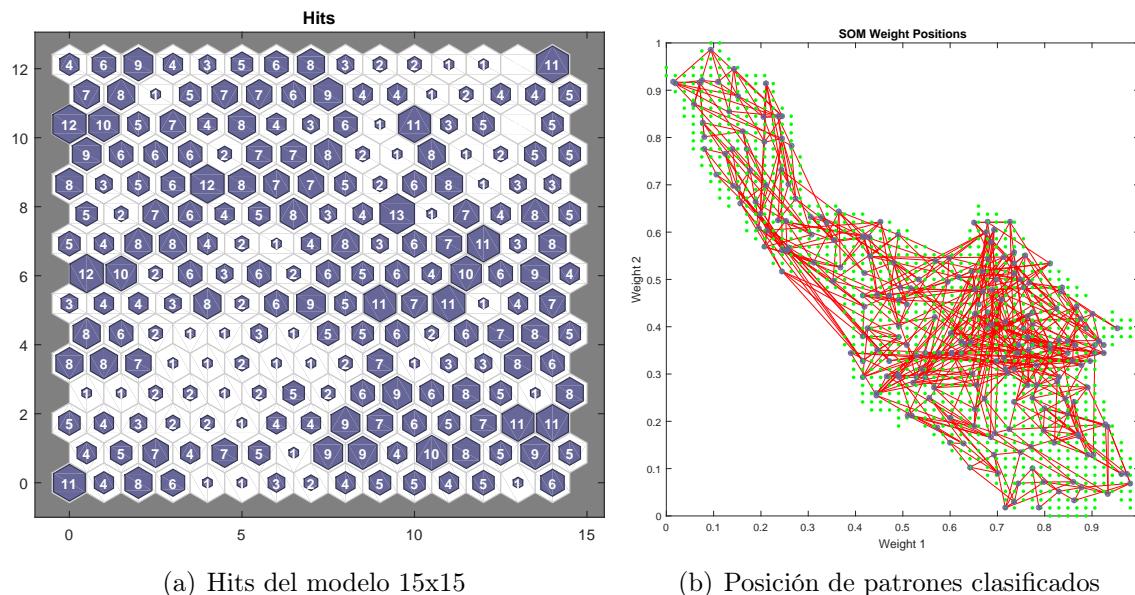
Fuente: Elaboración propia

color claro, significa que las neuronas están cerca unas de otras. Este mapa, es considerado el más importante para seleccionar la arquitectura y el número de épocas en el entrenamiento del modelo SOM.

La figura 4.33a, muestra el agrupamiento que presenta los mapas hits o ganadoras. El tamaño de la figura hexagonal es proporcional a las cantidades de los patrones o vectores de entrada, es decir, cada neurona representa el número de NE que han sido clasificados en las NS del mapa topológico. Se recomienda la normalización de los datos entre 0 y 1, sólo para propósitos de clasificación, es decir, sin resultados de los pesos w_i . Para obtener los pesos (ponderados de las variables), se debe realizar el entrenamiento SOM con los valores reales. En la misma figura, se observa que el mapa topológico, presenta al menos una NE, salvo dos en las neuronas de salida 194 y 224. Por lo tanto, la clasificación se llevó a cabo en 223 neuronas, considerados en posterior como 223 estaciones regionales reducidas.

En la misma figura, se muestra una clasificación variada. Existen NS que clasifican desde 1 NE hasta 13 NE (Neurona 145).

Figura 4.33: Hits del modelo auto organizado y posición de patrones clasificados

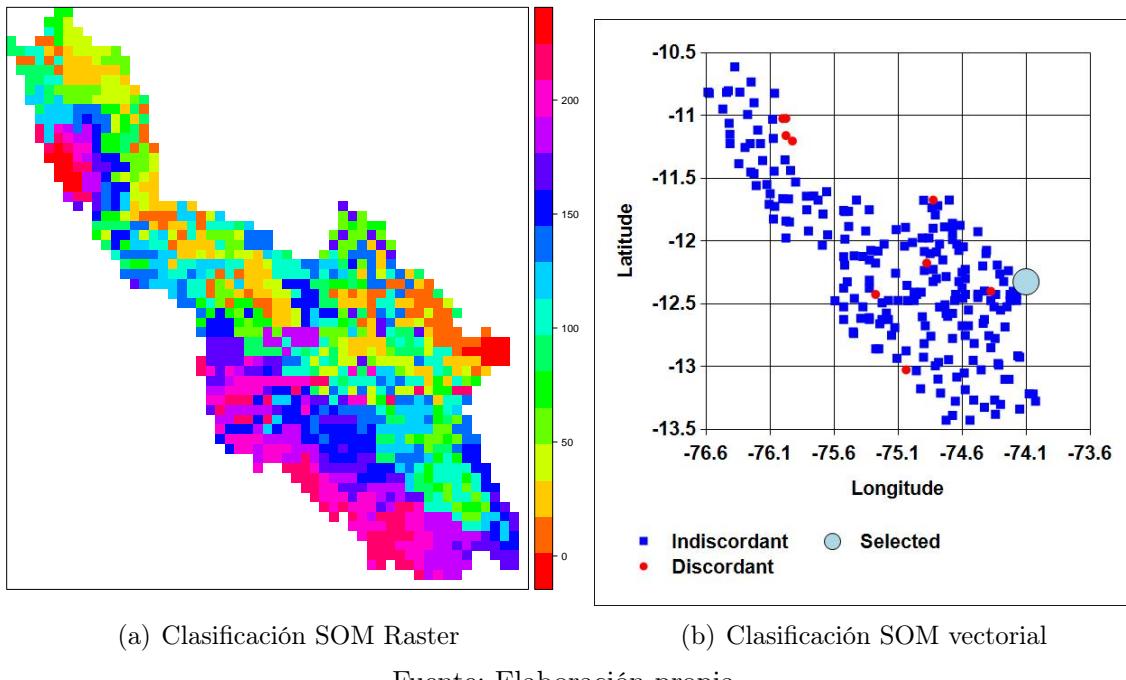


Fuente: Elaboración propia

La figura 4.33b, muestra el mapa de localización de los pesos de las neuronas y su correspondiente organización dentro del espacio real de la información. Éste gráfico descarta el espacio topológico. Con diferentes escenarios, se ha observado que a mayor neuronas de salida y épocas de entrenamiento, el clustering o agrupamiento mejora.

En la figura 4.34a, se muestra el mapa tipo raster de acuerdo a las neuronas de clasificación de la cuenca Mantaro, y la figura 4.34b muestra la clasificación tipo vector, donde se observa 223 neuronas o estaciones (2 neuronas nulas). En base al procedimiento *ensayo y error*, se determinó esta clasificación de la malla topológica igual a 225 neuronas y finaliza éste método (SOM) con los ponderados de los pesos de cada NE, conocidos como variables regionales de entrada.

Figura 4.34: SOM, clasificación regional según el número de neuronas



Fuente: Elaboración propia

Esta primera clasificación permitió: primero, la reducción espacial de los datos y segundo, el control y reducción de datos heterogéneos, como valores atípicos, tendentes, entre otros, que no permiten o alteran la utilización de metodologías estadísticas de manera eficiente. Por lo tanto, las estaciones de análisis son 223 distribuidos dentro de la cuenca del Mantaro.

4.3. Análisis regional de frecuencias de la sequía en la cuenca del Mantaro.

4.3.1. Identificación de regiones homogéneas

La región total o super región, presentó una discordancia D_i desde 0.01821 hasta 5.5783 y H_1 igual a 6.148, identificándose como región altamente heterogénea.

Para la regionalización, se ha realizado varias pruebas, en función al número de regiones y pesos de las características de las variables climáticas regionales. Inicialmente se probó con 12 regiones, sin embargo, en algunas regiones se clasifican sólo 6 estaciones y según Hosking y Wallis (1997), recomienda mínimo 7 para tener una buena estimación en la discordancia de las estaciones. Considerando un peso del 40 % a las PMA, siempre se clasifica una región con 6 grupos (estaciones 64, 206, 207, 220, 221 y 222) de PMA aproximado a 1400 mm a lo largo de la latitud 12.5° sur. Es así, que obliga a disminuir la cantidad de regiones a 10, consiguiendo el número de estaciones adecuados por región



y los test de la homogeneidad dentro de lo permitido según Wallis et al. (2007), tal como se muestran en las tablas posteriores.

Ha sido necesario reubicar, es decir, mover estaciones de una región a otra y eliminar algunas estaciones para evitar la discordancia de acuerdo al indicador D_i . Para realizar éste proceso de clasificación, se usó el software ICI-RAFT considerando los pesos de la alternativa 4 de la tabla 4.20.

Tabla 4.20: Variables climáticas regionales y asignación de pesos

Variable	Alternativa 1	Alternativa 2	Alternativa 3	Alternativa 4
LON	0.30	0.10	0.05	0.10
LAT	0.20	0.30	0.40	0.30
ALT	0.10	0.30	0.05	0.20
PMA	0.30	0.20	0.40	0.30
TMIN	0.05	0.05	0.05	0.05
TMAX	0.05	0.05	0.05	0.05

Fuente: Elaboración propia

En la figura 4.35, se muestra las 10 regiones con sus respectivos grupos o estaciones, ubicadas geográficamente dentro de la cuenca del Mantaro.

Región 1 : PMAR = 790.16 mm, geográficamente se ubica al este y agrupa a 11 estaciones. La discordancia máxima = 2.274 y mínima = 0.244 (Tabla E.2-sección E).

Región 2 : PMAR = 676.98 mm y agrupa la mayor cantidad de estaciones (46). Discordancia máxima = 2.447 y mínima = 0.052.

Región 3 : PMAR = 1010.29 mm, considerado como la segunda región húmeda y agrupa a la menor cantidad de estaciones (14). Discordancia máxima = 1.6993 y mínima = 0.088.

Región 4 : PMAR = 745.92 mm, geográficamente es discontinua y agrupa a 25 estaciones. Discordancia máxima = 1.796 y mínima = 0.042.

Región 5 : PMAR = 728.97 mm, y se ubica al norte de la cuenca con gran altitud y agrupa a 28 estaciones. Discordancia máxima = 2.212 y mínima = 0.06.

Región 6 : PMAR = 608.72 mm, considerado como la región seca o semiárida según el índice de Aridez descrito anteriormente, y se ubica al sur de la cuenca (departamento de Ayacucho) agrupando a 18 estaciones. Discordancia máxima = 1.936 y mínima = 0.205.

Región 7 : PMAR = 1280.66 mm, considerado la región más húmeda de la cuenca de estudio, y se ubica paralelo a la latitud 12.5° sur y agrupa a 17 estaciones. Discordancia máxima = 2.238 y mínima = 0.435.

Región 8 : PMAR = 896.03 mm, presenta agrupamiento discontinuo, distribuyéndose en la zona centro de la cuenca y agrupa a 21 estaciones. Discordancia máxima = 2.748 y mínima = 0.240.



Región 9 : PMAR = 735.19 mm, ubicado al oeste de la región sur con gran altitud y agrupa 15 estaciones. Discordancia máxima = 2.026 y mínima = 0.046.

Región 10 : PMAR = 697.10 mm, inclinado hacia la zona norte y agrupa 17 estaciones. Discordancia máxima = 2.578 y mínima = 0.063.

En la tabla 4.21, se muestra las principales características de los L-Momentos y su respectivo número de estaciones por cada región (No E). Las principales características son: la media (l_1), y los L-Momentos ratio (τ_1, τ_3, τ_4 y τ_5).

Tabla 4.21: Principales características L-Momentos Regionales de las regiones homogéneas

L-M	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
No E	11	46	14	25	28	18	14	21	15	17
l_1	790.16	676.976	1010.29	745.922	728.97	608.723	1280.66	896.04	735.19	697.10
τ_1	0.117	0.113	0.134	0.125	0.109	0.13	0.134	0.138	0.149	0.112
τ_3	-0.016	-0.054	-0.023	0.035	-0.038	0.016	-0.099	-0.009	-0.038	-0.011
τ_4	0.154	0.133	0.114	0.116	0.078	0.078	0.109	0.107	0.06	0.129
τ_5	-0.028	-0.027	-0.03	0.013	-0.025	-0.003	-0.033	-0.006	0.028	-0.024

Fuente: Elaboración Propia

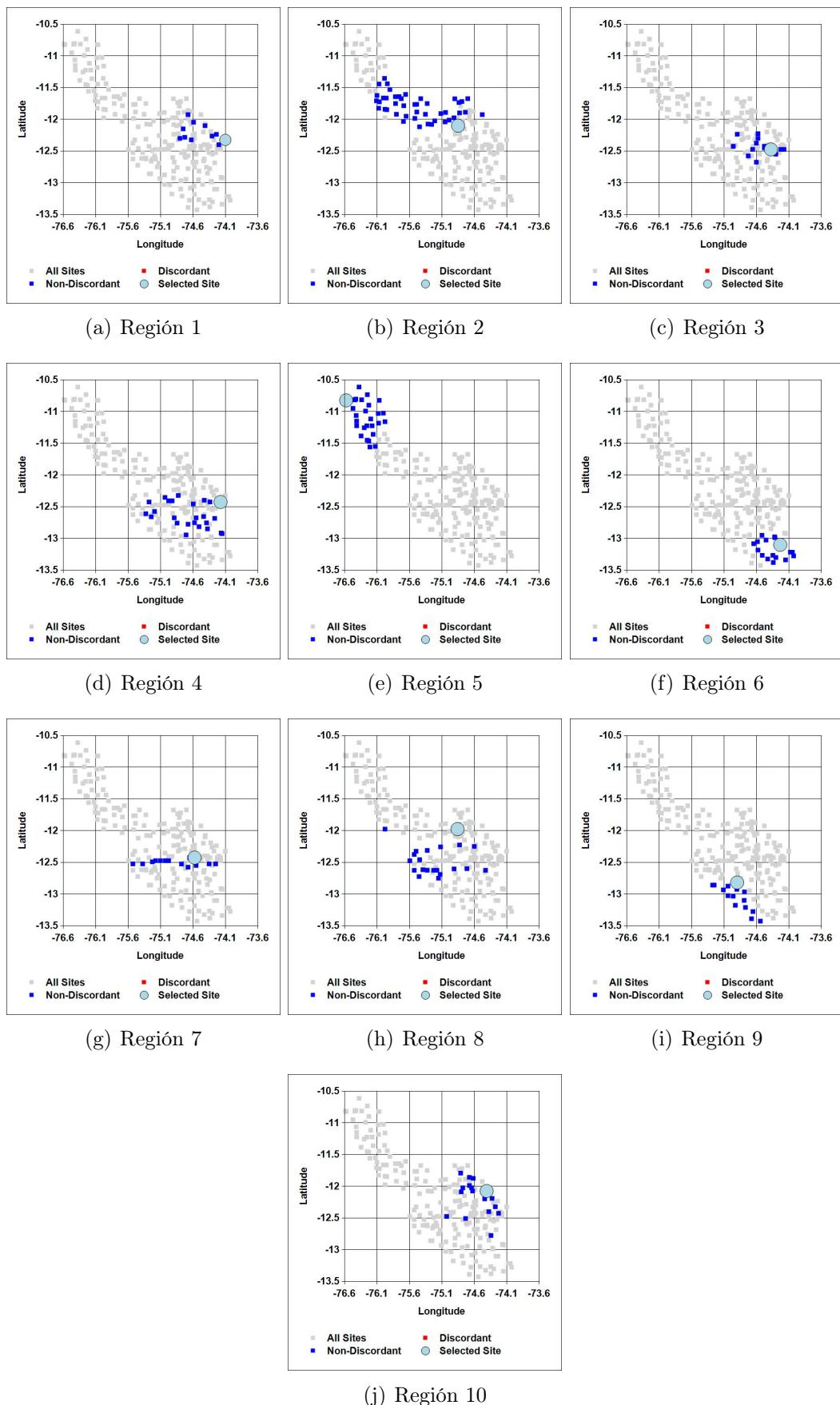
Una vez agrupado y aprobado mediante el test de discordancia (menores a 3 en todos los casos), se determinan el test de homogeneidad de cada región. En la tabla 4.22, se tienen los estadístico H_1 , H_2 y H_3 para cada región, siendo el estadístico H_1 el más riguroso para la evaluación de la homogeneidad. La variación de H_1 es desde -3.268 hasta 2.332 correspondientes al mínimo y máximo que se dan en las regiones 7 y 8 respectivamente. Todas las regiones presentan valores de H_1 menores a 3, por lo tanto, se consideran regiones homogéneas según Wallis et al. (2007).

Tabla 4.22: Estadísticos H para evaluar la homogeneidad de cada región

Estadístico	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
H_1	0.868	-2.424	-0.416	1.798	-2.117	-2.548	-3.268	2.332	0.24	-0.386
H_2	-1.753	-5.472	-2.542	-3.811	-4.146	-3.202	-1.867	-0.81	-1.245	-0.451
H_3	-1.736	-5.002	-2.941	-3.43	-4.153	-4.476	-2.862	-0.729	-1.671	-1.094

Fuente: Elaboración Propia

Figura 4.35: Ubicación de las regiones homogéneas con sus respectivas estaciones



Fuente: Software ICI-RAFT

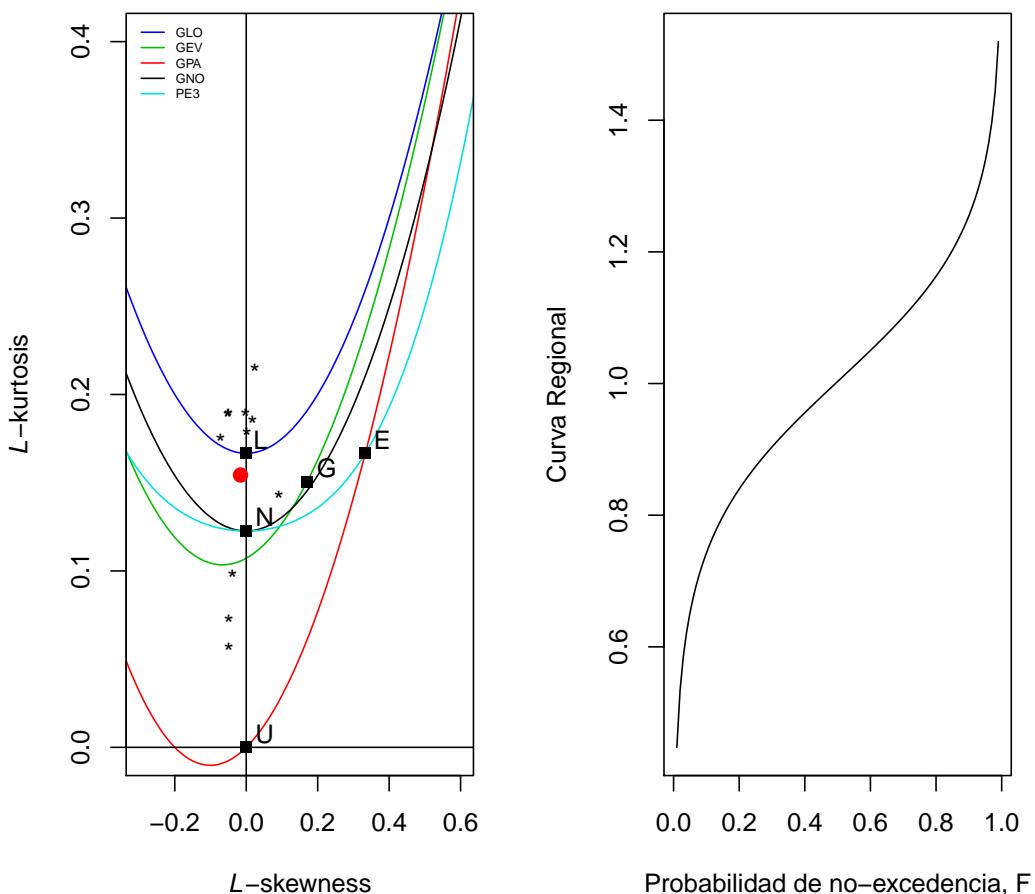
Predicción de sequías con Redes Neuronales Artificiales y Algoritmos Genéticos utilizando precipitación por percepción remota.

Iván Arturo Ayala Bizarro

4.3.2. Identificación de la función de distribución de probabilidades

Haciendo uso del software REFRAN-CV y R, se determinan las mejores distribuciones de probabilidad (cinco más comunes descritas en la sección 2.3.2) para cada región. La figura 4.36, representa el diagrama de L-Momentos para la región 1, (para las demás regiones ver anexo E) y se observa que la tendencia es hacia la distribución Logística Generalizada (GLO).

Figura 4.36: L Momentos Región 1



Fuente: Software ICI-RAFT

Tabla 4.23: Estadístico Z para evaluar la distribución de probabilidades

Distribución	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
GLO	0.597	4.089	3.707	4.766	9.598	7.33	4.485	5.068	8.706	2.693
GEV	-2.66	-3.41	-0.477	-0.195	2.937	2.656	-0.229	0.078	3.724	-1.541
GNO	-1.745	-0.971	0.734	0.82	4.991	3.757	1.506	1.436	5.259	-0.378
PE3	-1.752	-1.132	0.717	0.765	4.915	3.747	1.192	1.432	5.203	-0.382
GPA	-8.428	-16.175	-7.833	-9.495	-8.582	-5.918	-7.89	-8.824	-4.895	-9.077

Fuente: Elaboración Propia



El estadístico Z para cada distribución de probabilidad, se resume en la tabla 4.23, y se observa que la región 1 se ajusta mejor a la distribución Logística Generalizada (GLO), las regiones 2 y 10 a la distribución Normal Generalizada (GNO) y las demás regiones (3, 4, 5, 6, 7, 8 y 9) a la distribución Valores Extremos Generalizada (GEV).

En la tabla 4.24, se resume los parámetros de cada distribución de probabilidades elegidas de acuerdo al estadístico Z . El primer parámetro corresponde a la ubicación, el segundo a la escala y el último a la forma.

Tabla 4.24: Distribuciones de probabilidades para cada región

Característica	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
Distribución	GLO	GNO	GEV	GNO						
Param 1	1.003	1.011	0.921	0.916	0.939	0.916	0.938	0.916	0.916	1.002
Param 2	0.117	0.199	0.242	0.213	0.199	0.226	0.256	0.245	0.273	0.198
Param 3	0.016	0.11	0.325	0.221	0.354	0.255	0.469	0.301	0.354	0.023

Fuente: Elaboración Propia

4.3.3. Determinación de Cuantiles

Definido las distribuciones con sus respectivos ajustes, se calculan los cuantiles regionales para cada probabilidad de no excedencia (P_{NE}) y respectivo Periodo de retorno (T_R).

Tabla 4.25: Cuantiles Regionales asociados a distintos periodos de retorno

P_{NE}	T_R (Años)	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
0.99	1	1.765	1.532	1.587	1.669	1.452	1.65	1.463	1.63	1.62	1.594
0.5	2	1.003	1.011	1.005	0.991	1.007	0.995	1.024	1.001	1.01	1.002
0.33	3	0.922	0.923	0.898	0.896	0.919	0.895	0.913	0.893	0.89	0.916
0.25	4	0.874	0.872	0.837	0.844	0.87	0.839	0.847	0.832	0.821	0.867
0.2	5	0.84	0.836	0.796	0.809	0.835	0.802	0.801	0.791	0.774	0.834
0.167	6	0.813	0.808	0.765	0.784	0.81	0.774	0.765	0.76	0.739	0.808
0.143	7	0.791	0.786	0.741	0.764	0.789	0.752	0.737	0.735	0.711	0.788
0.125	8	0.773	0.767	0.721	0.747	0.772	0.734	0.713	0.715	0.688	0.771
0.11	9.	0.757	0.751	0.704	0.733	0.758	0.719	0.693	0.698	0.668	0.757
0.1	10	0.742	0.737	0.689	0.721	0.745	0.706	0.676	0.684	0.651	0.744
0.067	15	0.689	0.686	0.636	0.679	0.701	0.66	0.612	0.631	0.59	0.699
0.05	20	0.652	0.652	0.601	0.652	0.671	0.63	0.569	0.597	0.55	0.67
0.04	25	0.623	0.627	0.576	0.632	0.65	0.608	0.538	0.572	0.521	0.648
0.02	50	0.535	0.553	0.505	0.577	0.589	0.548	0.447	0.502	0.438	0.585
0.01	100	0.447	0.484	0.442	0.53	0.535	0.494	0.365	0.44	0.363	0.528
0.0067	150	0.396	0.445	0.408	0.504	0.506	0.466	0.32	0.407	0.323	0.497
0.005	200	0.359	0.419	0.385	0.487	0.486	0.447	0.289	0.385	0.296	0.475
0.004	250	0.331	0.399	0.367	0.474	0.471	0.432	0.265	0.368	0.276	0.459
0.002	500	0.242	0.338	0.316	0.437	0.427	0.39	0.196	0.319	0.215	0.411

Fuente: Elaboración Propia

En la Tabla 4.25, se muestran los cuantiles para los períodos de retorno desde 1 hasta 500 años (19 períodos), los cuales son empleados para general las curvas de crecimientos en cada región. Éstos valores están expresado en fracciones de la PMAR, por ejemplo,

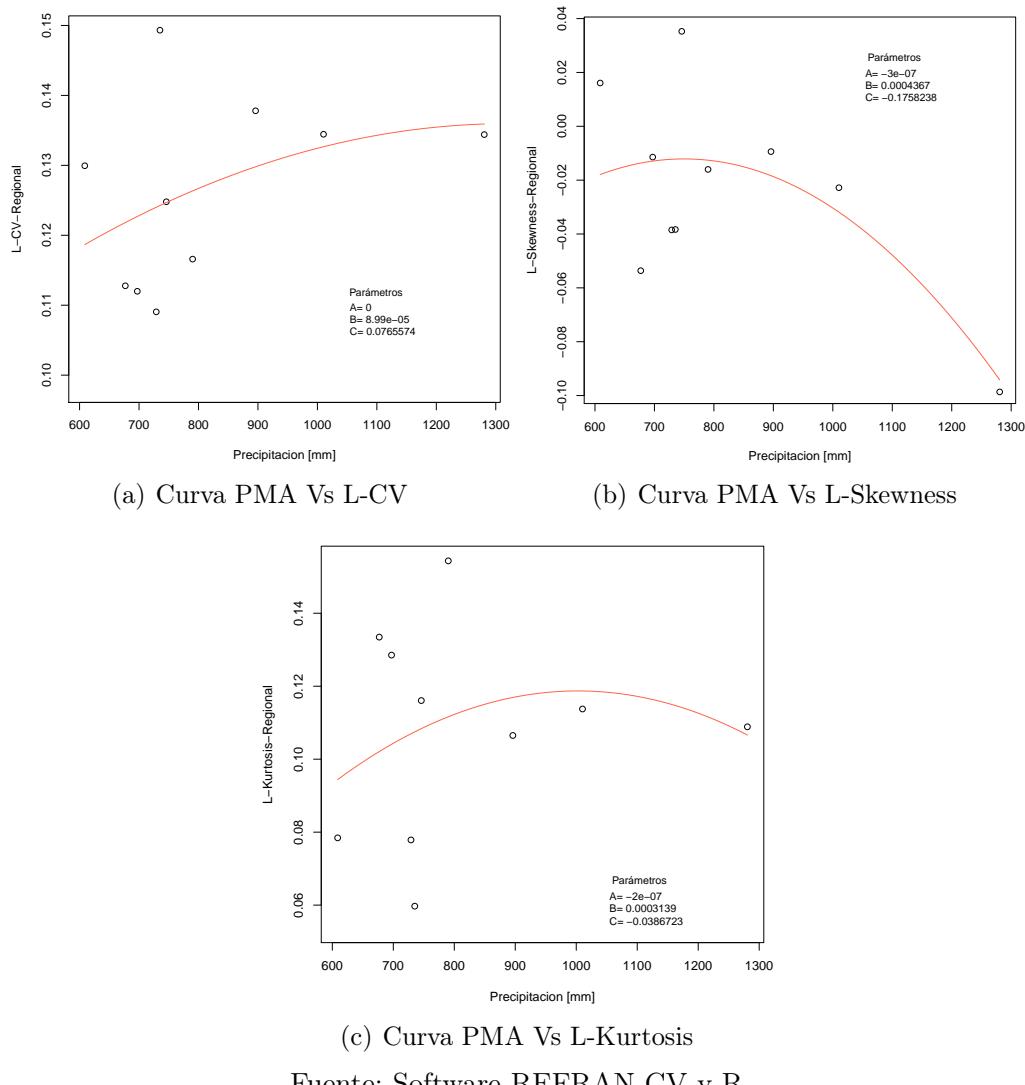
para el periodo de retorno igual a 100 años, en la región 1, se tiene 0.447 veces de la PMAR, es decir, una precipitación anual de 44.7 % de la PMAR.

El déficit de las precipitaciones, inicia a partir del periodo de retorno de 3 años, y para un 40 % de la PMAR conocido también como sequía del 40%, es decir, déficit del 60 % y recomendado para el análisis de sequías (Nuñez & Verbist, 2010), los periodos de retorno son: región 1, 3 y 8 $T_R = 150$ años, región 2 $T_R = 250$ años, región 4, 5, 6, y 10 T_R mayores a 500 años, región 7 y 9 T_R entre 75 años.

4.3.4. Mapeo de precipitaciones para períodos de retorno dados

Previo a realizar el mapeo de las precipitaciones, se construyen las curvas regionales que asocian los L-Momentos ratios principales y la variable auxiliar (considerada en este estudio la PMAR) obtenida en cada región.

Figura 4.37: Curvas de mejor ajuste de Precipitación media anual regional versus L-CV, L-Skewness, L-Kurtosis



Fuente: Software REFRAN-CV y R

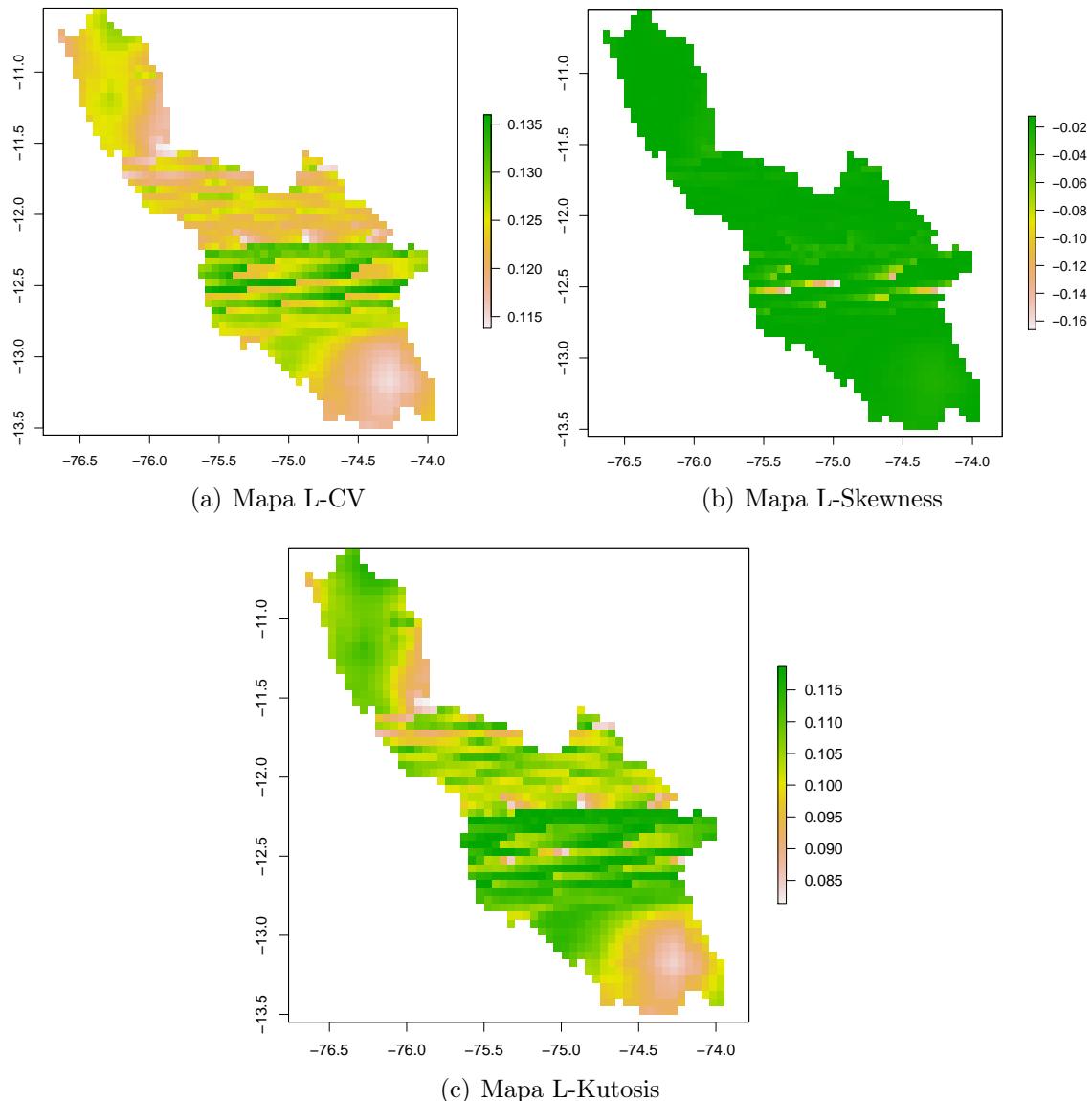
Haciendo uso del software REFRAN-CV y rutinas personalizadas en R, se calcularon los parámetros A , B y C de la función cuadrática $Ax^2 + Bx + C$ recomendada por Wallis et al. (2007). Para cada L-Momento ratio (L-CV, L-Skewness y L-Kurtosis) se han optimizado los parámetros, que se muestran en la figura 4.37 y en la Tabla 4.26.

Tabla 4.26: parámetros de ajuste curvas regionales

L-Momento Regional	Parámetro 1 (A)	Parámetro 2 (B)	Parámetro 3 (C)
$L - CV$	-3.40E-08	8.99E-05	0.076557415
$L - Skewness$	-2.91E-07	0.00043667	-0.175823842
$L - Kurtosis$	-1.57E-07	0.000313877	-0.038672302

Fuente: Elaboración Propia

Figura 4.38: Mapas L-Momentos regionales L-CV, L-Skewness y L-Kurtosis cuenca Mantaro

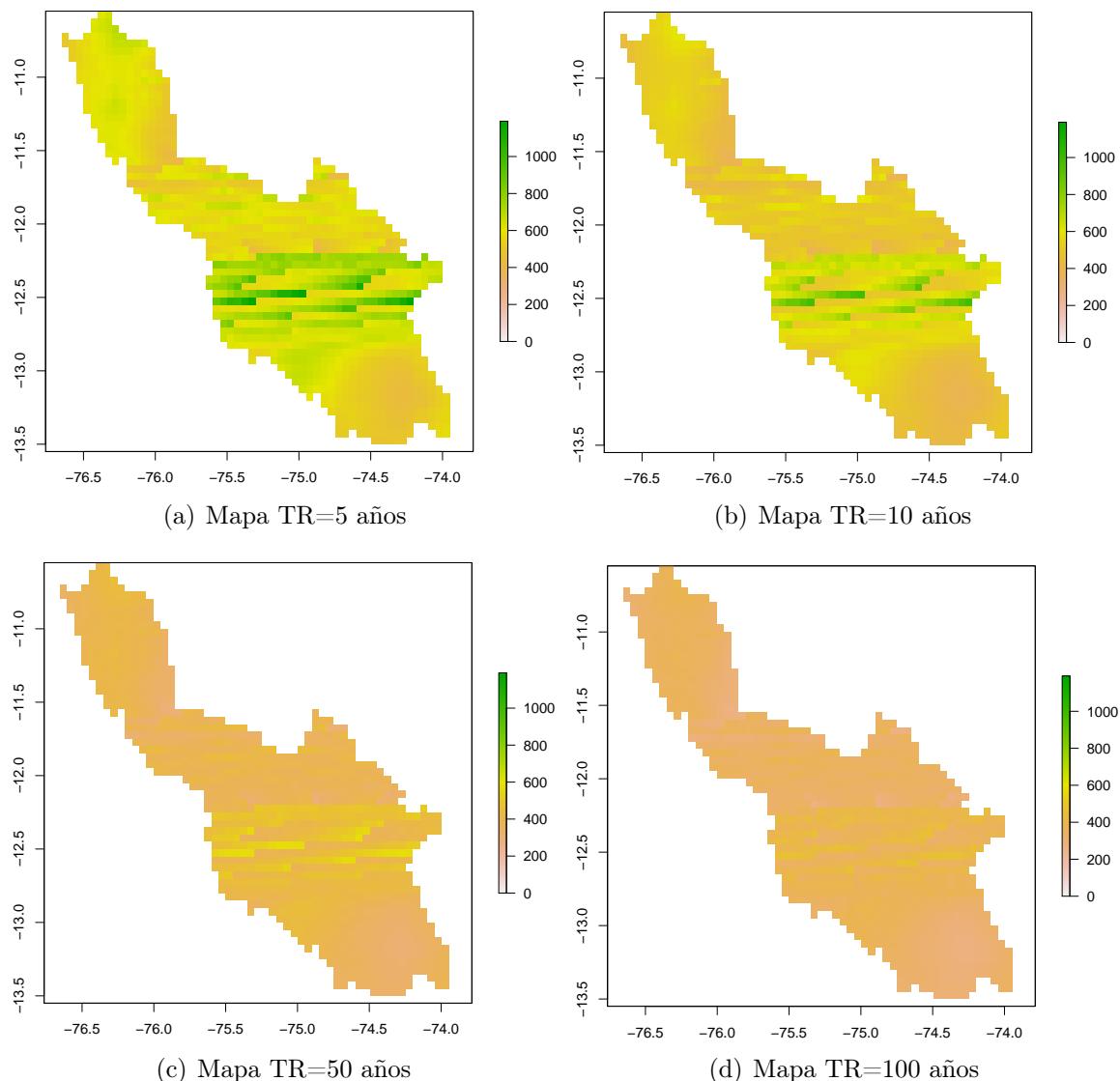


Fuente: Software REFRAN-CV y R

REFRAN-CV, tiene la opción de importar un mapa de extensión *.tif* con PMA distribuidos en celdas o grillas tipo raster para calcular los L-Momentos ratios en cada una de ellas. El procedimiento es: se calculan los parámetros de las distribuciones de probabilidad mediante los L-Momentos ratios que asocia la PMA de cada celda y la ecuación cuadrática calculada anteriormente. En cada una de las celdas se calculan las precipitaciones para diferentes períodos de retorno.

La figura 4.38 muestra las variaciones regionales de L-CV, L-Skewness y L-Kurtosis, los mismos que fueron obtenidos desde las ecuaciones regionales y sus respectivas PMA de cada celda. En la figura descrita, se observa en la región sur, un menor coeficiente de variación (0.115-0.120) y la parte central, mayor valores (0.130-0.135). Los valores de la curtosis es similar al anterior, región sur entre 0.085 y 0.090 y región central entre 0.110 y 0.115. En el sesgo, la distribución es uniforme 0.02-0.06, salvo las grillas que se ubican a lo largo de la latitud 12.5° sur, que varían entre 0.14-0.16.

Figura 4.39: Mapas de períodos de retorno y PMA



Fuente: Software REFRAN-CV y R

Las figuras 4.39a b, c y d, muestran la distribución de las precipitaciones anuales para los periodos de retorno 5, 10, 50 y 100 años respectivamente. En la región sur considerada como región semiárida, para el T_R 5 años, se distribuyen entre 400-600 mm de PMA, para 10 años, 400-500 mm de PMA, para 50 años, 300-400 mm de PMA y 100 años entre 200 y 300 mm de PMA.

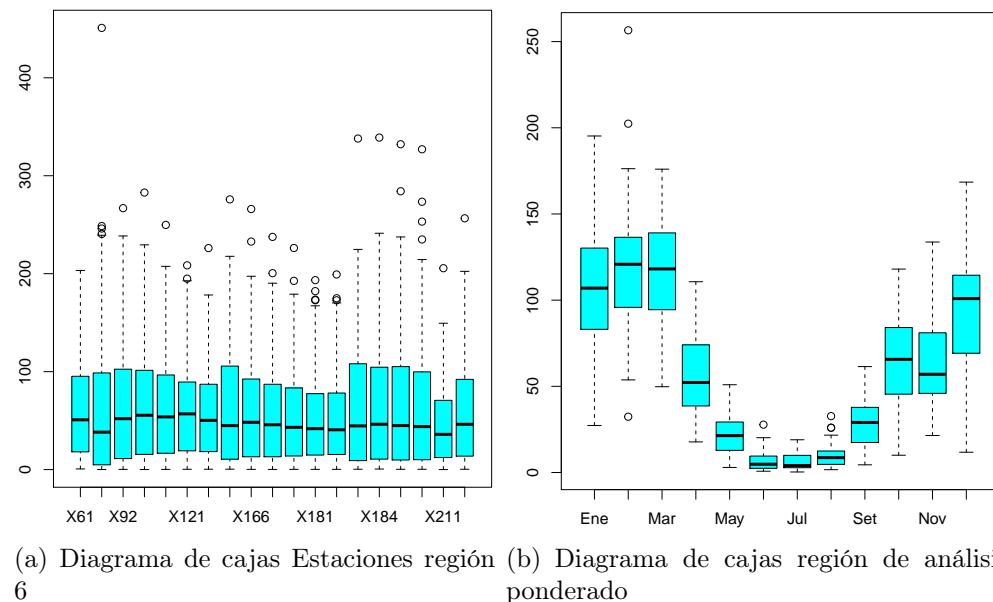
En general, en la cuenca del Mantaro para períodos de recurrencia de 5 años, se obtendrán precipitaciones anuales menores a 700 mm, salvo las regiones 3 y 7 y para períodos de recurrencia de 100 años, las precipitaciones tienen a distribuirse de manera uniforme con precipitaciones menores a 400 mm, salvo las regiones 3 y 7. La región más afectada corresponde a la región 6 zona sur de la cuenca estudiada.

4.4. Análisis estadístico de la precipitación en la región de análisis.

La zona de análisis, corresponde a la región 6, considerada como región semiárida, y se ubica en las planicies del departamento de Ayacucho, zona sur de la cuenca del Mantaro.

En la figura 4.40a, se muestra el diagrama de caja de cada una de las estaciones de la región de análisis, apreciándose cierta homogeneidad, que permitió el ponderado para el análisis de sequías. En la figura 4.40b, se muestra el diagrama de caja de la serie ponderada mensual, interpretándose que los meses de diciembre, enero, febrero y marzo son períodos húmedos; los meses abril, octubre y noviembre periodo semi-húmedos y el resto de los meses periodo secos, representando las estaciones típicas de la cuenca del Mantaro.

Figura 4.40: Diagrama de cajas región de análisis de sequías

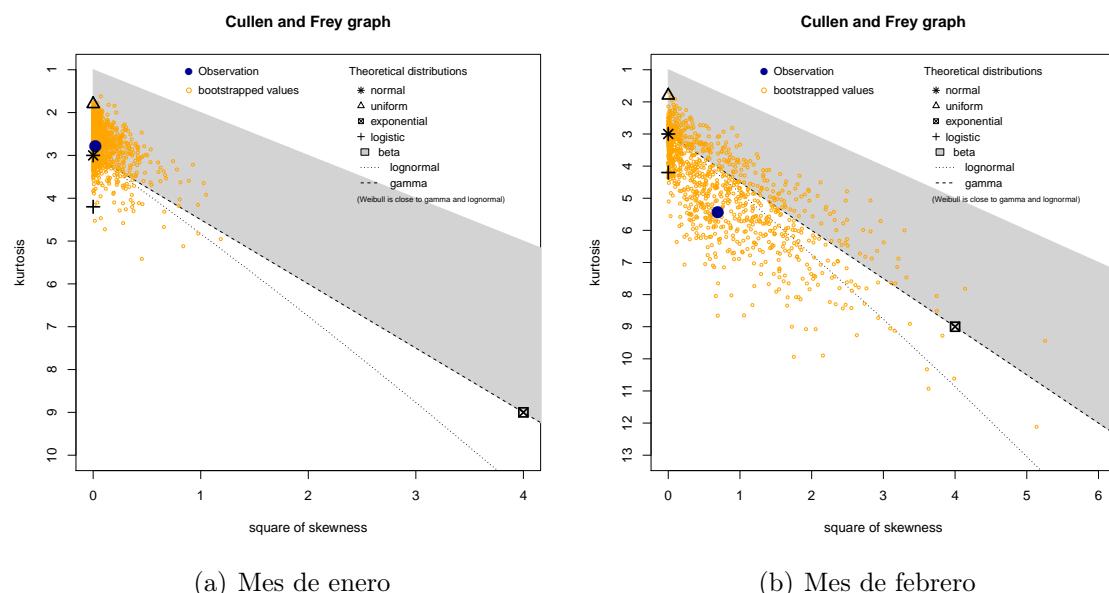


Fuente: Elaboración propia

Para calcular los SPIs, es necesario evaluar el ajuste de las distribuciones. Thomas B. McKee et al. (1993), recomienda para obtener los SPIs usar la función de distribución Gamma. Por otro lado, también es posible obtener los SPIs a partir de la distribución Normal, siendo el $SPI = Z = (x - \bar{x})/\sigma$, equivalente a la estandarización de la serie. Por lo tanto, se realiza la evaluación estadística de la serie mensual como corresponde los cálculo de los SPIs.

Existe una evaluación en primera instancia, a través del gráfico de Cullen y Frey, el cual consiste en la evaluación de 8 distribuciones de probabilidad (normal, uniforme, exponencial, logística, beta, lognormal, gamma y Weibull). La figura 4.41, muestra los gráficos de Cullen y Frey para los meses enero y febrero (en la sección Anexos A, se muestran los gráficos de Cullen y Frey y las distribuciones Empíricas y Teóricas para cada mes) y se aprecia que la serie del mes de enero se inclina hacia las distribuciones Nornal y Gamma, mientras que la serie del mes de febrero se inclina a la distribuciones Lognormal y Gamma. Este procedimiento contribuye en primera instancia a la selección de las distribuciones de probabilidad.

Figura 4.41: Gráfico de Cullen y Frey enero-febrero región de análisis



Fuente: Elaboración propia

Tabla 4.27: Estadísticos para las distribuciones Nornal y Gamma: Región de análisis

Estadísticos	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
$s - k(\text{Normal})$	0.07	0.13	0.10	0.12	0.06	0.17	0.21	0.18	0.11	0.11	0.17	0.12
$x - c(\text{Normal})$	1.68	6.09	4.08	3.99	1.53	12.95	20.67	14.22	5.73	13.65	10.03	7.80
$a - d(\text{Normal})$	0.13	0.48	0.35	0.45	0.19	1.74	1.46	1.30	0.40	0.31	0.84	0.41
$s - k(\text{Gamma})$	0.08	0.11	0.10	0.11	0.12	0.10	0.12	0.10	0.11	0.12	0.12	0.16
$x - c(\text{Gamma})$	3.20	7.39	5.67	2.95	3.09	2.29	8.05	6.17	3.95	18.58	8.21	13.56
$a - d(\text{Gamma})$	0.30	0.45	0.72	0.30	0.41	0.32	0.46	0.25	0.38	0.74	0.39	1.27

Fuente: Elaboración Propia

En la Tabla 4.27, se muestran 3 estadísticos (p-value), Smirnov Kolmogorov ($s - k$),

Predicción de sequías con Redes Neuronales Artificiales y Algoritmos Genéticos utilizando precipitación por percepción remota.



Chi cuadrado ($x - c$) y Anderson Darling ($a - d$), considerados como los test de bondad de ajuste, que permitieron evaluar las distribuciones Normal y Gamma. El estadístico $s - k$, asocia a la distribución Gamma a los meses enero, marzo, mayo, setiembre, octubre y diciembre, y en la distribución Normal asocia a los meses febrero, abril, junio, julio, agosto y noviembre. Es estadístico $x - c$, la función Gamma: enero, febrero, marzo, mayo, octubre, diciembre y el resto de los meses asocia a la distribución Normal. Por último, el estadístico $a - d$, la función Gamma: enero, marzo, mayo, octubre, noviembre y diciembre. Los resultados indican que la mitad de los meses se asocian a cada distribución (Gamma y Normal). Por lo tanto, para llevar a cabo los cálculos de los SPIs, se uniformiza a la distribución propuesta por Thomas B. McKee et al. (1993), es decir, la distribución **Gamma**.

4.5. Resultado de la escala e índices de sequía

En esta sección, se evaluó la escala temporal correspondiente a la respuesta de la vegetación dentro de la cuenca del Mantaro y el indicador (índice) cuantitativo adecuado para caracterizar las sequías dentro de la cuenca de estudio.

Respecto a la escala temporal, se tiene a la OMM et al. (2012, p. 7), y menciona que "...Las condiciones meteorológicas y de humedad del suelo (agricultura) responden a anomalías de precipitación en escalas temporales relativamente cortas, por ejemplo de entre 1 y 6 meses, mientras que los caudales fluviales, el almacenamiento en reservorios y las aguas subterráneas responden a anomalías de precipitación a más largo plazo, del orden de 6 meses y hasta 24 meses o más. Así, por ejemplo, habría que estudiar el SPI de 1 ó 2 meses para la sequía meteorológica, de entre 1 y 6 meses para la sequía agrícola, y de entre unos 6 y 24 meses o más para los análisis y aplicaciones de sequía hidrológica".

Para identificar el índice adecuado, Vicente Serrano et al. (2012, p. 687), realizaron un estudio para evaluar diferentes índices de sequía para la cuantificación de caudales, humedad del suelo, y crecimiento vegetativo. Los índices fueron el SPI, SPEI y PDSI y mencionan: "que la correlación más alta se produce entre 70-90 % de los casos con el SPI y SPEI en función de la variable y la estación del año, mientras que los índices PDSI representan menores del 15 %". También manifiestan que existen diferencias pequeñas entre el SPI y el SPEI. Es así, que en la presente investigación se realizaron las comparación de los índices SPI con metodologías numéricas e inteligencia artificial (algoritmos genéticos), SPEI de Vicente Serrano y los índices VCI como patrones de comparación.

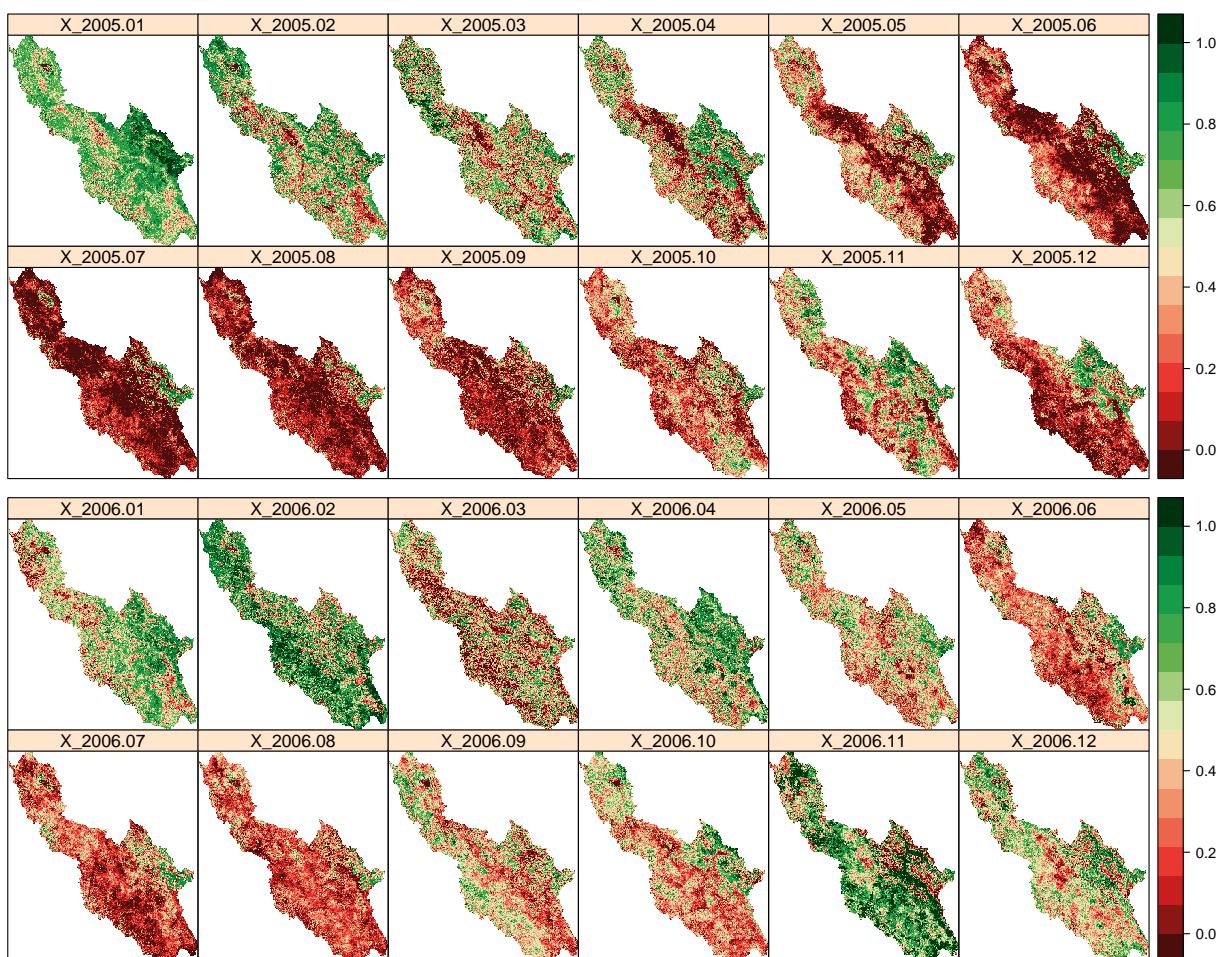
4.5.1. Índices de Vegetación

El índice VCI, deriva de los valores del NDVI y son distribuidos en el producto MODIS 13A3 del satélite terra. Están a disposición desde febrero de 2000 hasta la actualidad. Los datos corresponden a series mensuales con una resolución de $1km^2$. Debido a longitud corta de datos, este índice se toma sólo como patrón para identificar la mejor

correlación entre los índices descritos anteriormente. Por lo tanto, el índice seleccionando, es elegido para caracterizar las sequías en función de la respuesta de la vegetación en la zona de estudio, es decir, analizar las sequías agrícolas.

La figura 4.42, representan los mapas raster del VCI para los años 2005 y 2006 (periodos críticos de sequía), donde se observan regiones extensas con sequías severas y extremas en la escala temporal mensual. En la Sección Anexos B, se muestran el código fuente R y los mapas raster del índice VCI para una resolución de 1km^2 , desde marzo de 2000 hasta diciembre de 2015.

Figura 4.42: Índice de Condición de Vegetación periodos 2005 y 2006

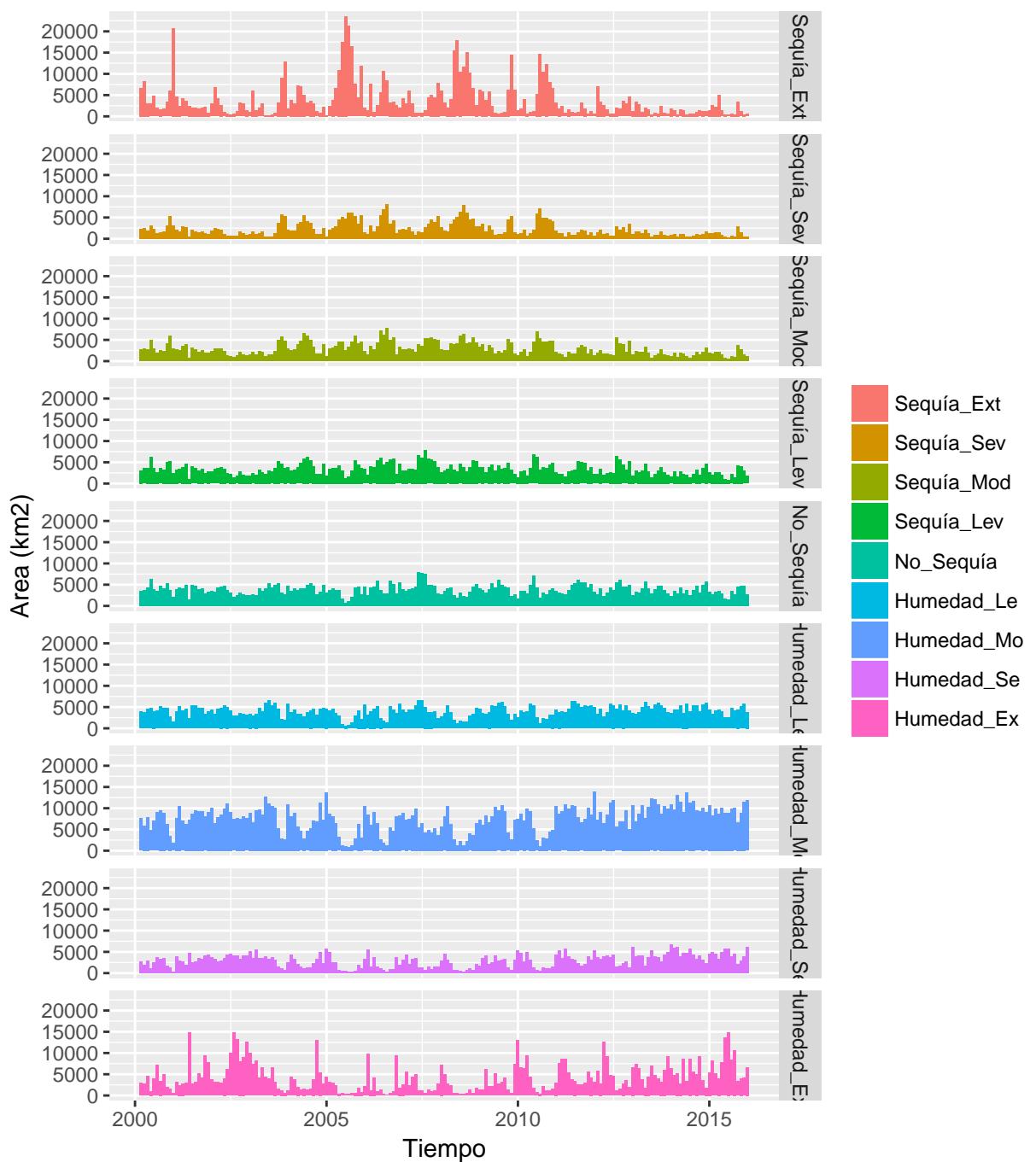


Fuente: Elaboración propia

La figura 4.43, permite identificar de manera eficiente la clasificación de las sequías y sus correspondientes áreas dentro del periodo de análisis. Tomando en cuenta las áreas que han sido afectadas dentro de la cuenca del Mantaro y de acuerdo a la clasificación de las sequías del índice VCI (Tabla 2.5).

Dentro del periodo 2000-2015 se observó sequías extremas intermitentes, siendo el periodo crítico entre el año 2005 y 2006, cuyas áreas involucradas han sido mayores a 20000 km^2 que representa mayor a 50 % del área de la cuenca.

Figura 4.43: Identificación de la sequía agrícola mediante el índice VCI cuenca del Mantaro



Fuente: Elaboración propia

Las sequías severas han afectado hasta 7500 km^2 aproximadamente y la mayor recurrencia han sido entre los años 2005 y 2010, con tendencia a disminución en los últimos años. Las sequías moderadas tienen el mismo comportamiento de las sequías severas.

Las sequías leves, han sido casi uniformes a lo largo del periodo de análisis.



Respecto a las humedades, existe un crecimiento entre los años 2010 y 2015, distribuyéndose hasta 15000 km^2 como máximo a lo largo de la cuenca. El promedio más elevado, respecto a las áreas de la cuenca, corresponden a la Humedad Moderada, con un promedio de 5000 km^2 aproximadamente.

La inexistencia de la sequías ni humedades extremas (clasificador No sequías), es altamente uniforme pero distribuido en áreas pequeñas, menores a 5000 km^2 .

4.5.2. Índices de precipitación estandarizada

Según la metodología propuesta, se estimaron los SPIs mediante dos metodologías, el primero que corresponde a la técnica numérica denominada Máxima Verosimilitud (SPIMV) y el segundo mediante la técnica de los Algoritmos Genéticos (SPIAG). Este último consistió en optimizar los parámetros α y β de la distribución Gamma.

Para determinar los SPIMV y SPIAG, se han utilizado el paquete SPIGA, el mismo que está escrito en lenguaje R y depositado en el repositorio CRAN. El código fuente resumidos está adjunto en las sección Anexos D. El análisis es mensual para diferentes escalas consideradas 1, 3, 6, 9 y 12 meses respectivamente.

Las figuras 4.45, muestran el desarrollo del desempeño del Algoritmo Genético para una población igual a 1000 y número de generaciones igual a 500 para los meses *enero-diciembre*. En todos los casos, se observa que el proceso llega a estabilizarse a partir de la generación 20, lo que indica, la obtención de buenos resultados (parámetros óptimos) a partir de éste ciclo.

El esfuerzo computacional, es mayor y altamente considerable para poblaciones mayores a 1000 y los resultados de los parámetros, muestran ser muy cercanos o iguales para poblaciones mayores a 100. Aproximadamente, el tiempo de ejecución para una población de 1000 y 100 generaciones, conlleva a 15 minutos de proceso para tres estaciones, equivalente a cinco minutos para una estación y característica de cómputo, procesador Core(TM) i7 2.4GHz y memoria RAM 12.0 GB.

La tabla 4.28, muestra los resultados de los parámetros estadísticos y el coeficiente de eficiencia (*CE*) para los SPIMV y SPIAG de escala temporal 6 meses. Se observan un promedio mensual de *CE* igual a 0.983 y 0.969 para los SPIAG y SPIMV respectivamente. Siendo el SPIAG ligeramente superior al SPIMV.

Tabla 4.28: Valores de parámetros estadísticos función Gamma por AG y MV, escala 6 meses

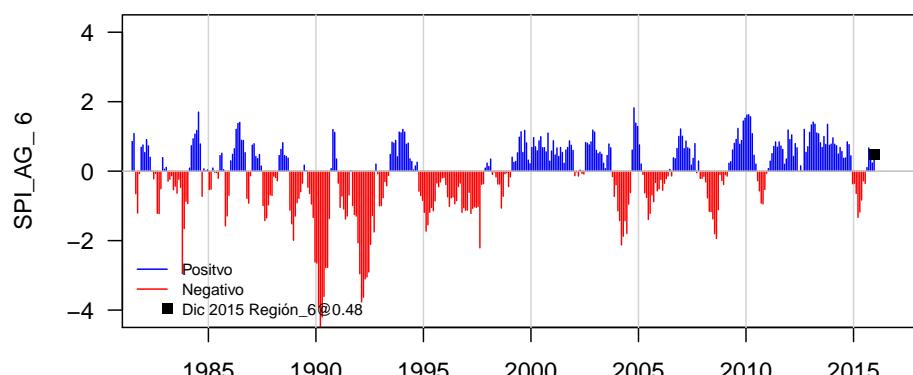
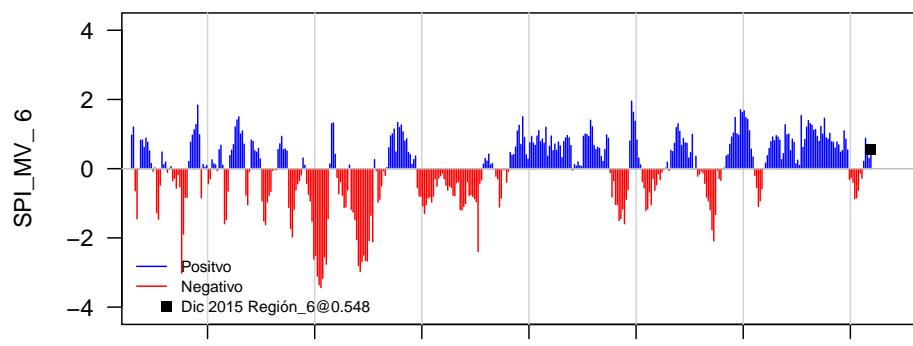
Parám.	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
α_{AG}	15.84	26.68	33.78	24.91	22.36	23.74	21.43	14.11	7.08	9.14	8.02	11.31
α_{MV}	14.98	17.93	20.74	18	18.52	21.81	22.37	17.61	10.59	9.81	10.64	11.67
β_{AG}	23.53	18.12	17.07	22.84	23.49	18.59	15.61	15.87	19.23	15.56	22.97	23.87
β_{MV}	24.35	26.28	26.86	30.53	27.52	19.66	14.67	12.37	12.47	14.22	16.98	22.71
CE_{AG}	0.995	0.988	0.988	0.979	0.969	0.968	0.990	0.997	0.981	0.971	0.982	0.992
CE_{MV}	0.986	0.962	0.972	0.959	0.946	0.950	0.986	0.995	0.963	0.955	0.969	0.989

Fuente: Elaboración Propia

De la metodología propuesta, con los parámetros óptimos conseguidos, se obtienen los SPIAG para distintas escalas (adjunto en la sección Anexos D), que posterior permitieron analizar mediante correlaciones y diagramas de Taylor la selección del mejor índices de sequías y escala temporal adecuado según la respuesta de la vegetación en la región de estudio.

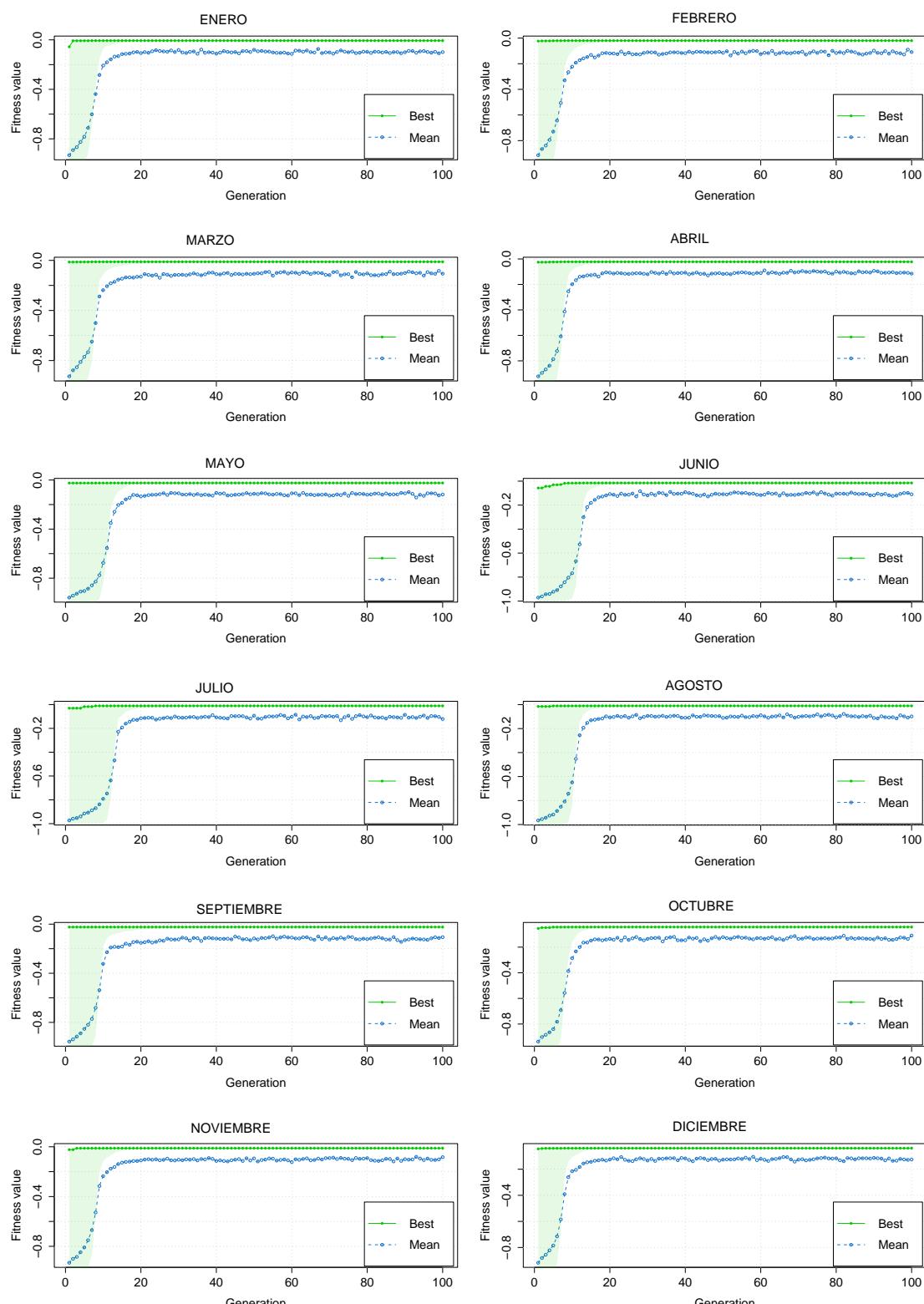
Las figuras 4.44a y b representan los SPIMV y SPIAG de escala 6 meses respectivamente, y se observa similitud entre sí. Dentro del periodo de análisis (1981-2015), se observan eventos de sequías en los años 1989-1993 (extrema), 1995-1998, 2004, 2005, 2008, 2010 e inicios del 2015.

Figura 4.44: Índices de precipitación estandarizada por Máxima Verosimilitud y Algoritmos Genéticos



Fuente: Elaboración propia

Figura 4.45: Performance Algoritmos Genéticos Población = 1000 y generación = 100



Fuente: Elaboración propia

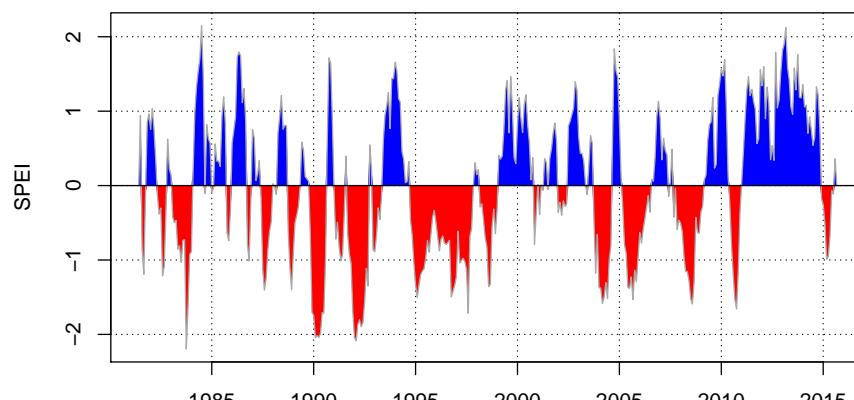
4.5.3. Índices de precipitación evapotranspiración estandarizada

Se ha empleado el paquete SPEI de Santiago Beguería y Vicente Serrano, alojado en el repositorio <https://cran.r-project.org/>. Al igual que los SPIs, se calcularon para diferentes escalas temporales. La Evapotranspiración Potencial (PET), se ha calculado mediante el método de Hargreaves, siendo los parámetros de ingreso, las temperaturas mínimas y máximas mensuales, y al igual que las precipitaciones mensuales fueron proporcionados por PISCO-SENAMHI en formatos .tif de resolución 5km^2 .

Las figuras 4.46, demuestra que los SPEIs son más conservadora que los SPIs, sin embargo, indican las mismas temporadas o eventos de sequías en el periodo de análisis. Los valores para distintas escalas temporales, también se adjuntan en la sección Anexos D.

Complementando al análisis anterior, en la figura 4.46, se aprecian períodos húmedos de mayor duración desde 2011 hasta finales de 2014 y presencia de periodos húmedos de duración intermitente.

Figura 4.46: Índices de precipitación evapotranspiración Escala 6 meses



Fuente: Elaboración propia

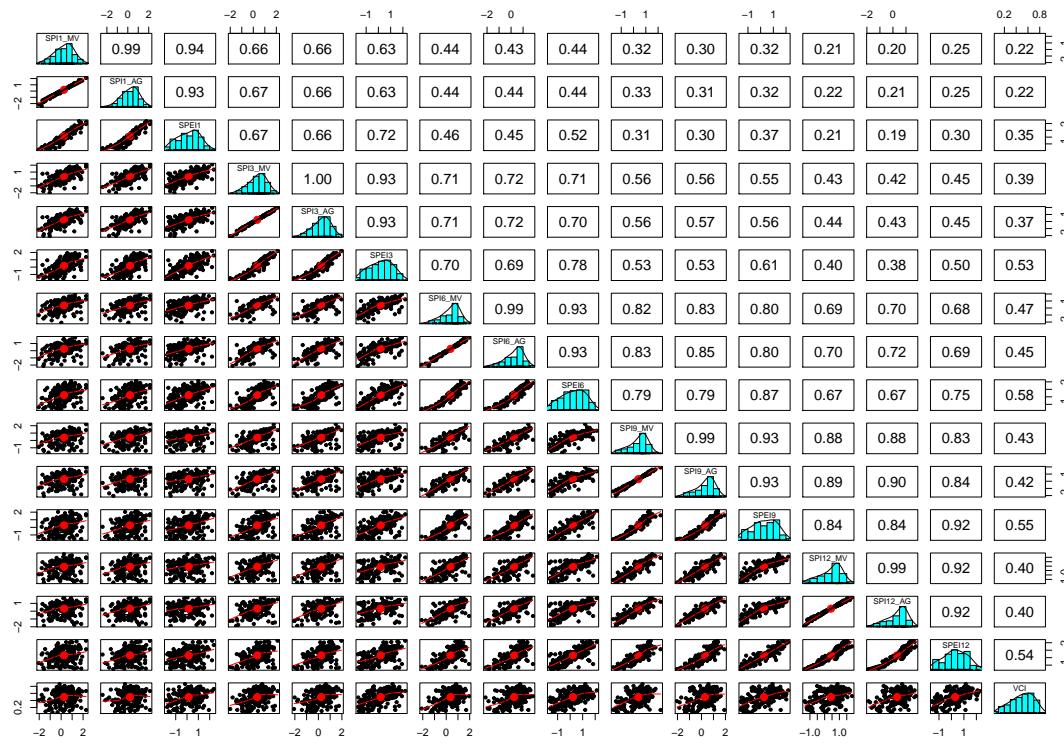
4.5.4. Correlación y estimación del mejor índices de sequía

Descrita anteriormente la metodología, se procede al análisis de correlación para y validez de los índices de sequías a través de los estadísticos correlación de Pearson y diagrama de Taylor.

Las variables que se correlacionan, corresponden a los índices SPIMV, SPIAG y SPEI, con sus respectivas escalas temporales de 1, 3, 6, 9 y 12 meses, haciendo un total de 15 índices.

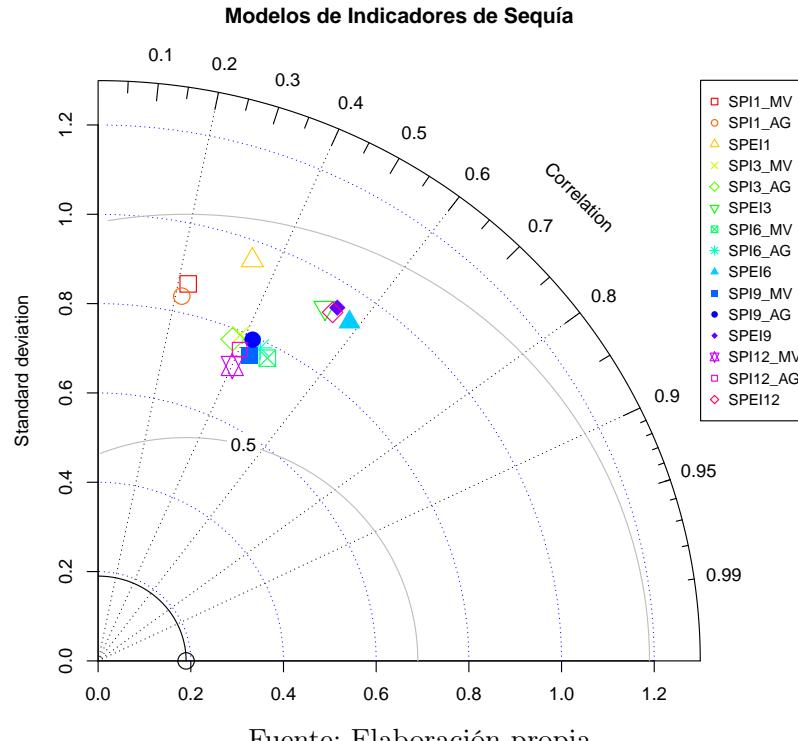
En la figura 4.47, se aprecian las correlaciones entre cada índice y el índice patrón (VCI, ubicado en la última columna). Los índices SPEI de 3, 6, 9, y 12 meses, correlacionan mejor (0.53, 0.58, 0.55 y 0.54 respectivamente) frente a los demás índices. La escala temporal de 6 meses presenta mayor correlación (0.58).

Figura 4.47: Correlación entre los índices de sequía SPIMV, SPIAG, SPEI y VCI para diferentes escalas



Fuente: Elaboración propia

Figura 4.48: Diagrama de Taylor SPIMV, SPIAG y VCI para diferentes escalas



Fuente: Elaboración propia



EL diagrama de Taylor, es una verificación rigurosa al proceso de correlación descrita anteriormente. La figura 4.48, muestra 15 variables (índices), para validar cada uno de ellos mediante los estadísticos, correlación, desviación estándar y el MSE. Se aprecia que el SPEI6 tiene mejor correlación frente a los demás índices. Según el gráfico de Taylor, los SPIs obtenidos mediante Máxima Verosimilitud y Algoritmos Genéticos, son similares (presentan correlaciones y errores cuadráticos medios relativamente bajos frente al índice VCI), por otro lado, los SPEIs 3, 6, 9 y 12 correlacionan muy bien, pero con errores cuadráticos medios relativamente superiores.

Por lo tanto, el índice adecuado y el que mejor responde al índice patrón es el **SPEI de escala 6 meses**. Este resultados corrobora a la descripción de la OMM et al. (2012, p. 7).

4.6. Predicción de sequías en la región de análisis.

El modelo para la predicción de sequías es mediante las Redes Neuronales Dinámicas NARX, que tiene la ventaja de asociar variables o entradas externas, contemplando los datos históricos para la generación de series de tiempo. Se han considerado como las variables externas a los ICG, los cuales han sido clasificados, de acuerdo a criterios estadísticos.

Por lo tanto, las variables de entrada corresponden a los ICG clasificados, y las salidas al índice SPEI de la región de análisis. Durante la construcción de modelo, se toma el 80 % de los datos (entradas y salidas) para el entrenamiento y el 20 % restante para la prueba o validez del modelo.

Este procedimiento se ha realizado mediante el programa Matlab, el mismo que presenta mayores ventajas frente a otros paquetes informáticos de RNA, por ejemplo en los tipo de redes considera 18 tipos, función de entrenamiento, considera 14 funciones, incluido Levenberg-Marquardt, adaptación de la función de aprendizaje (02), función de performance (03), funciones de transferencia (03), entre otros que lo clasifica entre los más robustos y completos para cumplir satisfactoriamente el modelado mediante las RNA.

Internamente, el programa divide el conjunto de datos en 70 %, 15 % y 15 % para el entrenamiento, validación y prueba respectivamente. El proceso de entrenamiento, corresponde al conjunto de ejemplos que se emplea para el proceso de aprendizaje, siendo el objetivo encontrar los pesos adecuados para la simulación. El proceso de validación, es el conjunto de ejemplos que se utiliza para ajustar los parámetros de un clasificador. También sirve para encontrar el número óptimo de unidades ocultas. Por último, el proceso de prueba o test, que corresponde al conjunto de ejemplos que se usa solo para evaluar el rendimiento del modelo entrenado. Solo es usado una vez finalizado el entrenamiento.

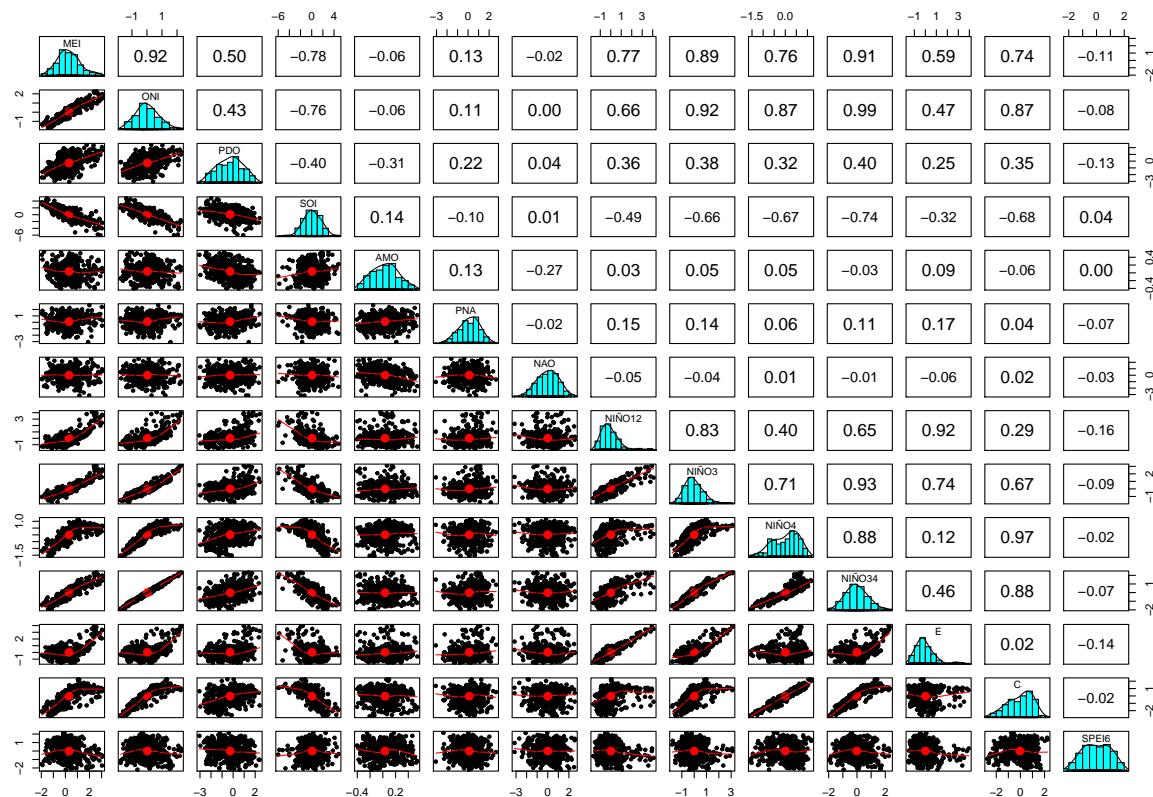
4.6.1. Selección de datos de ingreso

Para aplicar el modelo NARX, es necesario la presencia de variables externas consideradas como los datos de ingreso, los cuales en la investigación se han asignado a los ICG, sin embargo, no todos contribuyen en este proceso.

La selección de estas variables, se realiza a través de proceso de correlación entre sí y los SPEIs de la región de análisis, considerando menor correlación posible entre cada ICG y mayor correlación entre los SPEIs, el primero para evitar redundancia en el modelo y el segundo obtener mayor significancia entre los datos de ingreso y salida.

Inicialmente se consideran 18 ICG (MEI, ONI, PDO, SOI, AMO, PNA, NAO, NIÑO 1+2, NIÑO 3, NIÑO 4, NIÑO 3.4, BEST, NATL, SATL, TROP, E y C), de los cuales, se descartan algunos índices por la inexistencia de datos en los períodos 1981-06 al 1981-12, de esta manera se evita la reducción del registro de datos para el entrenamiento del modelo.

Figura 4.49: Correlaciones cruzadas, variables climáticas globales mes de diciembre y SPIs mes de enero



Fuente: Elaboración propia

En la figura 4.49, se observa que la correlación entre los ICG y el SPEI no son significativas, debido a la variabilidad estacional del clima (estaciones lluviosas y estaciones secas). Como alternativa externas se puede realizar el análisis trimestral, el mismo que



consiste en la obtención de los promedios de los tres meses consecutivos de cada año entre las variables climáticas y los indicadores de sequía, de esta manera mejora la correlación.

De acuerdo a la descripción anterior y la figura 4.49, los ICG clasificados son el MEI, PDO y el NIÑO 1+2, los cuales presentan menores correlaciones entre ICG y mayor entre el SPEI.

4.6.2. Arquitectura Redes Neuronales Artificiales

Del análisis anterior, se consideran 3 neuronas de entrada denominados índices MEI, PDO, y NIÑO 1+2 respectivamente y 1 neurona de salida correspondiente a los SPEI de la región de análisis. El número de neuronas en las capas ocultas conocidas como *Hidden Layer*, se determinan mediante el procedimiento *ensayo-error* progresivo conforme lo recomiendan ASCE (2000a, 2000b), Guoqiang et al. (1998), Hagan et al. (1996), entre otros.

Para establecer la arquitectura del modelo, se han realizado 50 simulaciones iniciales determinando el desempeño de cada modelo (valores cercanos a cero son los adecuados). Las arquitecturas analizadas son: 3-1, 6-1, 12-1, 24-1, 10-5-1, 20-10-1, 5-3-2-1, 10-5-2-1, 15-10-5-1, 20-10-5-1 y 35-20-10-1, donde el número extremo es la neurona de salida (1) y los demás son las neuronas en las capas ocultas. Se han considerado hasta 3 capas ocultas.

Se implementó esta técnica para reducir el espacio de búsqueda del mejor modelo. En la tabla 4.29, se aprecia los desempeños para cada arquitectura y se concluye que las arquitecturas con número de neuronas grandes, el desempeño asciende, es decir, se obtienen modelos ineficientes, y viceversa para obtener modelos eficientes. También se observa mayor estabilidad y correlaciones durante el entrenamiento para arquitecturas mayores iguales a 2 capas ocultas.

Tabla 4.29: Performance del modelo para distintas arquitecturas

No	3-1	6-1	12-1	24-1	10-5-1	20-10-1	5-3-2-1	10-5-2-1	15-10-5-1	20-10-5-1	35-20-10-1
1	1.336	1.919	21.165	21.997	2.004	3.735	0.954	1.426	1.913	1.5	2.266
2	1.88	1.603	2.31	2.717	1.624	1.863	3.504	0.869	1.931	1.157	1.43
3	2.622	1.207	1.173	10.607	1.801	2.328	2.004	1.299	1.607	1.829	3.2
4	3.013	1.016	1.811	3.429	2.92	1.093	1.876	1.25	1.772	1.471	0.873
5	1.567	0.97	4.866	11.708	1.61	7.18	0.769	1.62	3.658	1.851	2.733
6	1.765	1.036	3.903	6.923	2.779	2.826	2.551	1.687	1.185	1.234	1.559
7	2.668	2.084	3.069	2.371	2.876	3.279	1.989	1.383	0.74	2.295	1.635
8	0.884	1.403	1.061	6.278	3.316	1.903	3.036	1.25	2.23	5.116	1.732
9	1.773	1.133	2.174	6.316	1.328	1.467	0.94	1.418	3.195	0.788	2.403
10	1.921	2.311	1.521	9.316	2.22	3.516	1.13	2.688	3.177	1.227	1.35
11	2.035	3.832	1.334	2.204	2.199	1.767	1.356	0.759	1.276	3.785	1.047
12	0.932	0.874	2.113	5.828	2.399	2.756	2.656	0.808	0.949	3.712	2.332
13	1.127	1.785	2.15	4.413	1.086	1.481	3.161	1.997	1.207	2.13	5.216
14	2.004	2.467	1.473	14.193	3.541	1.783	0.956	3.417	2.747	0.93	1.617
15	1.247	1.356	5.419	2.742	2.418	4.509	4.248	3.308	1.122	3.331	2.078
16	1.475	2.86	0.959	3.749	1.381	1.075	3.399	3.705	2.509	2.141	4.467
17	1.755	4.575	1.918	2.062	1.168	1.145	2.217	2.336	4.069	0.998	2.124
18	1.494	1.661	1.931	7.945	0.892	1.991	1.369	4.791	3.31	4.086	3.303
19	2.694	1.942	1.048	5.662	0.953	1.146	1.729	2.231	2.359	2.402	2.12
20	2.33	1.783	4.383	3.206	2.303	3.26	0.827	1.801	0.762	1.608	1.293
21	1.118	1.239	3.775	9.554	1.957	0.997	2.371	1.181	3.932	1.957	1.668
22	1.707	1.604	2.998	11.915	1.144	3.358	0.929	3.248	2.173	5.227	1.613
23	1.413	1.72	4.624	10.397	3.419	6.943	2.453	2.586	1.269	1.27	1.958
24	1.116	0.818	2.245	6.577	1.118	2.146	1.346	3.788	0.994	0.897	2.305
25	2.01	1.797	1.465	6.32	5.156	3.801	1.337	2.231	3.976	2.79	3.898
26	1.837	2.789	1.861	5.939	2.551	2.798	2.363	1.101	1.261	0.925	0.88
27	1.584	2.096	4.375	14.213	1.785	1.976	1.165	1.695	1.77	2.934	1.463
28	2.118	2	2.424	2.357	0.916	0.873	2.432	2.181	2.963	2.816	1.793
29	1.597	1.748	2.591	5.499	2.728	0.922	1.661	1.372	1.197	3.215	1.897
30	1.858	1.535	0.991	6.389	1.342	2.338	2.371	1.9	3.587	1.187	2.342
31	1.597	2.187	2.139	11.783	1.503	1.514	1.594	0.71	1.487	0.948	3.989

(continúa en la página siguiente)

Tabla 4.29: Continuación de la tabla

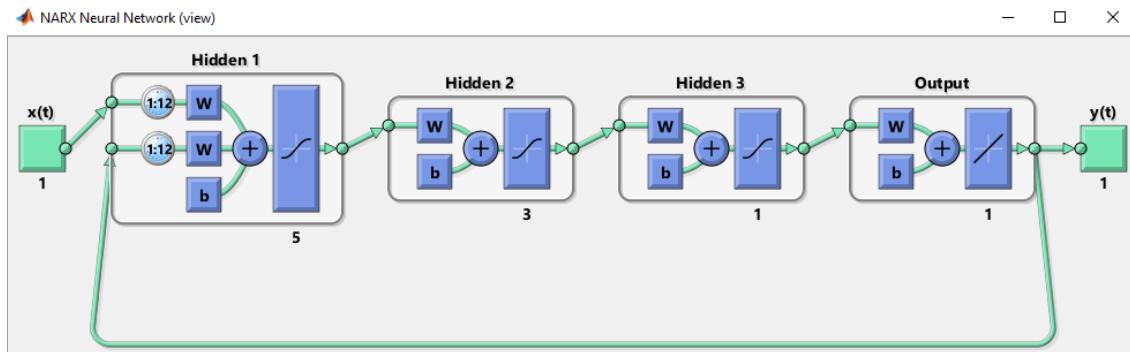
No	3-1	6-1	12-1	24-1	10-5-1	20-10-1	5-3-2-1	10-5-2-1	15-10-5-1	20-10-5-1	35-20-10-1
32	2.499	2.983	5.629	2.88	2.041	1.476	1.352	1.427	2.263	1.895	2.059
33	3.008	1.51	1.922	11.424	1.667	3.39	1.95	1.826	1.881	2.749	3.043
34	2.56	2.089	2.051	3.675	1.103	4.178	1.474	1.461	1.164	1.186	4.255
35	1.117	1.085	0.942	4.702	1.343	2.15	2.166	2.352	2.142	1.866	2.155
36	1.221	1.831	1.296	6.233	1.808	1.37	2.632	2.835	2.081	2.703	4.464
37	1.462	1.71	1.698	11.578	2.277	2.221	1.03	0.946	2.32	3.219	1.125
38	1.528	1.361	1.313	2.287	2.206	1.076	1.098	1.644	1.429	3.103	6.391
39	1.02	3.24	2.642	1.732	1.502	1.792	1.369	1.697	3.717	2.398	1.746
40	2.76	1.613	2.084	17.988	1.478	1.983	1.712	0.92	1.338	1.784	2.094
41	2.561	1.882	4.713	4.319	0.735	0.896	1.302	3.388	0.727	3.438	3.031
42	1.976	2.086	1.33	1.686	2.218	3.91	2.523	3.129	3	0.972	1.031
43	1.821	1.206	1.072	1.987	0.986	2.938	2.693	3.009	1	1.784	3.304
44	1.26	3.033	2.234	3.378	1.588	2.155	2.266	2.481	3.604	2.372	1.846
45	2.581	0.969	2.333	17.083	2.612	2.268	3.653	3.955	2.923	3.459	1.313
46	1.197	2.203	2.649	7.043	1.839	2.371	0.895	3.006	2.451	4.165	2.361
47	2.054	1.861	3.172	6.925	1.948	2.573	1.982	1.408	1.514	2.708	1.598
48	1.18	2.037	1.322	3.89	1.064	1.399	0.993	1.442	2.525	1.59	2.148
49	1.239	0.991	1.447	7.51	1.728	2.241	1.441	1.466	3.43	1.166	1.482
50	1.556	1.905	1.836	31.955	1.385	0.791	2.108	3.463	3.029	0.951	1.537
Prom	1.782	1.859	2.739	7.458	1.919	2.379	1.907	2.078	2.177	2.225	2.311
Desv	0.572	0.758	2.931	5.816	0.837	1.356	0.829	0.98	0.986	1.124	1.166

Fuente: Elaboración Propia

Conociendo el rango de las arquitecturas y los desempeños correspondientes, se realizaron las simulaciones con número de neuronas relativamente bajos (proporcional a la cantidad de neuronas de ingreso) para obtener el modelo adecuado.

La figura 4.50, muestra la arquitectura del modelo obtenida a través del procedimiento *ensayo-error* y sometida a 5000 ciclos de evaluación. De esta manera, la arquitectura del modelo es: 5-3-1-1, es decir, los tres primeros corresponden a las tres capas ocultas y la última neurona (1) a la capa salida.

Figura 4.50: Arquitectura Redes Neuronales Artificiales Modelo NARX



Fuente: Software Matlab

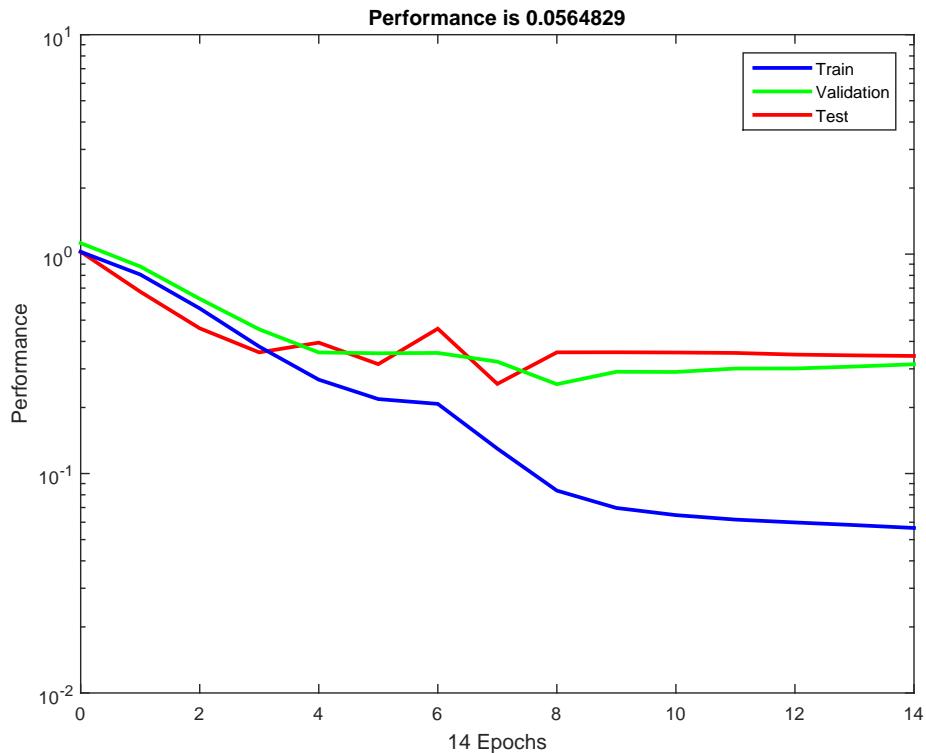
4.6.3. Resultados Redes Neuronales Artificiales

Los datos de entrada y salida, son series de tiempo con registros mensuales desde 1981 hasta 2015. Para la validez del modelo, el conjunto de datos (entradas y salidas), se han dividido en dos grupos. El primero que corresponden al 80 % (1981-2008) para la construcción del modelo y el segundo del 20 % restante (2009-2015) para la validez del modelo.

Los datos no requieren estandarizarse, ya que los valores de los ingresos y salidas de la presente investigación son estandarizadas por defecto, sin embargo, Matlab realiza este

procedimiento en el rango de -1 y 1, mediante el comando *mapminmax*. También realiza de manera aleatoria la división del conjunto de datos en tres grupos, entrenamiento (70 %), validación (15 %) y prueba (15 %), para evaluar el desempeño del modelo según la arquitectura propuesta.

Figura 4.51: Curvas de desempeño del modelo NARX, arquitectura 5-3-1-1



Fuente: Software Matlab

La figura 4.51, muestra el desempeño del modelo durante las tres etapas, entrenamiento, validación y prueba, y se observa que el desempeño es descendente hasta el ciclo 14, etapa que corresponde a finalizar el entrenamiento del modelo. También se observa, a partir del ciclo 8, el proceso de validación y prueba permanecen constantes, es un indicador de parada en el entrenamiento.

El entrenamiento se lleva a cabo hasta obtener los coeficientes de regresión adecuados. Por defecto, el entrenamiento, se detiene después de seis incrementos consecutivos en error de validación y el mejor rendimiento se toma de la época o ciclo con el error de validación mínimo. Este procedimiento se aplicó en ciclos grandes, hasta lograr la validación adecuada con los datos del segundo grupo.

La figura 4.52, muestra cuatro correlaciones y pertenecen: al entrenamiento (70 % del registro), validación (15 % del registro), prueba (15 % del registro) y el modelo total (100 % del registro). El modelo total presenta 0.91335 de correlación, indicándonos **significancia** del modelo propuesto.

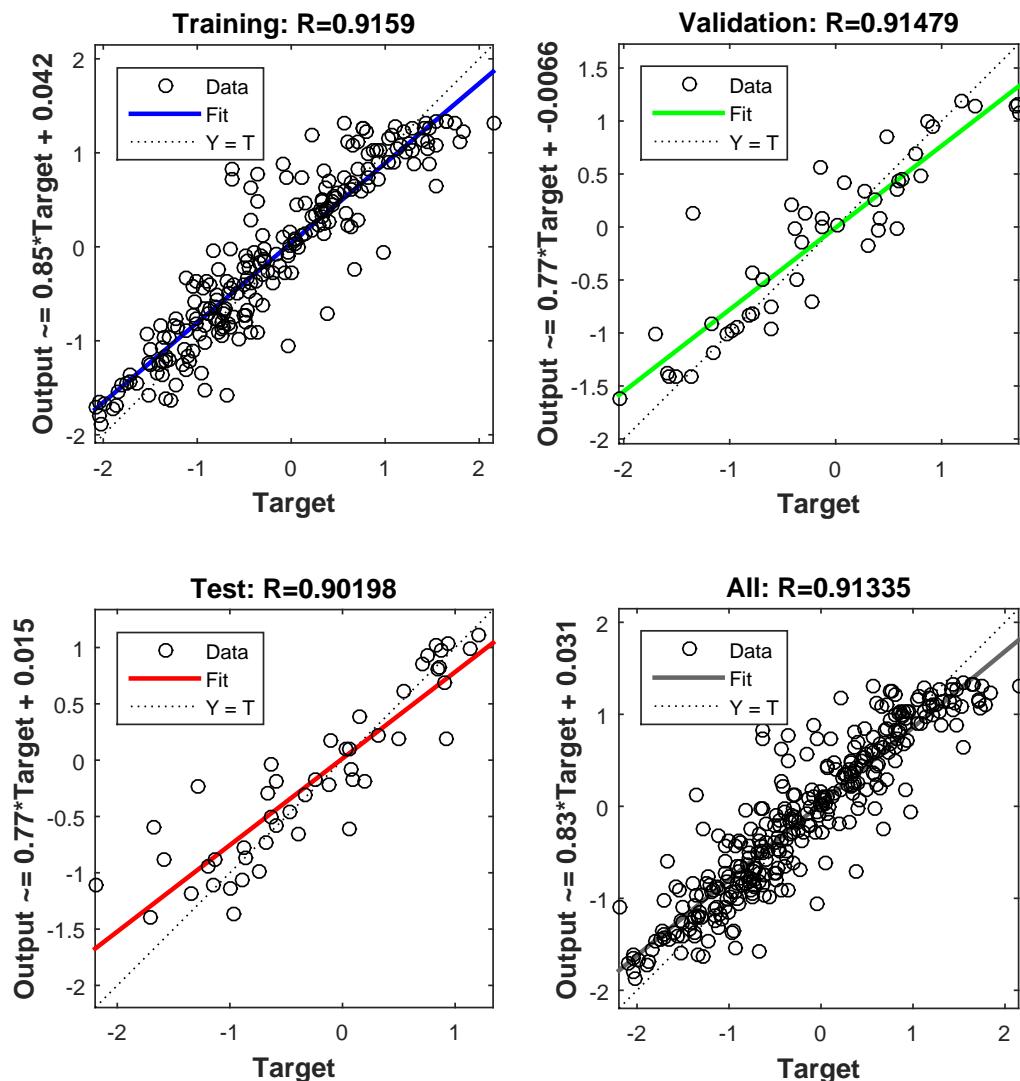
La figura 4.53, muestra la autocorrelación de los errores percibidos durante la etapa del entrenamiento. Este gráfico es un indicador de la eficiencia del modelo predictivo

en series de tiempo. Sólo se debe aceptar un valor alto de correlación en desplazamiento nulo para indicarnos el modelo de predicción perfecta. En gráfico, se observa este comportamiento, y los demás correlaciones están dentro del intervalo de confianza de 95 %. Por lo tanto, el presente modelo es adecuado para realizar las predicciones.

Dentro del primer grupo de la serie, Matlab realiza la simulación con los datos de salida y los resultados del modelo entrenado. La figura 4.54, muestra esta serie de tiempo, con sus respectivos errores respecto al tiempo.

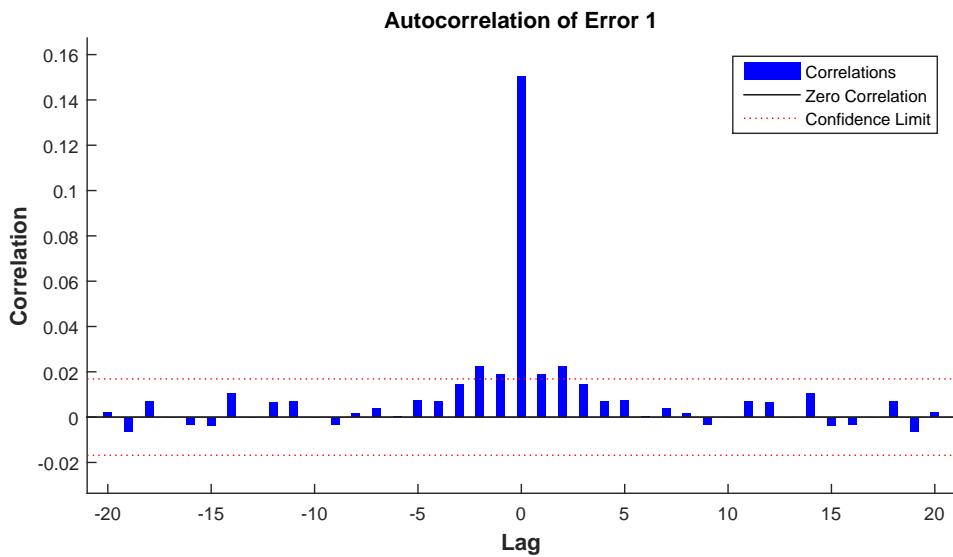
Finalmente, se construye la serie de tiempo del segundo grupo (2009-2015), para validar el modelo y visualizar la similitud o diferencia entre los resultados obtenidos mediante el modelo NARX. La figura 4.55, muestra la predicción de los SPEIs y los resultados esperados. El estadístico error cuadrático medio (MSE) indica 0.626, siendo este valor el mínimo durante 5000 ciclos y pruebas con diferentes arquitecturas.

Figura 4.52: Análisis de regresión para el entrenamiento, validación, prueba y toda la serie, arquitectura 5-3-1-1



Fuente: Software Matlab

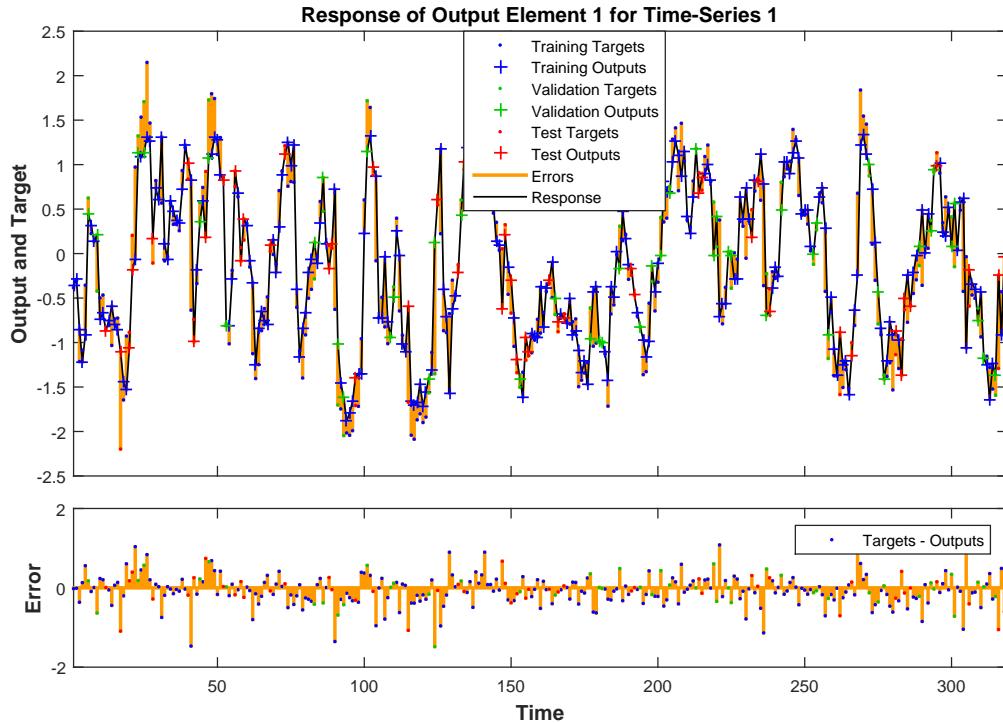
Figura 4.53: Autocorrelación de los errores desplazados, arquitectura 5-3-1-1



Fuente: Software Matlab

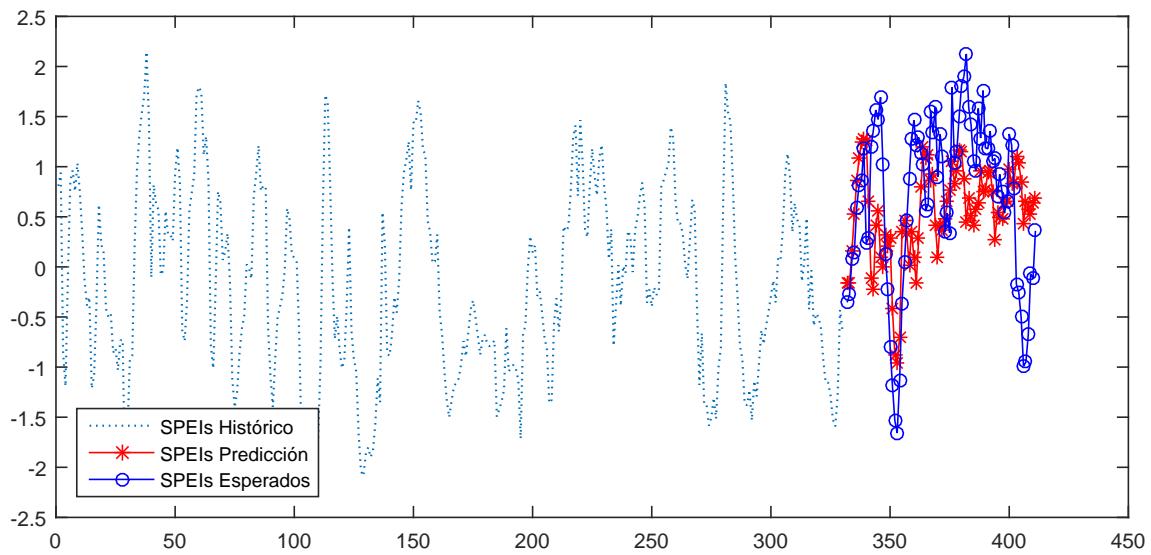
Por lo tanto, el modelo RNA-NARX de arquitectura 5-3-1-1 presenta un desempeño de MSE igual a 0.626 para la predicción de los SPEIs con las variables externas MEI, PDO y NIÑO1+2.

Figura 4.54: Series de tiempo entre los SPEIs y errores entre los valores entrenados, arquitectura 5-3-1-1



Fuente: Software Matlab

Figura 4.55: Serie de tiempo entre los valores simulados y esperados, arquitectura 5-3-1-1



Fuente: Elaboración propia

4.6.4. Prueba de hipótesis

Para la prueba de hipótesis se presentan dos grupos a comparar, los SPEIs calculados y los SPEIs simulados mediante el modelo de las RNAs. Previo a este procedimiento, la diferencia de estos dos grupos deben de admitir un cierto grado de normalidad para la eficacia estadística.

En el presente trabajo de investigación, la hipótesis plateada es:

H_0 : $\mu =$ El modelo *RNA-NARX* permite la predicción de la sequía agrícola, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro.

H_1 : $\mu \neq$ El modelo *RNA-NARX* no permite la predicción de la sequía agrícola, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro.

Código Fuente 4.1: Prueba de Hipótesis R

```

1 t.test(x,y,"two.sided",paired=TRUE)
2 Paired t-test
3 data: x and y
4 t = 1.645, df = 79, p-value = 0.1039
5 alternative hypothesis: true difference in means is not equal to 0
6 95 percent confidence interval: -0.03023939  0.31821439
7 sample estimates: mean of the differences  0.1439875

```

$p - value > 0.05$, por lo tanto, **se acepta** la hipótesis nula (H_0)

Conclusiones y recomendaciones

Conclusiones

- Aplicando la estadística inferencial, la medida de la evidencia (*p-value*) que aporta el resultado en el modelo RNA-NARX es de 0.104, demostrando ser mayor al nivel de confianza igual a 0.05. Por lo tanto, se acepta el modelo RNA-NARX para la predicción de la sequía agrícola, utilizando precipitaciones mensuales por percepción remota y los índices climáticos globales como entradas externas, en la cuenca del Mantaro.
- El índice de sequía más representativo en función al crecimiento y desarrollo de la vegetación dentro de la cuenca del Mantaro es el *SPEI de escala temporal 6 meses*, el mismo que presenta 0.58 de coeficiente de correlación y representa *correlación positiva moderada*.
- La relación entre los índices SPIAG y SPIMV, han demostrado ser similares. El promedio mensual del coeficiente de Nash para los índices SPIAG y SPIMV es de 0.983 y 0.969 respectivamente. La diferencia de los resultados entre éstas dos técnicas, numérica (Máxima Verosimilitud) e inteligencia artificial (Algoritmos Genéticos), es ligera (0.014).
- Mediante la técnica de L-Momentos, se han identificado los periodos de recurrencia de 5 años, donde se obtienen precipitaciones anuales menores a 700 mm, salvo las regiones 3 y 7 y para periodos de recurrencia de 100 años, las precipitaciones tienen a distribuirse de manera uniforme con precipitaciones menores a 400 mm, excepto las regiones 3 y 7. La región más afectada corresponde a la región 6 zona sur de la cuenca estudiada.

Recomendaciones

- De los resultados obtenidos, para grandes volúmenes de información, se recomienda el uso del índice SPI por Máxima Verosimilitud, debido al alto esfuerzo computacional por el método de los Algoritmos Genéticos.
- Para investigaciones futuras, se recomienda la implementación de un modelo de *optimización paralela*, para el análisis regional de frecuencia en la etapa de selección de las regiones homogéneas y evitar el procedimiento *prueba-error*, que en muchos casos conlleva tiempo y resultados no esperados.



- Dentro del territorio peruano, se recomienda el uso de las precipitaciones mensuales del producto PISCO, los mismos que están calibrados y validados con más de 420 estaciones que controla SENAMHI- Perú. Para otros territorios, las precipitaciones por sensoramiento remoto deberá ser calibrada y validada mediante las estaciones puntuales de los Servicios Meteorológicos e Hidrológicos Nacionales propios de cada territorio.
- Las Redes Neuronales SOM, clasifican una región de manera eficiente, sin emplear ningún patrón estadístico. De los ensayos realizados, se recomienda usar las Neuronas de Salida (mapas topológicos) mayores iguales al 10 % del número de Neuronas de Entrada y grandes ciclos de cálculo (mayores a 5000).
- Para realizar modelos de predicción en series de tiempo, utilizando técnicas de Inteligencia Artificial, se recomienda el uso de los modelos dinámicos, es decir, las *Redes Neuronales Recurrentes* como Nonlinear AutoRegressive Network (NAR) o NARX, donde las salidas no dependen únicamente de las entradas (redes estáticas), sino también de las entradas, salidas y estados de la red actual o anterior.
- Se recomienda tomar mayor interés en la gestión local, regional y nacional para la mitigación adecuada de la región sur de la cuenca de estudio, que corresponden al departamento Ayacucho, el mismo que presenta mayores frecuencias o recurrencias a las sequías.
- Se recomienda usar la escala temporal de 6 meses, para el monitoreo de las sequías agrícolas dentro de la cuenca del Mantaro, empleando el índice SPEI.
- Realizar modelos completos definiendo una relación entre las sequías climáticas e hidrológicas para determinar el criterio de un diagnóstico real desde los parámetros climatológicos. De esta manera, crear sistemas de alarmas cuando es considerado sequías severas y responder a través de una gestión local (operacional y demanda), gestión regional y gestión de sequía Nacional.
- La aplicación del modelo de predicción RNA-NARX, se puede dar en las distintas regiones dentro del territorio peruano, siguiendo la metodología propuesta en la tesis, estableciendo la actualización de la arquitectura de la red y seleccionando de manera adecuada los indicadores de sequía.

Glosario

caja negra : aquel elemento que es estudiado desde el punto de vista de las entradas que recibe y las salidas o respuestas que produce, sin tener en cuenta su funcionamiento interno. 35

capa : agrupación de neuronas dispuestas en un nivel de una red neuronal. 39, 40

darwiniano : término con el que se describen las ideas de *Charles Darwin*, especialmente en relación a la evolución biológica por selección natural. 43

entrenamiento : proceso en el cual una red neuronal modifica sus pesos sinápticos en respuesta a una entrada, para proporcionar la salida adecuada. 34, 37

evaluacion : a cada uno de los cromosomas de esta población se aplicará la función de aptitud para saber cómo de *bueno* es la solución que se está codificando. 47

función de transferencia : función matemática que transforma la entrada total en la respuesta de una neurona artificial. 36

individuos : elemento de la población, cada individuo es representado por una cadena de caracteres o genes. 44

metaheurística : es un método heurístico para resolver un tipo de problema computacional general, usando los parámetros dados por el usuario sobre unos procedimientos genéricos y abstractos de una manera que se espera eficiente. 5

multiobjetivo : la necesidad de tener en cuenta muchas variables en estudio a fin de obtener una solución óptima. 49

neurona : célula Nerviosa, con capacidad de procesar, modificar y manipular, generando señales de salida (información). 35

outlier : valores atípicos en una serie de tiempo. 17

perceptrón : modelo más simple de una red neuronal, monocapa sin presencia de capas ocultas.. 38, 39

perceptrón multicapa : modelo de redes neuronales artificiales, se caracteriza por la organización de sus neuronas en capas de distintos niveles. 38, 39



Python : lenguaje de programación con licencia de código abierto. 48

R : lenguaje de programación principalmente orientado al análisis estadístico y visualización de información, que incluye herramientas de análisis de datos y generación de gráficas y mapas. Esta publicado como software libre con licencia GNU-GPL y funciona bajo Windows, MAC OS y Linux. 11, 48

Acrónimos

P_{NE}	probabilidad de no excedencia. 79
T_R	Periodo de retorno. 79, 83
AG	Algoritmos Genéticos. 5, 11, 28, 43, 44, 46, 48, 49, 53, 55
AMO	Atlantic Multidecadal Oscillation. 10
ANA	Autoridad Nacional del Agua. 1, 6, 13
ARF	Análisis Regional de Frecuencias. 2, 24
ARF-LM	Análisis Regional de Frecuencias-L Momentos. 16, 17, 20
ASCE	Sociedad Americana de Ingenieros Civiles. 31
AVHRR	Advanced Very High Resolution Radiometer. 5
CCR	Curva de Crecimiento Regional. 23
CHG	Climate Hazards Group. 6
CHIRPS	Climate Hazards Group Infrared Precipitation with Stations. 1, 12
CPC	Climate Prediction Center. 10
DANNGA	Drought Artificial Neural Network and Genetic Algorithm. 11
ENSO	El Niño Southern Oscillation. 5, 58
FAO	Organización de las Naciones Unidas para la Agricultura y la Alimentación. 31
IA	Inteligencia Artificial. 1, 8
IAR	Índice de Aridez. 32, 68
ICG	Índices Climáticos Globales. 1, 2, 10, 11, 54, 60, 93–95
ICID	Comisión Internacional de Riegos y Drenajes. 31
ICIWaRM	Centro Internacional de Gestión Integrada de Recursos Hídricos. 24
IE	Índice de Estacionalidad. 33, 70



IES	Índices Efectivos de Sequías. 4
IFM	Índice de Fournier Modificado. 33
MSE	Error Cuadrático Medio. 62, 63, 93
MVS	Máquinas de Vector Soporte. 4
NAR	Nonlinear AutoRegressive Network. 102
NARX	Nonlinear Autoregressive Network with Exogenous Inputs. 2, 42, 43, 53, 54, 60, 93, 94, 98, 99, 101, 102
NASA	National Aeronautics and Space Administration. 6, 48
NDVI	Índice de Vegetación de Diferencia Normalizada. 5, 10, 11, 27, 85
NE	neuronas de entrada. 40, 73
NOAA	National Oceanic and Atmospheric Administration. 5, 10, 59
NS	neuronas de salida. 40, 71, 73
OMM	Organización Meteorológica Mundial. 28
ONI	Oceanic Niño Index. 10
PDO	Pacific Decadal Oscillation. 10
PDSI	Índice de Severidad de Sequía de Palmer. 26, 85
PET	Evapotranspiración Potencial. 30, 31
PISCO	Peruvian Interpolation data of the SENAMHI's Climatological and Hidrological Observations. 1, 6, 10, 11, 13, 14, 51, 65, 91, 102
PM	Penman-Moteith. 31
PMA	Precipitación Media Anual. 53, 66, 68, 74, 82, 83
PMAR	Precipitación Media Anual Regional. 23, 75, 76, 79, 80
RNA	Redes Neuronales Artificiales. 2, 4, 5, 11, 33–39, 53, 54, 93, 99–101
RNAM	Redes Neuronales Artificiales Multicapa. 38
SENAMHI	Servicio Nacional de Meteorología e Hidrología. 1, 6, 10, 11, 13, 51, 52, 65, 91, 102
SMHN	Servicios Meteorológicos e Hidrológicos Nacionales. 28
SOI	Southern Oscillation Index. 10
SOM	Self Organizing Map. 2, 20, 40, 42, 53, 70, 71, 73, 102



SPEI	Índices de Precipitación Evapotranspiración Estandarizada. 2, 10, 26, 28, 30, 32, 53, 62, 85, 91, 93–95, 98–102
SPI	Índices de Precipitación Estandarizada. 4, 5, 7, 8, 10, 11, 15, 16, 26, 28–30, 32, 53, 62, 84, 85, 88, 91, 93, 101
SPIAG	Índices de Precipitación Estandarizada Algoritmos Genéticos. 2, 53, 88, 89, 101
SPIGA	Standardized Precipitation Index using Genetic Algorithm Method. 55
SPIMV	Índices de Precipitación Estandarizada Máxima Verosimilitud. 2, 53, 88, 89, 101
TDL	Línea de retardo. 43
TRMM	Tropical Rainfall Measuring Mission. 5
TSM	Temperatura de la Superficie del Mar. 57, 58
VCI	Índice de Condición de Vegetación. 10, 11, 26–28, 53, 62, 85, 86, 91, 93
VCR	Variables Climáticos Regionales. 51, 52
ZCI	Zona de Convergencia Intertropical. 1

Símbolos

- A parámetro auxiliar para determinar α . 29
- $\Gamma(\alpha)$ Función Gamma, $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$. 28
- $G(x)$ función de distribución acumulada Gamma. 29
- $H(x)$ función de distribución acumulada Gamma Completa. 29
- IP interpolación. 13
- P_{CHIRPS} precipitación Mensual CHIRPS. 13
- P_C precipitación con corrección. 13
- $P_{SENAMHI}$ precipitación Mensual de las estaciones SENAMHI. 13
- $Q^{(i)}$ cuantil de la estación i . 23
- w_{ij} pesos de las conexiones de la neurona i hacia la neurona j . 38
- $|Z^{DIST}|$ Estadístico de ajuste distribución L-momentos. 23
- α parámetro de forma de la distribución gamma. 28
- β parámetro de escala de la distribución gamma. 28
- η tasa de aprendizaje. 42
- $g(x)$ función de densidad de probabilidad Gamma. 29
- $l_1^{(i)}$ media de los datos observados en la estación i . 23
- nn número de datos no nulos. 29
- $q(F)$ curva de crecimiento regional. 23
- q probabilidad de valores nulos. 29
- σ_4 desviación estándar promedio regional. 23
- τ_3 L-Momento sesgo. 18
- τ_4^{DIS} L-Kurtosis de la distribución adaptada. 23
- τ_4^R L-kurtosis promedio regional. 23
- τ_4 L-Momento curtosis. 18
- τ L-Momento Coeficiente de variación. 18



\bar{x} media aritmética de la variable. 29

x precipitación ó variable. 28

Referencia Bibliográfica

Precipitación Percepción Remota

- Caicedo Carrascal, F. M. (2008). *Asimilación de precipitación estimada por imágenes de satélite en modelos hidrológicos aglutinados y distribuidos, caso de estudio afluencias al embalse de betania (huila, colombia)* (MA. Thesis, Pontificia Universidad Javeriana, Facultad de Ingeniería).
- Funk, C. C., Peterson, J., Pete, Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Shukla, S., ... Andrew, H. (2015). The climate hazards infrared precipitation with station a new environmental record for monitoring extremes. *Science data*, 2(150066). doi:10.1038/sdata.2015.66
- Hong, Y. & Gourley, J. J. (2015). *Radar hydrology principles, models, and applications*. Taylor y Francis Group. Recuperado desde <https://www.crcpress.com/Radar-Hydrology-Principles-Models-and-Applications/Hong-Gourley/9781466514614>
- Masud Cheema, M. J. & G., M. B. W. (2012). Local calibration of remotely sensed rainfall from the trmm satellite for different periods and spatial scales in the indus basin. *International Journal of Remote Sensing*, 33, 2603-2627. doi:10.1080/01431161.2011.617397
- Ponce, V. M. (1989). *Engineering hydrology principles and hydrology*. Prentice hall.
- Shaofeng, J., Wenbin, Z., Aifeng, L. & Tingting, Y. (2011). A statistical spatial downscaling algorithm of trmm precipitation based on ndvi and dem in the qaidam basin of china. *Remote Sensing of Environment*, (115), 3069-3079. doi:10.1016/j.rse.2011.06.009
- Takahashi, K., Montecinos, A., Gobanova, K. & Dewite, B. (2011). Enso regimes: reinterpreting the canonical and modoki el niño. *Geophysical Reserach letters*, 38, 1-5. doi:10.1029/2011GL047364

Sequías

- Acuña, J., Felipe, O., Ordoñes, J. & Arboleda, F. (2011). Análisis regional de frecuencia de precipitación anual para la determinación de mapas de sequías. *Revista peruana geo-atmosférica RPGA*, (3), 104-115.
- Andreu, J., Solera, A., Paredes Arquiola, J., Haro Moteagudo, D. & Van Lanen, H. (Eds.). (2015). Drought research and science policy interfacing.



- Belayneh, A., Adamowski, J., Khalil, B. & Ozga-Zielinski, B. (2013). Long-term spatiotemporal drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508, 418-429. doi:10.1016/j.jhydrol.2013.10.052
- Campos, D. F., Aranda. (2014). Comparación de tres métodos estadísticos para detección y monitoreo de sequías. *Agrociencia*, 48(5), 463-476. doi:48:463-476.2014
- Dastorani, M. T. & Afkhami, H. (2011). Application of artificial neural networks. *Desalination*, (16), 39-48.
- Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P. & Huang, Y. (2013, agosto). A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation*, 23, 245-253. doi:10.1016/j.jag.2012.09.010
- Edwards, C. D. & McKee, T. B. [T. B.]. (1997). Characteristics of 20th century drought in the united stated at multiple time scales. *Atmospheric Science*, (634).
- Farahmand, A. & AghaKouchak, A. (2015). A generalized framework for deriving non-parametric standardized drought indicators. *Advances in Water Resources*, 76, 140-145. doi:<http://dx.doi.org/10.1016/j.advwatres.2014.11.012>
- Guttman, N. B. [N. B.], Hosking, J. R. M. & Wallis, J. R. (1993). Regional precipitation quantile values for the continental united states computed from l-moments. *American Meteorological Society*, 6, 2326-2340.
- Guttman, N. B. [Nathaniel B.]. (1999). Accepting the standardized precipitation index: a calculation algorithm. *Journal of the American Water Resources Association*, 35(2).
- Hao, Z. & Singh, V. P. [Vijay P.]. (2015, agosto). Drought characterization from a multivariate perspective: a review. *Journal of Hydrology*, 527, 668-678. doi:<http://dx.doi.org/10.1016/j.jhydrol.2015.05.031>
- Haro Monteagudo, D., Momblanch Benavent, A., Andreu Álvarez, J., Solera Solera, A., Paredes Arquiola, J. & Van Lanen, H. (2015). *Technical report no 23 drought conference*. Universidad Politecnica de Valencia.
- Hassanzadeh, Y., Abdi Kordani, A. & Fakheri Fard, A. (2012). Drought forecasting using genetic algorithm and conjoined model of neural network-wavelet. *Journal of Water and Wastewater*, 23(3), 48-49.
- Hosking, J. R. M. & Wallis, J. R. (1997). *Regional frequency analysis an approach based on l -moments*. Cambridge University Press.
- Keskin, M. E., Terzi, O., Taylan, D. & Kucukyaman, D. (2011). Meteorological drought analysis artificial neural networks. *Scientific Research and Essays*, 6(21), 4469-4477. doi:10.5897/SRE10.1022
- Kogan, F. N. (1995). Application of vegetation index and brightness temperature for drought detection. *National Oceanic and Atmospheric Administration*, 15(11), 91-100.
- Kogan, F. N. & Sullivan, J. (1993). Development of global drought - watch system using noaa/avhrr data. *National Oceanic and Atmospheric*, 13(5), 219-222.
- Li, X., Zhang, Q. & Ye, X. (2013). Dry/wet conditions monitoring based on trmm rainfall data and its reliability validation over poyang lake basin, china. *Water*, 5, 1848-1864. doi:10.3390/w5041848



- Liu, Z., Wang, Y., Shao, M., Jia, X. & Li, X. (2016, marzo). Spatiotemporal analysis of multiscalar drought characteristics across the loess plateau of China. *Journal of Hydrology*, 534, 281-299. doi:10.1016/j.jhydrol.2016.01.003
- Lobo, L. D. (2006). *Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y el Caribe*.
- Marcos, V. O. (2001). Sequía: definiciones, tipologías y métodos de cuantificación. *Investigaciones Geográficas*, (26), 59-80.
- McKee, T. B. [Thomas B.], Doesken, N. J. & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Eighth Conference on Applied Climatology*.
- Mishra, A. K. & Singh, V. P. [V. P.]. (2010). A review of drought concept. *Journal of Hydrology*, 391. doi:10.1016/j.jhydrol.2010.07.012
- Morid, S., Smakhtin, V. & Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*, (27), 2103-2111. doi:10.1002/joc.1498
- Mutuga, K. J., Nyadawa, M. O. & Home, P. G. (2014). Use downscaled tropical rainfall measurement mission data for meteorological drought monitoring: case study of narumoru catchment. *International Journal of Advances in Engineering and Technology*, 7(5), 1375-1385. Recuperado desde <https://doaj.org/article/626becfad6e8430fb9520c42e3584155>
- Nagaranjan, R. (2009). *Drought assessment*. Capital Publishing Company.
- Núñez, J. H., Verbist, K., Wallis, J. R., Chaefer, M. G., Morales, L. & Cornelis, W. M. (2011, agosto). Regional frequency analysis for mapping drought events in north-central Chile. *Journal of Hydrology*, 405(3-4), 352-366. doi:10.1016/j.jhydrol.2011.05.035
- Núñez, J. & Verbist, K. (2010). *Guía metodológica para la aplicación del análisis regional de frecuencia de sequías basado en l-momentos y resultados de aplicación en América Latina* (UNESCO, Ed.).
- OMM, Svoboda, M., Hayes, M. & Wood, D. (2012). *Guía de usuario sobre el índice normalizado de precipitación*. Ginebra.
- Palmer, W. C. (1965). Meteorological drought. *U.S. Department of Commerce*. (45).
- Quiring, S. M. & Ganesh, S. (2010, marzo). Evaluating the utility of the vegetation condition index (VCI) for monitoring meteorological drought in Texas. *Agricultural and Forest Meteorology*, 150(3), 330-339. doi:10.1016/j.agrformet.2009.11.015
- Rojas Acuña, J., Rufasto Campos, E., Eche Llenque, J. C. & Carrillo Gomero, F. (2008). Modelo de la sequía en la costa norte de Perú usando los índices ENOS y las imágenes NDVI del sensor AHRR a bordo del satélite NOAA. *Electrónica - UNMSM*, (21), 12-23.
- Rojas, O., Vrielink, A. & Rembold, F. (2011). Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. *Remote Sensing of Environment*, 115, 343-252. doi:10.1016/j.rse.2010.09.006
- Rossi, G., Vega, T. & Brunella, B. (2007). *Methods and tool for drought analysis and management*. Springer.
- Sánchez Romero, C. O. . (2009). *Influencia de variables macroclimáticas en la sequía meteorológica en el sur del Perú* (MA. Thesis, Universidad Politécnica de Valencia).



- Schwabe, K., Albiac, J., Conor, J. D., Hassan, R. M. & Meza Gonzales, L. (2011). *Drought in arid and semi-arid regions*. Springer Science+Business Media Dordrecht. doi:10.1007/978-94-007-6636-5
- Serianldi, F., Bonaccorso, B., Cancelliere, A. & Grimaldi, S. (2009). Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth*, 34, 596-605. doi:10.1016/j.pce.2008.09.004
- Velasco, I., Ochoa, L. & Gutiérrez, C. (2005). Sequía, un problema de perspectiva y gestión. *Región y sociedad*, 17(34), 35-71.
- Vicente Serrano, S. M., Beguería, S., Lorenzo La Cruz, J., Camarero, J. J., López Moreno, J., Azorín Molina, C., ... Sánchez Lorenzo, A. (2012). Análisis comparativo de diferentes índices de sequía para aplicaciones ecológicas, agrícolas e hidrológicas. *Asociación Española de Climatología*, 679-689. Recuperado desde <http://hdl.handle.net/10261/62168>
- Vicente-Serrano, S. M. (2006, octubre). Evaluating the impact of drought using remote sensing in a mediterranean, semi-arid region. *Nat Hazards*, 40(1), 173-208. doi:10.1007/s11069-006-0009-7
- Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. (2010, abril). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696-1718. doi:10.1175/2009jcli2909.1
- Wallis, J. R., Schaefer, M. G., Barker, B. L. & Taylor, G. H. (2007). Regional precipitation - frequency analysis and spatial mapping for 34-hour and 3-hour duration for washington state. *Hydrology and Earth System Sciences*, 1(11), 415-442.
- Wambua, R. M., Mutua, B. M. & Raude, J. M. (2014). Drought forecasting using indices and artificial neural networks for upper tana river basin, kenya-a review concept. *Civil & Environmental Engineering*, 4(4). doi:10.4172/2165-784X.1000152
- Wilhite, D. A. & Glantz, M. H. (1985). Understanding the drought phenomenon: the role of definition. *Water International*, 10(3), 111-120.
- Yang, T., Shao, Q., Hao, Z., Chen, X., Zhang, Z., Xu, C. & L., S. (2010, enero). Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the pearl river basin, China. *Journal of Hydrology*, 380(3-4), 386-405. doi:10.1016/j.jhydrol.2009.11.013
- Yevjevich, V. (1967). An objetive aproach to definitions and investigations of continental hydrologic drought. *Hydrology papers Colorado State University*, (23).
- Zargar, A., Sadiq, R., Naser, B. & Khan, F. I. (2011, diciembre). A review of drought indices. *Environmental Reviews*, 19(NA), 333-349. doi:10.1139/A11-013
- Zhang, A. & Jia, G. (2013, julio). Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. *Remote Sensing of Environment*, 134, 12-23. doi:10.1016/j.rse.2013.02.023

Algoritmos Genéticos

- Beasley, D., Bull, D. R. & Martin, R. R. (1993). An overview of genetic algorithm: part 1, fundamentals. *University Computing*, 15(2).
- De Jong, K. A. (1975). *Analysis of the behaviour of a class of genetic adaptive systems* (Tesis doctoral, University of Michigan).



- E. Goldberg, D. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wensley Publishing Company.
- Manoj, K., Mohammad, H., Naveen, U. & Deepti, G. (2010). Genetic algorithm: review and application. *International Journal of Information Technology and Knowledge Management*, 2(2), 451-454.
- El-Mihoub, T. A., Hopgood, A. A., Nolle, L. & Battersby, A. (2006). Hybrid genetic algorithms: a review. *Engineering Letters*, 13(2), 12-25.
- Mitchell, M. (1995). Genetic algorithms: an overview. *Complexity*, (1), 31-39.
- Mitchell, M. (1996). *An introduction to genetic algorithms* (First, Ed.). Massachusetts Institute of Technology.
- Mujahid, T. & Mathew, K. (2014). A genetic algorithm analysis towards optimization solutions. *International Journal of Digital Information and Wireless Communications*, (4).
- Xin She, Y., Amir Hossein, G., Siamak, T. & Amir Hossein, A. (2012). *Metaheuristics in water, geotechnical and transport engineering* (First, Ed.). Elsevier. Recuperado desde <http://store.elsevier.com/Metaheuristics-in-Water-Geotechnical-and-Transport-Engineering/isbn-9780123982964/>

Redes Neuronales Artificiales

- ASCE. (2000a, abril). Artificial neural networks in hydrology. i : preliminary concepts. *Journak of Hydrologic Engineering*, 5(2), 115-123. doi:[http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(115\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115))
- ASCE. (2000b, abril). Artificial neural networks in hydrology. ii : hydrologic applications. *Journak of Hydrologic Engineering*, 5(2), 115-123. doi:[http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(124\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124))
- Basheer, I. A. & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design and application. *Jouurnal of Microbiological Methods*, (43), 3-31.
- Basogain Olabe, X. (2000). *Redes neuronales artificiales y sus aplicaciones*. Ingeniería de Sistemas y Automática.
- Beale, M. H., Hagan, M. T. & Demuth, H. B. (2014). *Neural network toolbox user guide* (R2014a). The MathWorks, Recuperado desde www.mathworks.com
- Guoqiang, Z., Patuwo., B. E. & Y, H. M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 1(14), 35-62.
- Gutiérrez, J. M., Cano, R., Cofiño, A. S. & Sordo, C. M. (2004). *Redes probabilísticas y neuronales en las ciencias atmosféricas* (M. de Medio Ambiente - España, Ed.).
- Hagan, M. T., Demuth, H. B., Beale, M. H. & De Jesús, O. (1996). *Neural network design* (2da). Boston PWS Publishing. Recuperado desde hagan.okstate.edu/nnd.html
- Haykin, S. (1999). *Neural networks a comprehensive foundation* (Second, Ed.). Pearson Education. Recuperado desde <http://www.amazon.com/Neural-Networks-A-Comprehensive-Foundation/dp/B000O8QMAU>
- J. Abrahart, R., M. See, L. & P. Solomatine, D. (2008). *Practical hydroinformatics computational intelligence and technological developments in water applications*. Springer-Verlag. Recuperado desde <http://www.springer.com/us/book/9783540798804>



- Kohonen, T. [T.]. (1990). The self-organizing map. *IEEE*, 78(9).
- Kohonen, T. [T.]. (2001). *Self-organizing maps* (Third Edition).
- Kohonen, T. [Teuvo]. (2013). Essential of the self-organizing map. *Neural Networks*, 37, 52-65. doi:10.1016/j.neunet.2012.09.018
- Krasnopolksky, V. M. (2013). *The application of neural networks in the earth system sciences*. Springer Science. Recuperado desde <http://www.springer.com/us/book/9789400760721>
- Lek, S. & Guégan, J. (1999). Artificial neural networks as a tool in ecological model an introduction. *Ecological modeling*, (120).
- Li, E. Y. (1994). Artificial neural networks and their business applications. *Information Management*, (27), 303-313.
- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Pelaéz Chávez, N. (2012). *Aprendizaje no supervisado y el algoritmo wake-sleep en redes neuronales*.
- Tayfur, G. (2012). *Soft computing in water resources engineering artificial neural networks, fuzzy logic and genetic algorithm* (B. WIT Press Southampton, Ed.).

Metodología de Investigación y Estadística

- Forte, R. M. (2015). *Mastering predictive analytics with r*. Open Source.
- Hernández Sampieri, R., Hernández Collado, C. & Baptista Lucio, P. (2014). *Metodología de la investigación* (6ta). Mc Graw Hill.

Fuentes Electrónicas

- Cano, J. L. (2012). Introducción a python para científicos e ingenieros. Recuperado el 28 de septiembre de 2015, desde <http://pybonacci.org/2012/03/16/introduccion-a-python-para-cientificos-e-ingeneros/>
- Soto P., D. & Del Caltillo P., L. (2013). Informe nacional del perú. Recuperado el 21 de junio de 2016, desde http://www.ais.unwater.org/ais/pluginfile.php/571/mod_page/content/88/PERU_2_2.pdf
- Team, T. O. D. M. (1997). Oklahoma drought management plan. Recuperado el 1 de octubre de 2015, desde https://www.owrb.ok.gov/supply/drought/reports/drought_plan.pdf

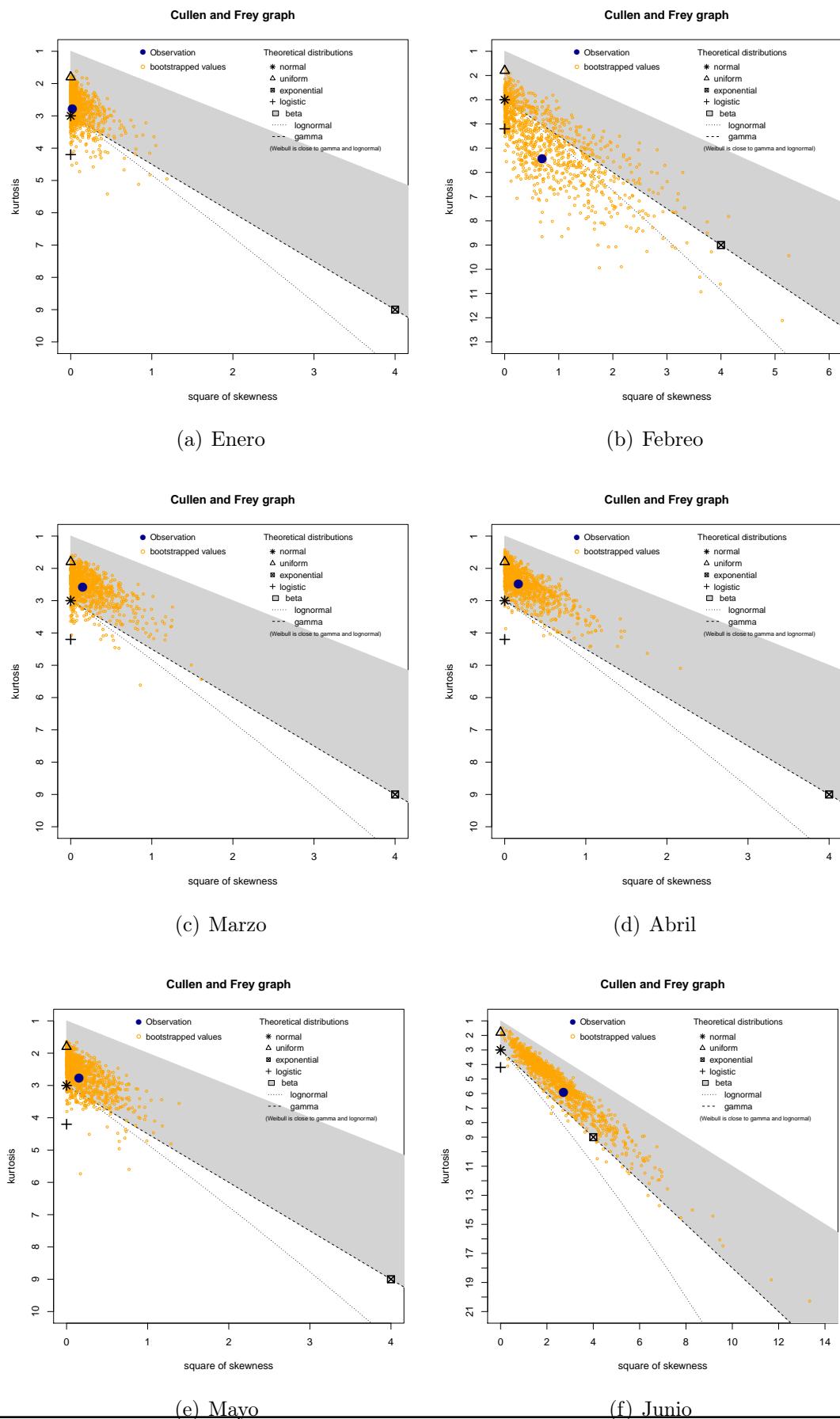
Anexo A

Análisis Estadístico de precipitaciones

Código Fuente A.1: Rutinas R para El Análisis Estadístico

```
1 library(fitdistrplus)
2 library(RHydrology)
3 library(xtable)
4 dataF<- read.table('Data_Pmensual.txt', header=T)
5 P <-dataF[,2:ncol(dataF)]
6 n_col <-ncol(P)
7 # Ploteo de Cullen y Frey
8 for(i in 1:n_col){
9   titleCullenFrey <- paste0("CF_", colnames(P[i]))
10  pdf(paste(titleCullenFrey, ".pdf"))
11  descdist(P[,i], boot=1000);dev.off()
12  boxplot(P[,1:n_col], col="cyan")
13  sk_n<-c();sk_ln<-c();sk_g<-c();xc_n<-c();xc_ln<-c();xc_g<-c();ad_n<-c();ad_ln<-
14    c();ad_g<-c()
15  for(i in 1:n_col){
16    fn <- fitdist(P[,i],"norm");fln <- fitdist(P[,i],"lnorm");fg <- fitdist(P[,i] ,
17      "gamma")
18    titleCDF <- paste0("CDF_", colnames(P[i]))
19    pdf(paste(titleCDF, ".pdf"))
20    cdfcomp(list(fn,fln,fg),legendtext=c("Normal", "Lognormal", "Gamma"),xlab="x",
21      lwd=2);grid();dev.off()
22    gn<-gofstat(fn); gln<-gofstat(fln);gg<-gofstat(fg)
23    sk_n[i]<-gn$ks; sk_ln[i]<-gln$ks; sk_g[i]<-gg$ks
24    xc_n[i]<-gn$chisq; xc_ln[i]<-gln$chisq;xc_g[i]<-gg$chisq
25    ad_n[i]<-gn$ad; ad_ln[i]<-gln$ad;ad_g[i]<-gg$ad
26  }
27 estNames<-c('s-k(Normal)', 'x-c(Normal)', 'a-d(Normal)', 's-k(Log Normal)', 'x-c(←
28   Log Normal)', 'a-d(Log Normal)', 's-k(Gamma)', 'x-c(Gamma)', 'a-d(Gamma)')
29 distMatrix <-rbind(sk_n,xc_n,ad_n,sk_ln,xc_ln,ad_ln,sk_g,xc_g,ad_g)
30 print(distMatrix)
31 rownames(distMatrix)<-estNames;colnames(distMatrix)<-colnames(P);print(←
32   distMatrix)
33 write.table(distMatrix, file = "O_Statistic.txt", row.names=T, col.names=T, sep="\t")
34 xtable(distMatrix,caption ='Estadiisticos',digits = 2 )
```

Figura A.1: Gráfico de Cullen y Frey región de análisis



Predicción de sequías con Redes Neuronales Artificiales y Algoritmos Genéticos utilizando precipitación por percepción remota.

Iván Arturo Ayala Bizarro

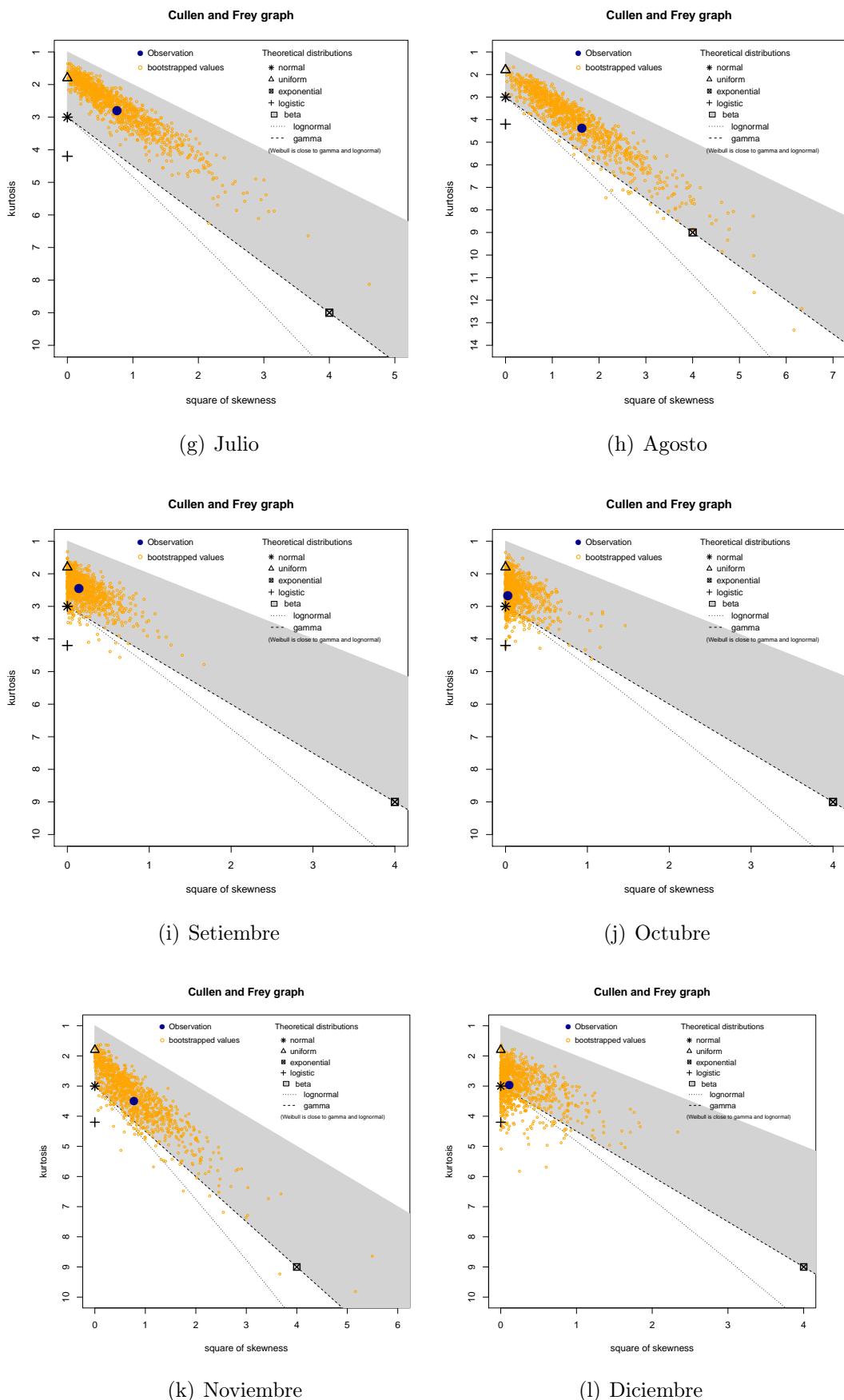
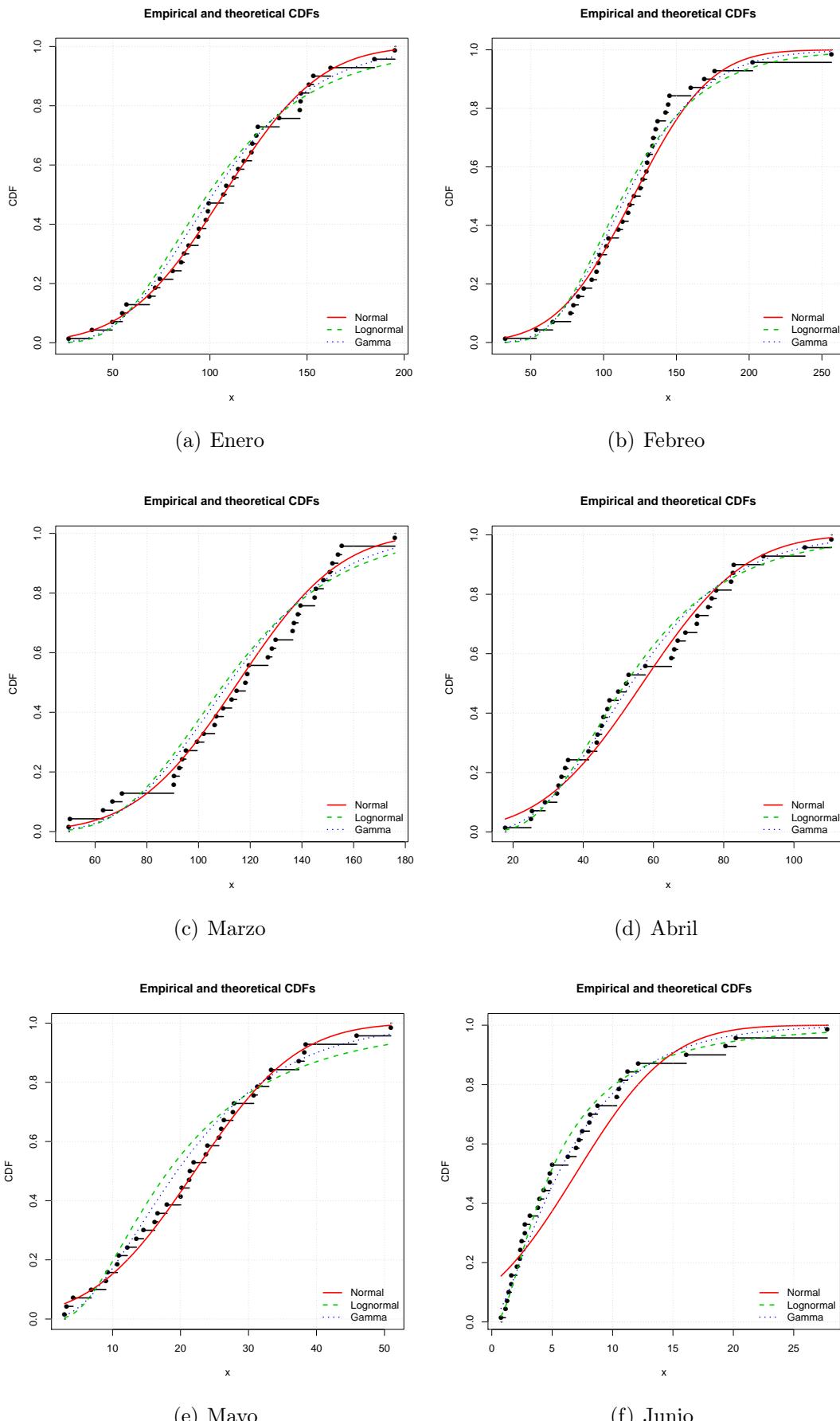
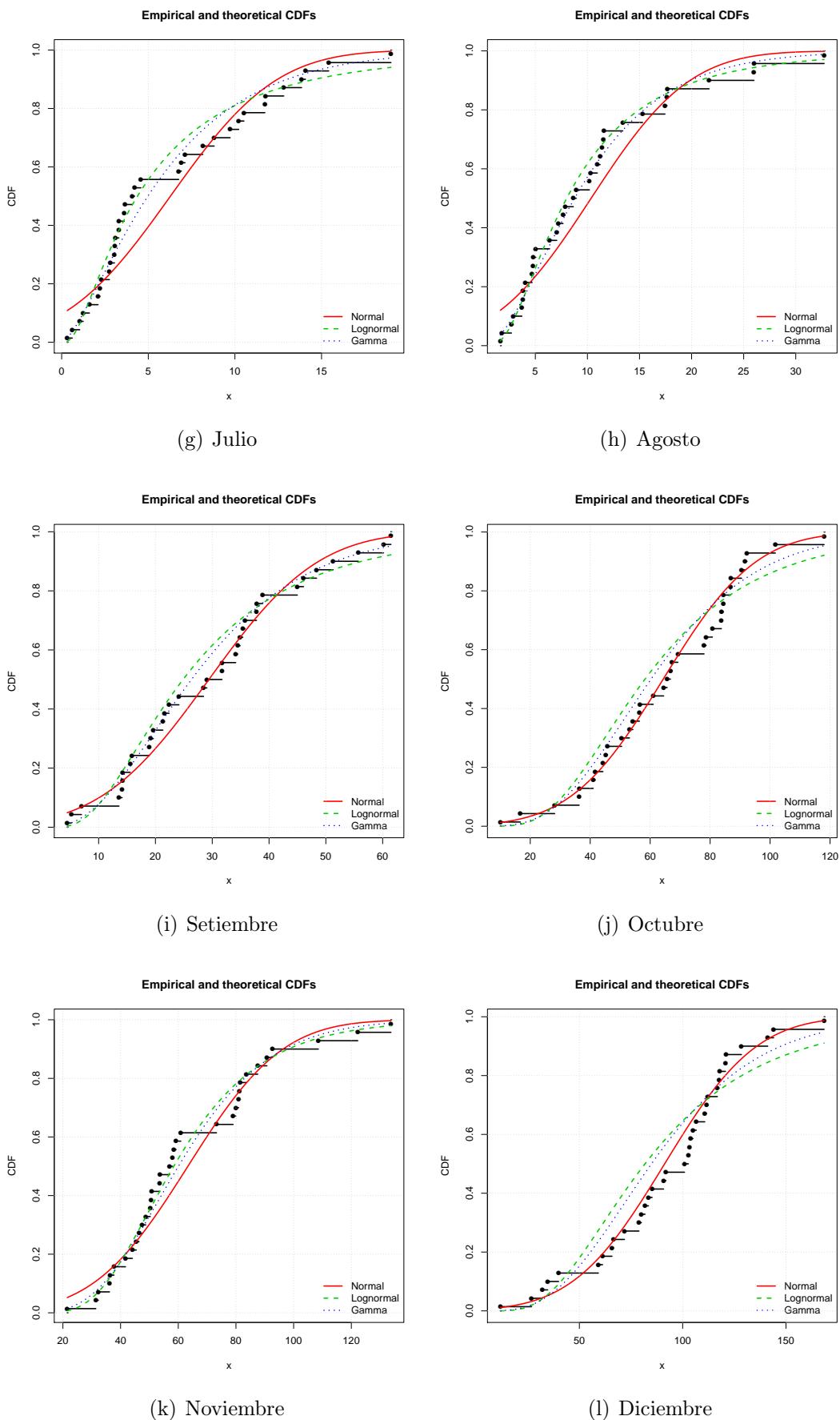


Figura A.2: Distribuciones Empírica y Teórica Acumulada región de análisis



Predictión de sequías con Redes Neuronales Artificiales y Algoritmos Genéticos utilizando precipitación por percepción remota.

Iván Arturo Ayala Bizarro



Fuente: Elaboración propia

Predicción de sequías con Redes Neuronales Artificiales y Algoritmos Genéticos utilizando precipitación por percepción remota.

Iván Arturo Ayala Bizarro

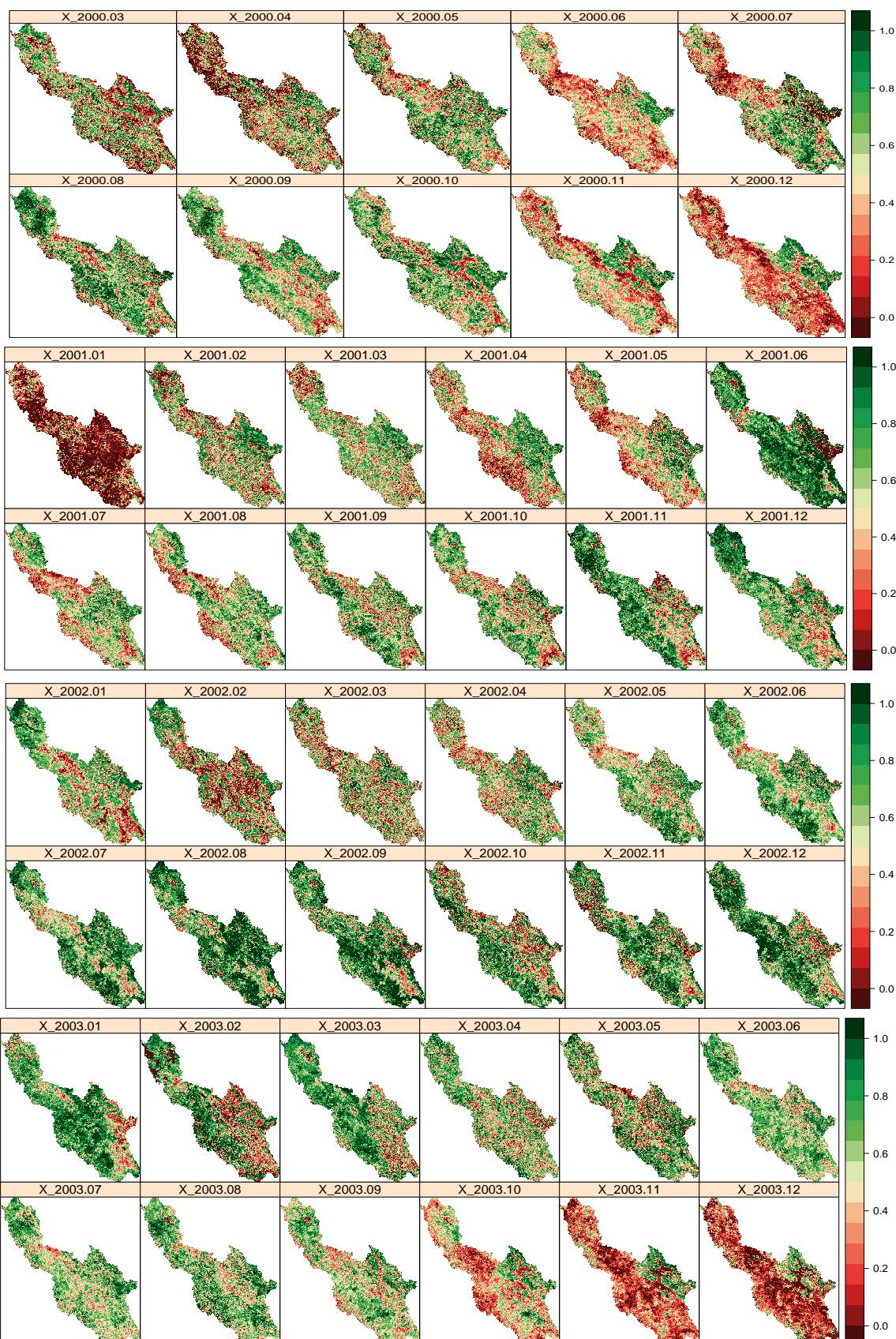
Anexo B

Resultados Índices de condición de Vegetación cuenca Mantaro

Código Fuente B.1: Rutinas R para obtener el VCI en formatos rasters

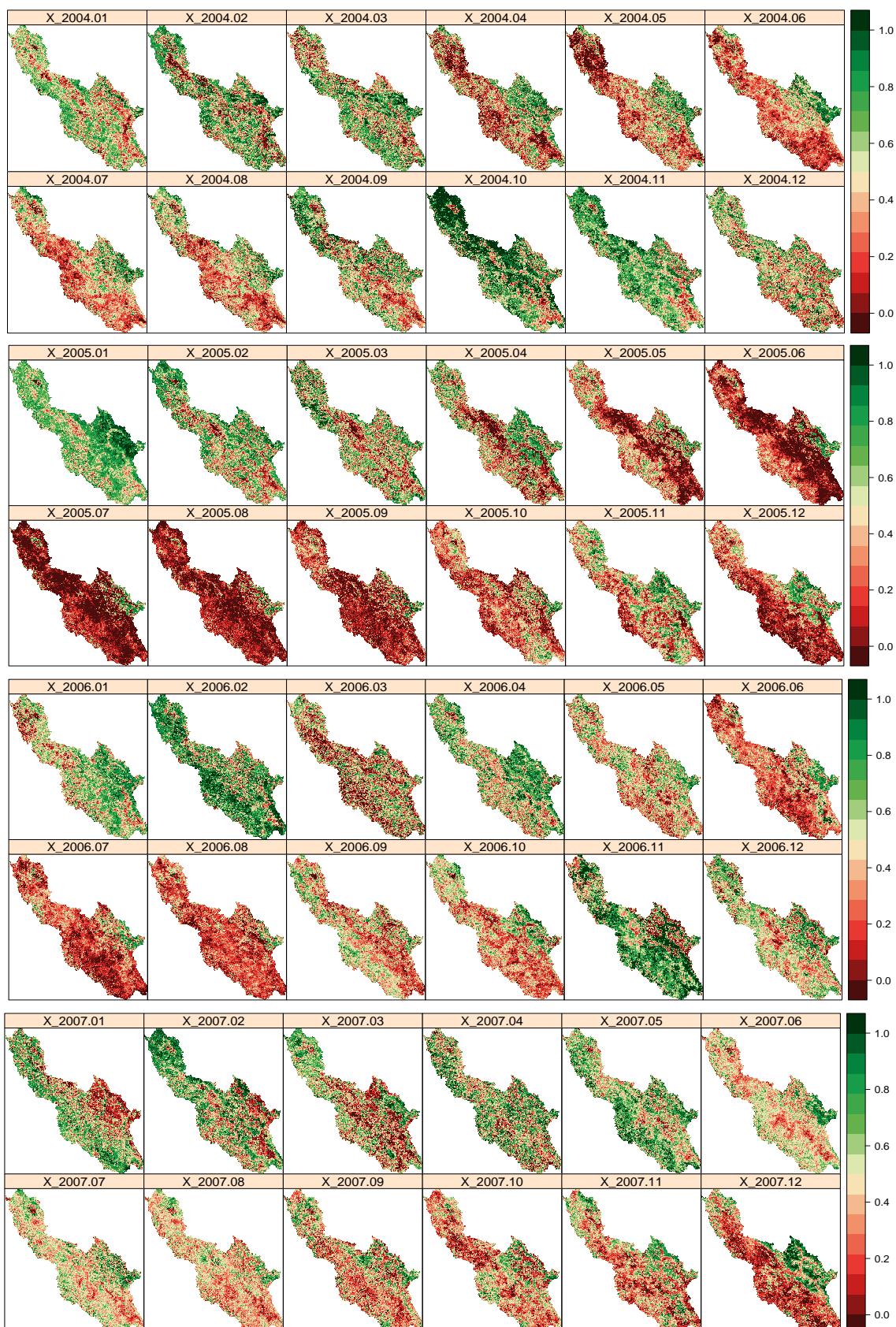
```
1 library(raster);library(rgdal);library(sp);library(rasterVis)
2 n_cuenca <- 'AB_Limite_Mantaro';cuenca<- readOGR(' ',n_cuenca)
3 cuenca<- spTransform(cuenca, CRS('+proj=longlat +datum=WGS84'))
4 dimname<-c(6,21)
5 ##### Cargar el Raster para el tratamiento de datos
6 files_1<- list.files(pattern='01.tif', full.names = T)
7 files_2<- list.files(pattern='02.tif', full.names = T)
8 files_3<- list.files(pattern='03.tif', full.names = T)
9 files_4<- list.files(pattern='04.tif', full.names = T)
10 files_5<- list.files(pattern='05.tif', full.names = T)
11 files_6<- list.files(pattern='06.tif', full.names = T)
12 files_7<- list.files(pattern='07.tif', full.names = T)
13 files_8<- list.files(pattern='08.tif', full.names = T)
14 files_9<- list.files(pattern='09.tif', full.names = T)
15 files_10<- list.files(pattern='10.tif', full.names = T)
16 files_11<- list.files(pattern='11.tif', full.names = T)
17 files_12<- list.files(pattern='12.tif', full.names = T)
18 for (mes in 1:12){
19 cat(sprintf('***** Calculando Mes %d \n',mes))
20 n_files<-paste('files_',as.character(mes), sep="")
21 files <-eval(as.name(n_files))
22 NDVI<-stack(files); e <- extent(cuenca)
23 NDVI_val <- extract(NDVI,e, nl=nlayers(NDVI), na.rm=T)
24 nrow <- nrow(NDVI_val);ncol <- ncol(NDVI_val);Lmax <-c();Lmin <-c()
25 for(i in 1:nrow){
26 Lmax[i] <- max(NDVI_val[i,]);Lmin[i] <- min(NDVI_val[i,])
27 ##### Caculo del VCI
28 setTxtProgressBar (pb, 0)
29 dif <- (Lmax-Lmin)
30 for(i in 1:ncol){
31 VCI <- (NDVI[[i]]-Lmin)/dif
32 names(VCI) <- paste("VCI",substr(files[i], dimname[1], dimname[2]))
33 writeRaster(VCI, names(VCI), format = "GTiff", overwrite=T)}}
```

Figura B.1: Índices de condición de Vegetación: 2000-2003



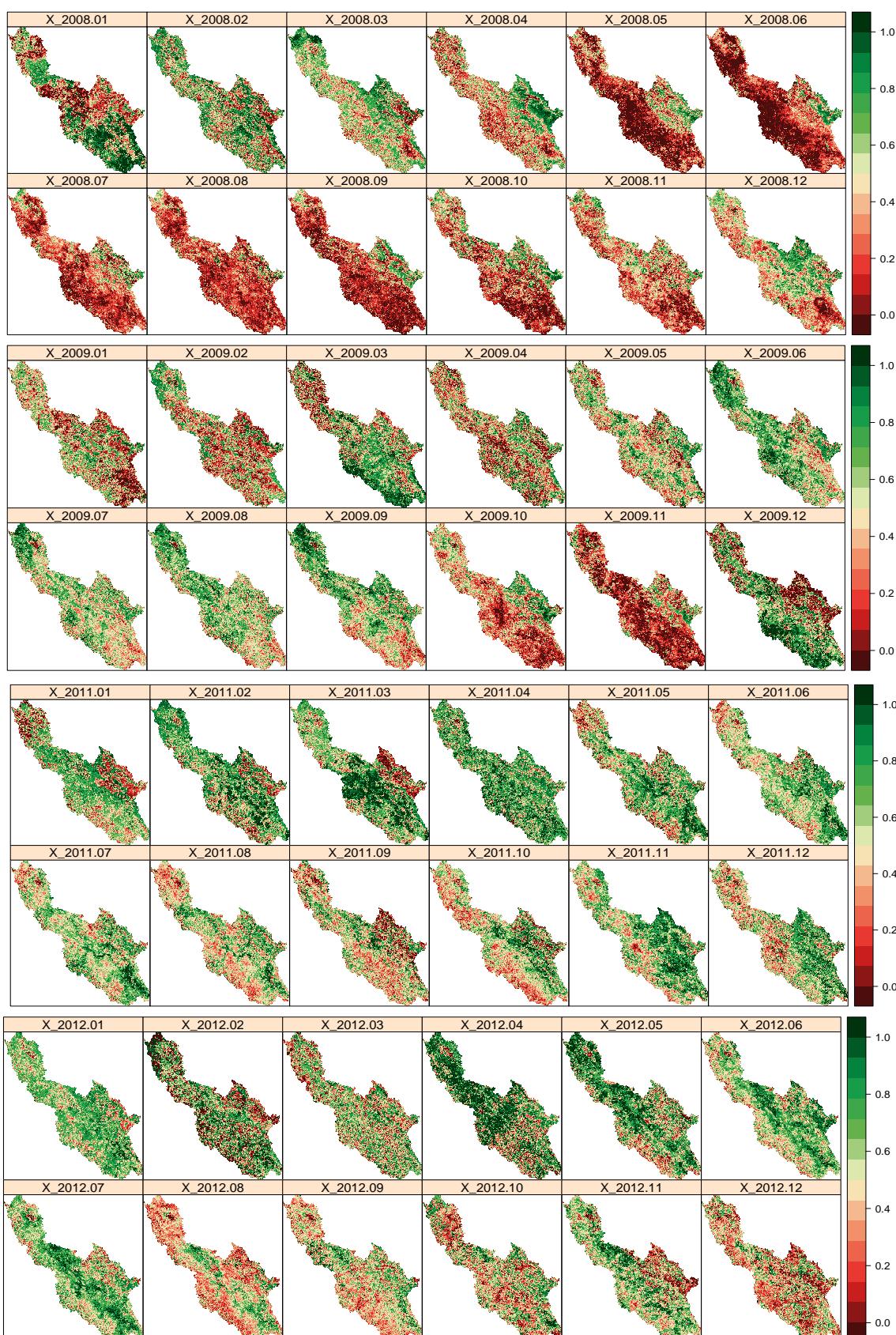
Fuente: Elaboración propia

Figura B.2: Índices de condición de Vegetación: 2004-2007



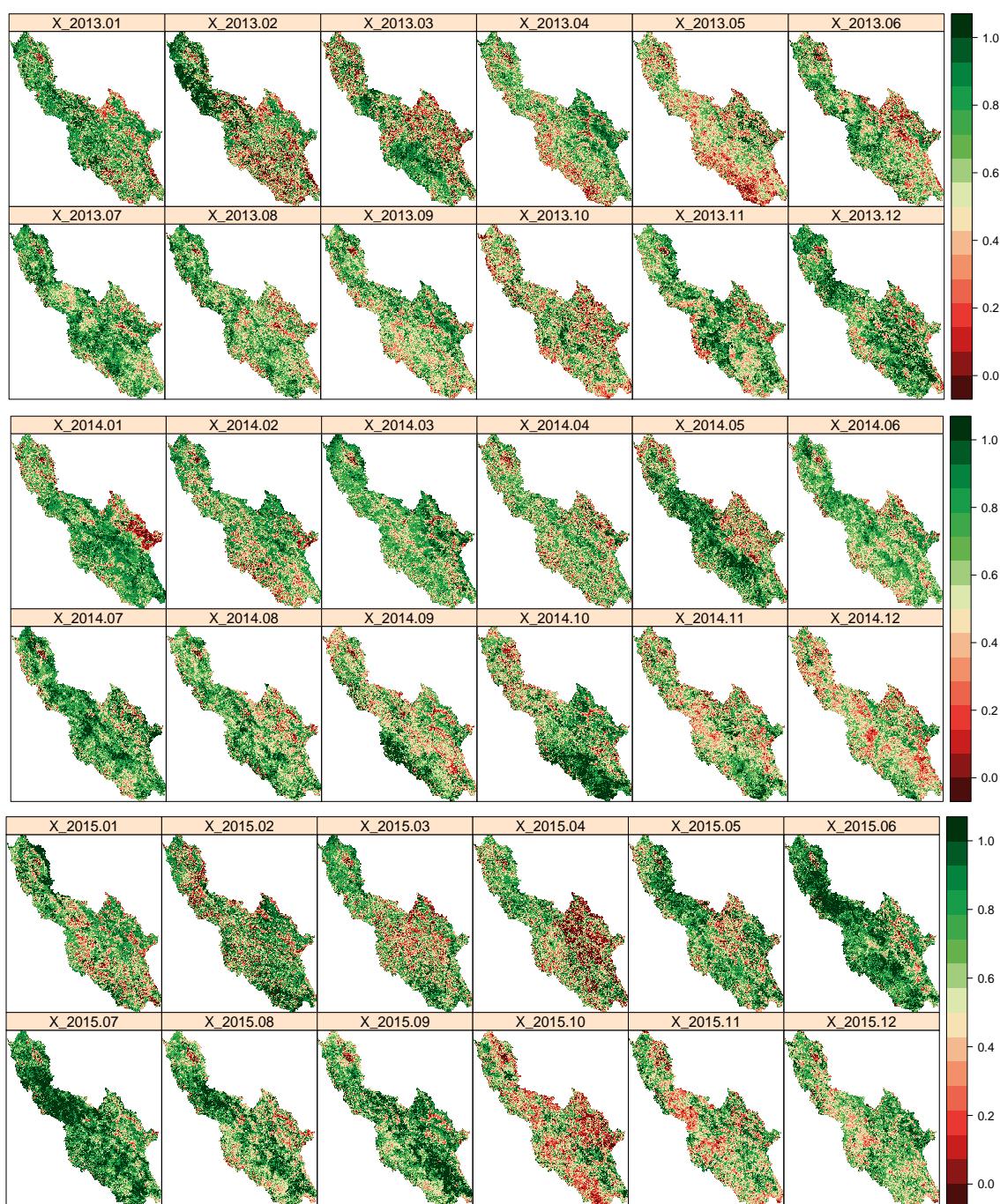
Fuente: Elaboración propia

Figura B.3: Índices de condición de Vegetación: 2008-2012



Fuente: Elaboración propia

Figura B.4: Índices de condición de Vegetación: 2013-2015



Fuente: Elaboración propia

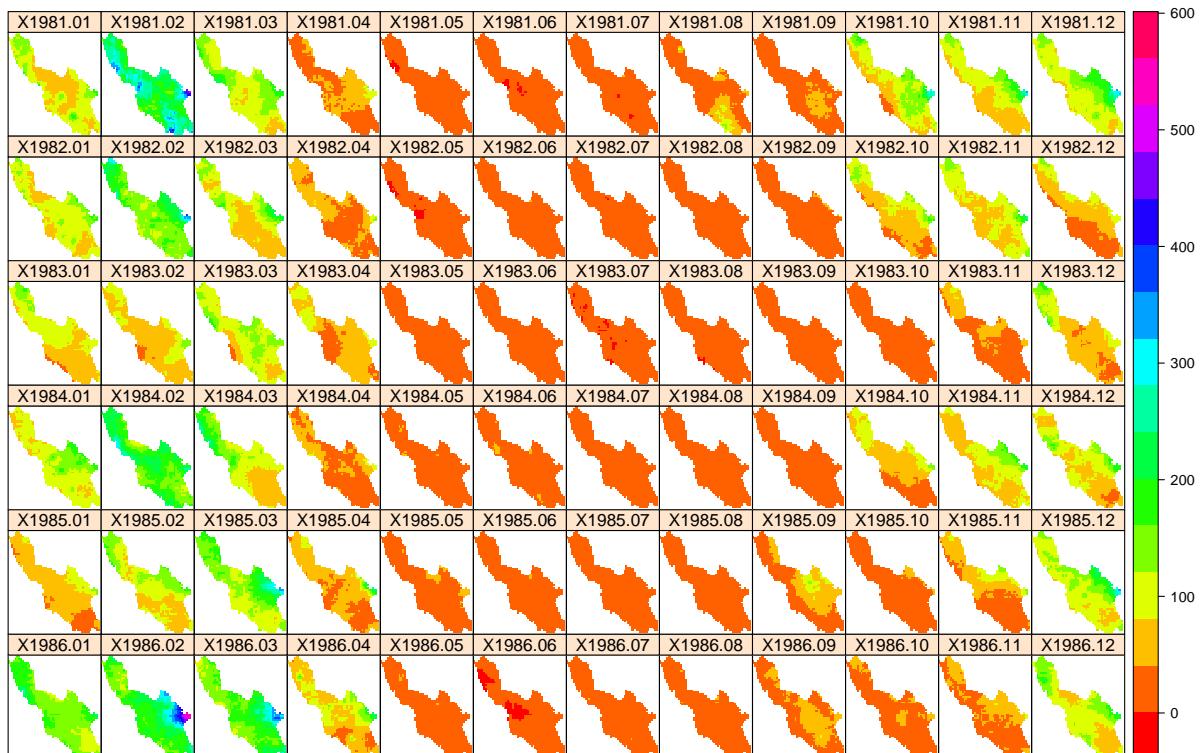
Anexo C

Precipitaciones Mensuales PISCO

Código Fuente C.1: Código fuente R para el análisis de la precipitación desde formato raster

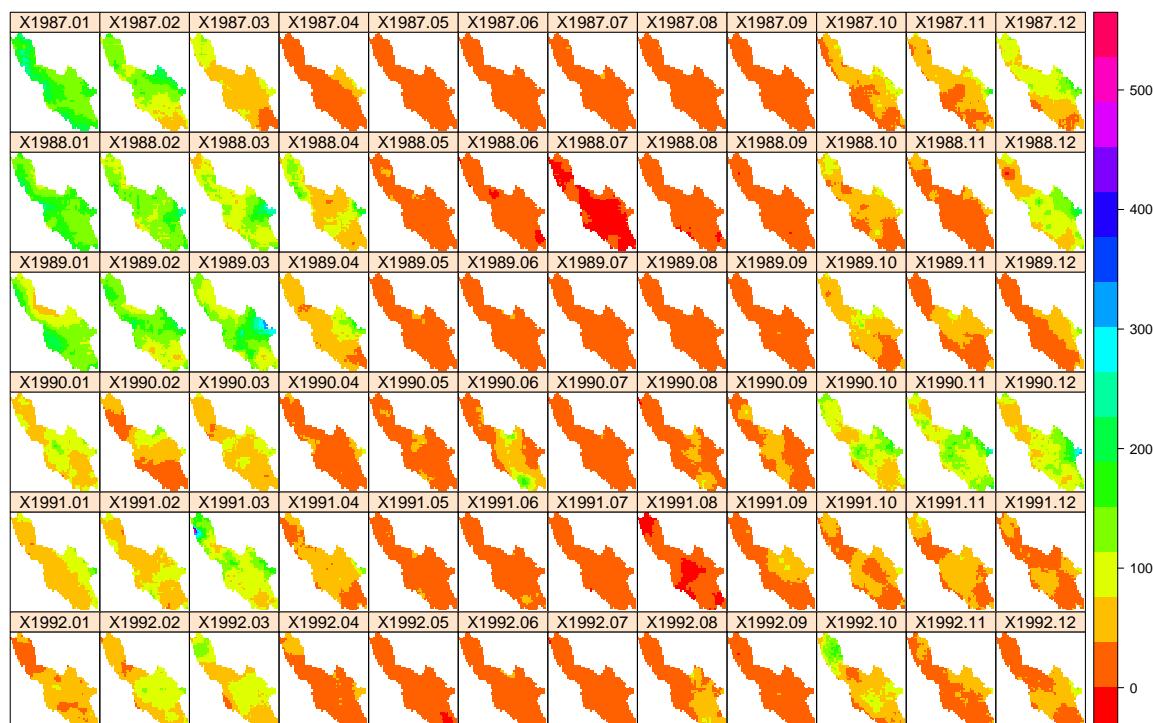
```
1 rtPlotLimit <- function(n_cuenca,iv=FALSE, dimplot = c(3,4), color =c('green','←
  blue'), zlim=c(-10,2000), dimname=c(17,23)){←
2 #-----←
3 # AUTOR      : Ivan Arturo Ayala Bizarro←
4 # PROPOSITO   : Plotea exporta como series de tiempo formatos rasters←
5 #           con limite poligonal denominado cuenca←
6 #-----←
7 cuenca<-readOGR('. ',n_cuenca)←
8 cuenca<- spTransform(cuenca, CRS('+proj=longlat +datum=WGS84'))←
9 files<- list.files(pattern='.tif') # Extraer listado de archivos←
10 imageraster <- raster(files[1])←
11 imageraster <- crop(imageraster, cuenca)←
12 imageraster <- mask(imageraster, cuenca)←
13 names(imageraster) <- substr(files[1], dimname[1], dimname[2])←
14 cat(sprintf('Procesando raster %s \n',names(imageraster)))←
15 rbrick <- imageraster←
16 if(iv){←
17 vals<-extract(imageraster,cuenca) # extract(imageraster, cuenca)←
18 coord<-cellFromPolygon(imageraster,cuenca) # Tomar las coordenadas←
19 combine<-cbind(coord[[1]],vals[[1]]) # Combinar}←
20 if(length(files)>=2){←
21 for (i in 2:length(files)){←
22 imageraster <- raster(files[i])           # cargar imagen←
23 imageraster <- crop(imageraster, cuenca)  # cortar para la cuenca←
24 imageraster <- mask(imageraster, cuenca)  # hacer mascara para la cuenca←
25 names(imageraster) <- substr(files[i], dimname[1], dimname[2])←
26 rbrick <- addLayer(rbrick,imageraster) # acumular raster←
27 if(iv){←
28 vals<-extract(imageraster,cuenca) # extract(imageraster, cuenca)←
29 combine<-cbind(combine,vals[[1]])}←
30 cat(sprintf('Procesando raster %s \n',names(imageraster))))}←
31 if(iv){←
32 write.table(combine,paste('Poligono_',n_cuenca,'.txt'))←
33 cat('Los valores se imprimieron correctamente') }←
34 spplot(rbrick,layout=dimplot, zlim = zlim, col.regions=colorRampPalette(color)←
  (255))}
```

Figura C.1: Precipitaciones mensuales PISCO (mm) 1981-01 al 1986-12



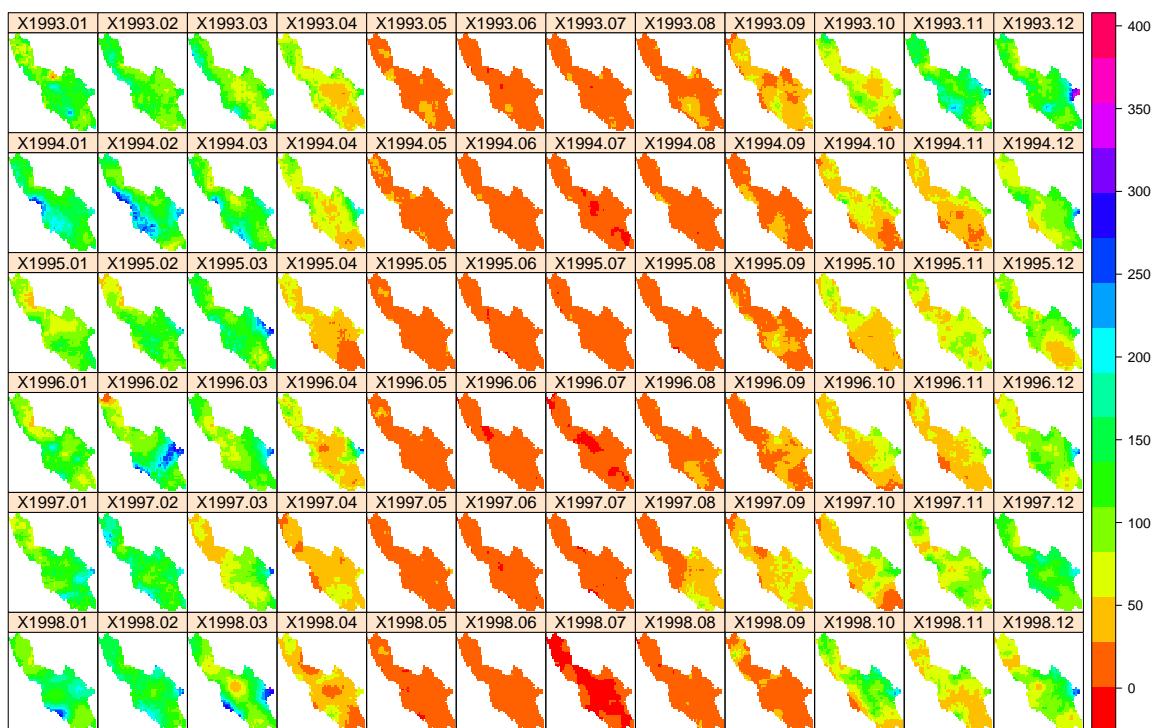
Fuente: Elaboración propia realizada con R

Figura C.2: Precipitaciones mensuales PISCO (mm) 1987-01 al 1992-12



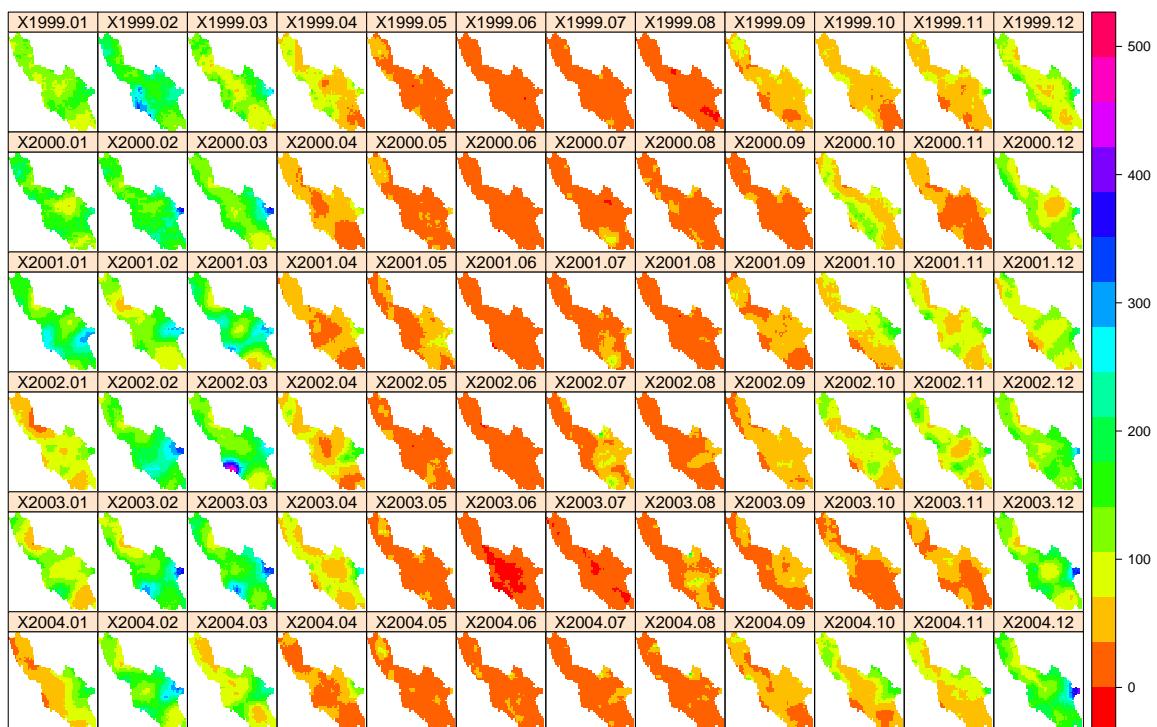
Fuente: Elaboración propia realizada con R

Figura C.3: Precipitaciones mensuales PISCO (mm) 1993-01 al 1998-12



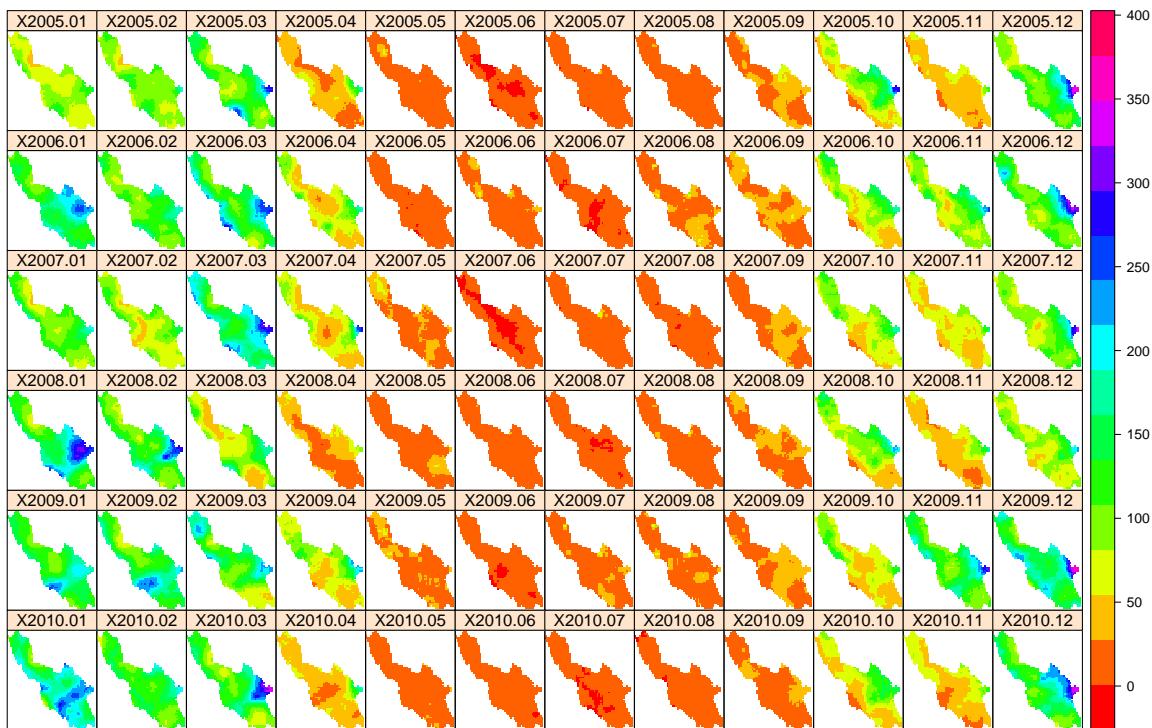
Fuente: Elaboración propia realizada con R

Figura C.4: Precipitaciones mensuales PISCO (mm) 1999-01 al 2004-12



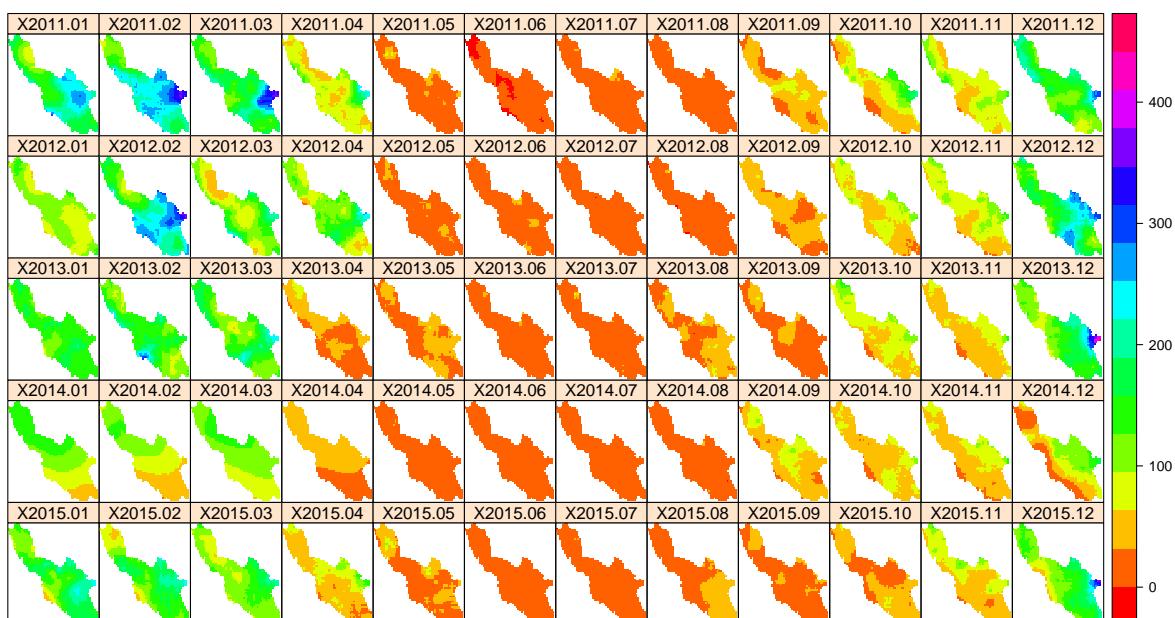
Fuente: Elaboración propia realizada con R

Figura C.5: Precipitaciones mensuales PISCO (mm) 2005-01 al 2010-12



Fuente: Elaboración propia realizada con R

Figura C.6: Precipitaciones mensuales PISCO (mm) 2011-01 al 2015-12



Fuente: Elaboración propia realizada con R

los raster presentan una resolución de 5kmx5km

Anexo D

Índices de Precipitación-Evapotranspiración Estandarizada

Código Fuente D.1: Rutinas R para obtener el SPEI

```
1 library(SPEI)
2 Pm <- read.table('Datos_Precipitaciones_Octubre.txt', header=T)
3 Tmin <- read.table('Datos_Tmin_Region6.txt', header=T)
4 Tmax <- read.table('Datos_Tmax_Region6.txt', header=T)
5 lat_pma <- read.table('Datos_Ubicacion.txt', header=T)
6 dat<-Pm[,1:2]
7 Pm<-Pm[,3:ncol(Pm)]
8 #Pmt<-Pmt[,3:ncol(Pmt)]
9 Tmim<-Tmin[,3:ncol(Tmin)]
10 Tmax<-Tmax[,3:ncol(Tmax)]
11 summary(Pm)
12 # PET Hargreaves
13 PET_h<-c()
14 for(i in 1:ncol(Pm)){
15   h <- hargreaves(Tmim[,i],Tmax[,i],lat=lat_pma$Lat[i] )
16   h<-as.vector(t(h));PET_h<-cbind(PET_h,h});result<-cbind(dat,PET_h)
17 write.table(result,paste('0_Hargreaves.txt'))
18 ## SPEI
19 PET <- read.table("Datos_PET.txt", header=TRUE)
20 PET<-PET[,3:ncol(PET)]
21 escala <- 6
22 SPEI_list<-c()
23 for(i in 1:ncol(Pm)){
24   P <- Pm[,i];Pe <- PET[,i];spei_c <- spei(P-Pe,escala)
25   valores<-spei_c$fitted
26   SPEI_list<-cbind(SPEI_list,valores)
27 # Plot SPEI
28 titleSPEI <- paste("SPEI_",as.character(escala),"_",colnames(Pm[i]), sep="")
29 pdf(paste(titleSPEI, ".pdf"),width=7,height=4)
30 plot(spei(ts(P-Pe,freq=12,start=c(1981,1)),escala))
31 dev.off()
32 write.table(SPEI_list,paste('0_SPEI',escala,'.txt'))
```



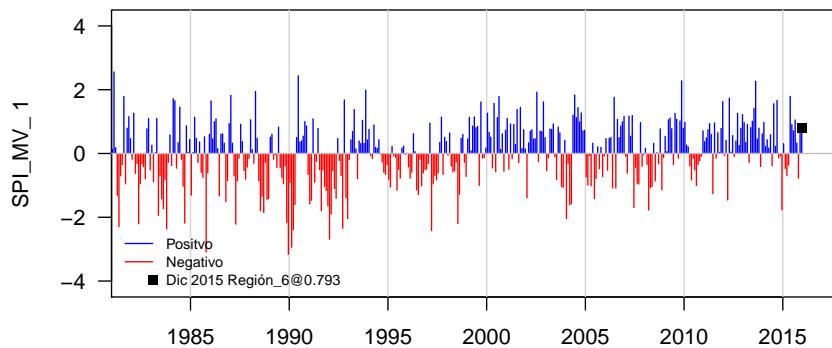
Código Fuente D.2: Rutinas R para obtener el SPI MV y AG

```
1 library(RHydrology)
2 ##### Importar datos desde archivos
3 Pmes <- read.table("Datos_Precipitaciones_Region6.txt", header=TRUE)
4 boxplot(Pmes[,3:ncol(Pmes)], col="cyan")
5 summary(Pmes)
6 ### SPIs Algoritmos Geneeticos
7 spiAG(Pmes, scale =6, poblacion=1000, maxIter = 100, plotAG=F, plotFDA=F)
8
9 ### SPIs Maaxima Verosimilitud
10 spiMV(Pmes, scale =12)
```

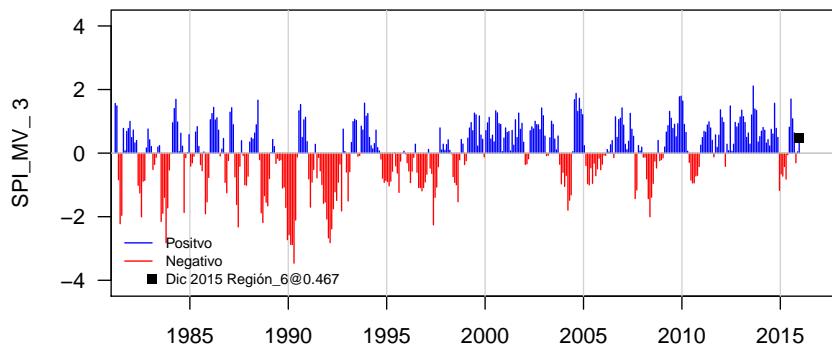
Código Fuente D.3: Optimización de parámetros con Algoritmos Genéticos en R

```
1 fitnessGamma <- function(x1,x2,P){
2 #-----
3 # AUTOR : Ivan Arturo Ayala Bizarro
4 # PROPOSITO : Calcula la funcioon a minimizar
5 #
6 n <- length(P)
7 # Paraametros
8 alpha <- x1; beta <- x2
9 # Funcion de distribucioon acumulada
10 Gx <- sort(pgamma(P, shape = alpha, scale = beta)); Gxm <- mean(Gx)
11 # Funcion de coste para la optimizacioon
12 sum1<-0.; sum2<-0.
13 for(i in 1:n){
14 Px <- i/(n+1) # Probabilidad empiirica Weibull
15 #Px <- (i-0.44)/(n+12) # Probabilidad empiirica Gringorten
16 sum1 <- sum1+ (Px-Gx[i])**2;sum2 <- sum2+ (Px-Gxm)**2
17 fit <- sum1/sum2
18 return (fit)}
19 distEmperica <- function(x,m=1){
20 #
21 # AUTOR : Ivan Arturo Ayala Bizarro
22 # PROPOSITO : Calcula la distribucioon empiirica por el metodo elegido
23 #
24 n <- length(x);Px<-vector()
25 if (m == 2) { # Probabilidad empiirica Gringorten
26 for(i in 1:n) {Px[i]<- (i-0.44)/(n+12)}}
27 else if (m == 3) { # Probabilidad empiirica por implementar
28 for(i in 1:n) {Px[i] <- i/(n+1)}}
29 else { # Probabilidad empiirica Weibull
30 for(i in 1:n) {Px[i] <- i/(n+1)}}
31 return (Px)}
32 ##### Caalculo de los paraametros por Algoritmos Geeneticos
33 GA <- ga(type = "real-valued",
34 fitness = (function(x)-fitnessGamma(x[1], x[2],Pnn)),
35 min = c(0.1, 0.1), max = c(alpha+rango, beta+rango),
36 popSize = 500, maxiter = 50)
37 out <-summary(GA)
38 plot(GA)
39 print(out)
40 str(out)
```

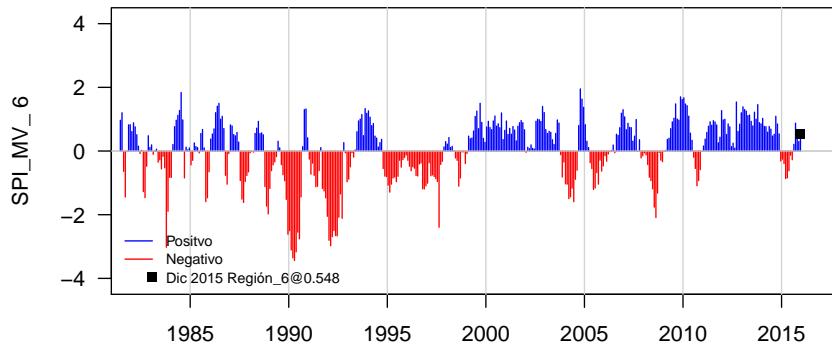
Figura D.1: Índices de precipitación y evapotranspiración estandarizada



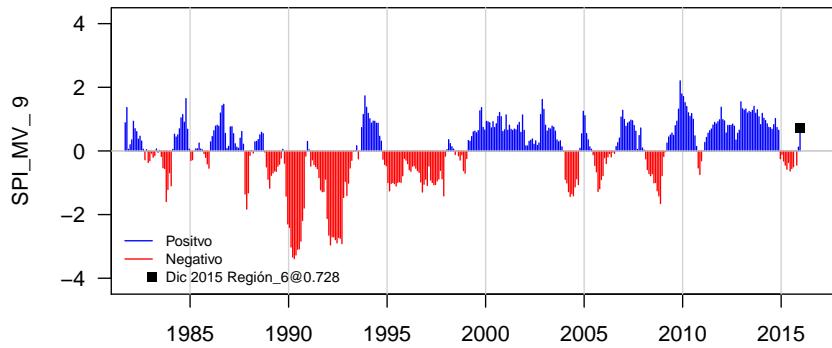
(a) SPI Máxima Verosimilitud, escala 1 mes



(b) SPI Máxima Verosimilitud, escala 3 meses

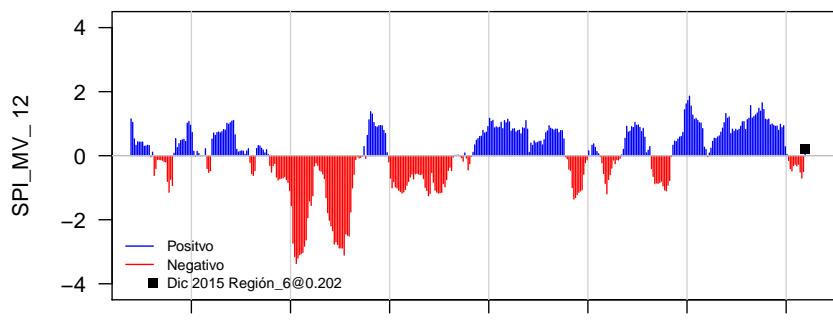


(c) SPI Máxima Verosimilitud, escala 6 meses

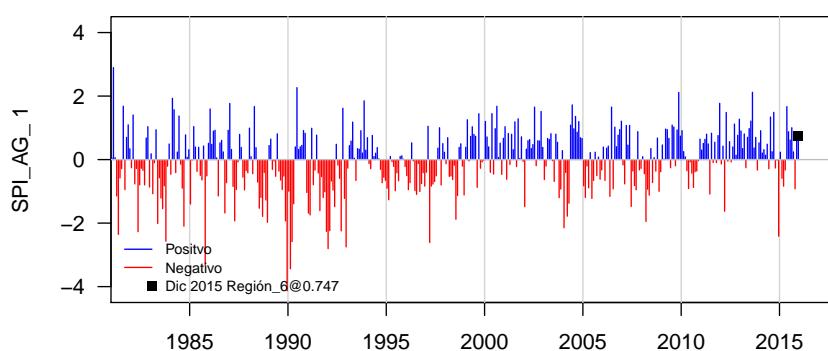


(d) SPI Máxima Verosimilitud, escala 9 meses

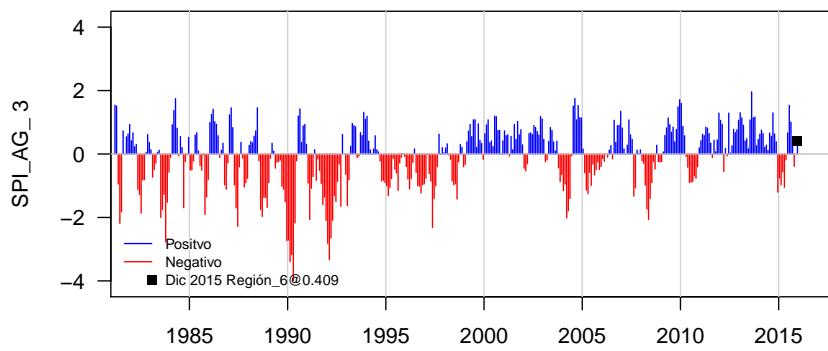
Fuente: Elaboración propia



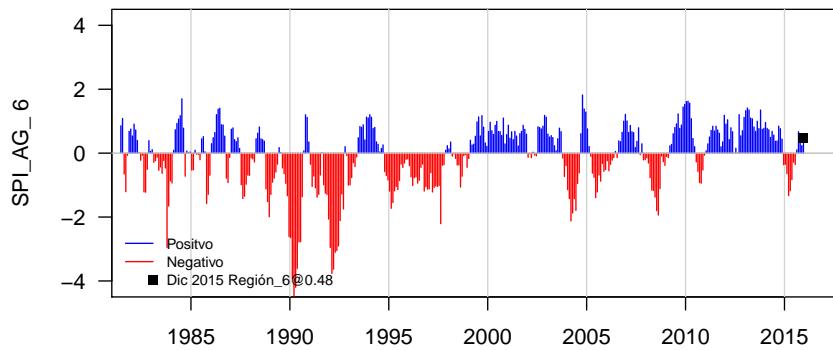
(e) SPI Máxima Verosimilitud, escala 12 meses



(f) SPI Algoritmos Genéticos, escala 1 mes

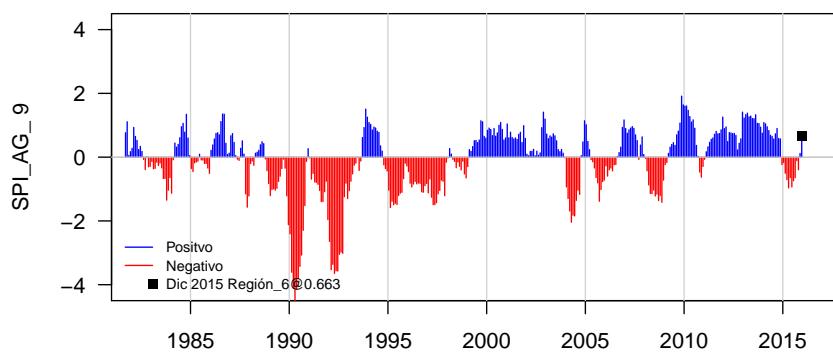


(g) SPI Algoritmos Genéticos, escala 3 meses

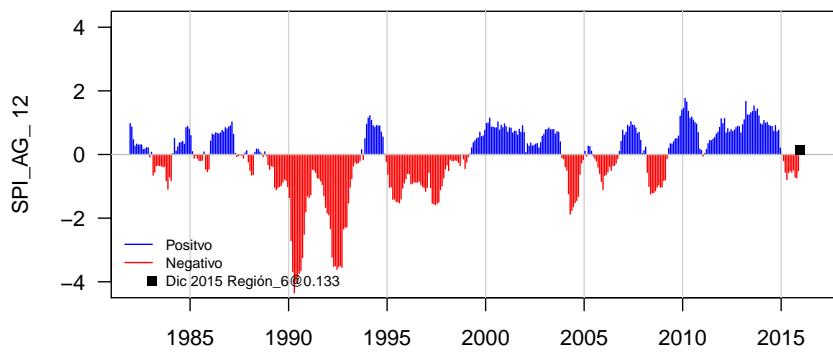


(h) SPI Algoritmos Genéticos, escala 6 meses

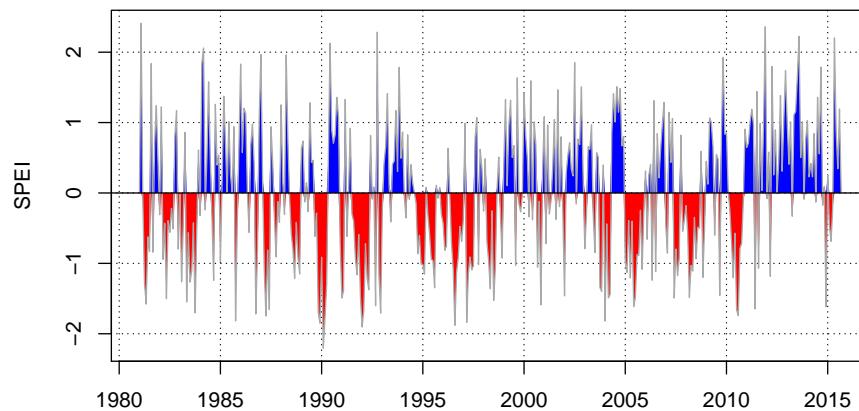
Fuente: Elaboración propia



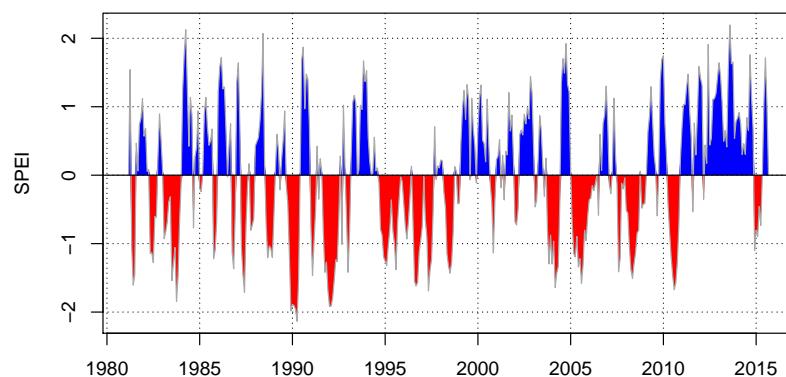
(i) SPI Algoritmos Genéticos, escala 9 meses



(j) SPI Algoritmos Genéticos, escala 12 meses

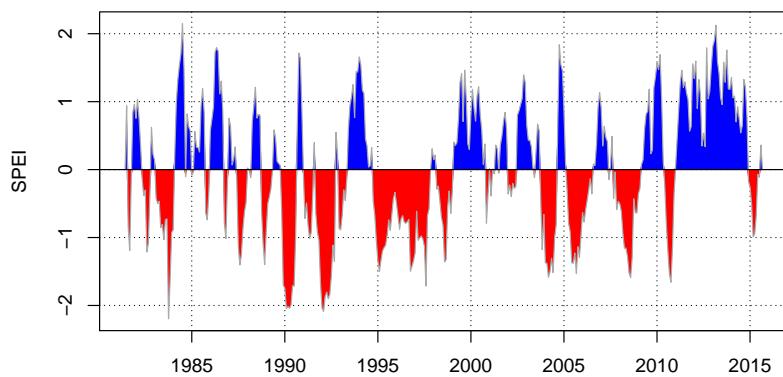


(k) SPEI, escala 1 mes

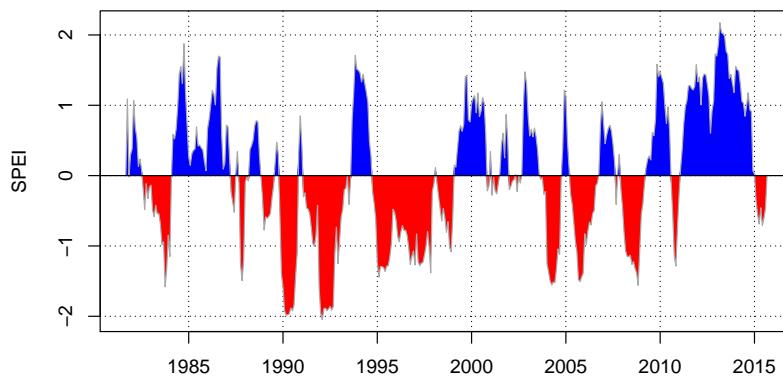


(l) SPEI, escala 3 meses

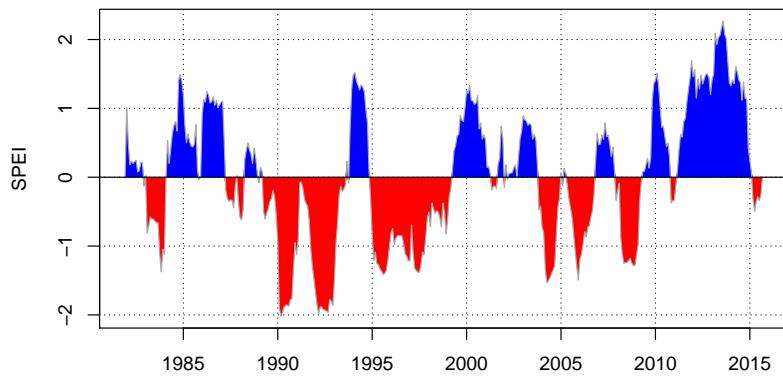
Fuente: Elaboración propia



(m) SPEI, escala 6 meses



(n) SPEI, escala 9 meses



(ñ) SPEI, escala 12 meses

Fuente: Elaboración propia

Anexo E

Resultados L Momentos

Tabla E.1: Resultados L-Momentos

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-1	-12.325	-74.1	35	801.98	0.103	0.017	0.186	-0.018
st-PER-2	-12.1	-74.413	35	624.944	0.147	-0.048	0.073	-0.051
st-PER-3	-12.046	-74.589	35	690.407	0.124	-0.038	0.098	-0.061
st-PER-4	-12.075	-74.409	35	630.741	0.13	-0.048	0.087	-0.054
st-PER-5	-12.324	-74.274	35	761.23	0.102	0.027	0.192	0.047
st-PER-6	-12.425	-74.174	35	751.715	0.112	0.084	0.122	-0.026
st-PER-7	-12.025	-74.775	35	660.493	0.122	-0.093	0.095	-0.043
st-PER-8	-12.1	-74.849	35	637.24	0.137	-0.011	0.098	-0.027
st-PER-9	-12.074	-75.313	35	666.387	0.117	-0.045	0.072	-0.072
st-PER-10	-11.986	-75.515	35	695.481	0.096	-0.079	0.177	0.013
st-PER-11	-11.645	-75.815	35	744.864	0.136	0.004	0.141	0.034
st-PER-12	-11.675	-75.725	35	648.462	0.108	-0.02	0.139	-0.028
st-PER-13	-11.204	-75.925	35	601.834	0.099	-0.006	0.006	0.032
st-PER-14	-10.824	-76.575	35	631.882	0.115	-0.026	0.122	-0.045
st-PER-15	-10.805	-76.426	35	725.389	0.108	-0.068	0.111	-0.032
st-PER-16	-12.237	-74.237	35	953.618	0.111	-0.072	0.176	-0.045
st-PER-17	-12.265	-74.305	35	934.218	0.108	-0.051	0.189	-0.053
st-PER-18	-12.283	-74.725	35	883.87	0.103	-0.049	0.19	-0.032
st-PER-19	-12.087	-74.8	35	664.025	0.128	-0.031	0.082	-0.048
st-PER-20	-12.075	-74.625	35	681.951	0.119	-0.027	0.093	-0.04
st-PER-21	-11.875	-74.615	35	712.155	0.094	-0.125	0.187	-0.008
st-PER-22	-12.324	-74.826	35	788.919	0.101	0.013	0.183	-0.02
st-PER-23	-12.081	-75.257	35	660.23	0.12	-0.025	0.087	-0.038
st-PER-24	-11.92	-75.352	35	752.414	0.109	-0.101	0.173	-0.031
st-PER-25	-11.641	-75.758	35	733.048	0.13	-0.004	0.136	0.021
st-PER-26	-11.786	-75.686	35	686.014	0.095	-0.096	0.197	-0.053
st-PER-27	-11.531	-75.9	35	558.626	0.123	-0.018	0.057	-0.021
st-PER-28	-11.025	-76.001	35	682.201	0.095	-0.036	0.022	0.002
st-PER-29	-10.9	-76.225	35	754.026	0.101	-0.081	0.068	-0.031
st-PER-30	-10.815	-76.336	35	766.531	0.105	-0.089	0.089	-0.039
st-PER-31	-12.401	-74.2	35	742.483	0.108	0.091	0.144	-0.033
st-PER-32	-12.401	-74.375	35	697.703	0.1	0.106	0.163	0.064

(continúa en la página siguiente)



Tabla E.1: Continuación de la tabla

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-33	-12.475	-74.375	35	899.154	0.133	0.017	0.103	-0.058
st-PER-34	-12.299	-74.8	35	857.682	0.108	-0.002	0.19	-0.037
st-PER-35	-12.424	-74.476	35	1091.252	0.135	-0.019	0.118	-0.066
st-PER-36	-11.926	-74.675	35	719.882	0.095	0.024	0.216	0.096
st-PER-37	-12.2	-74.438	35	711.127	0.127	-0.003	0.103	-0.017
st-PER-38	-12.475	-74.658	35	951.264	0.131	0.013	0.12	-0.058
st-PER-39	-12.408	-74.919	35	806.409	0.113	0.045	0.162	-0.05
st-PER-40	-12.025	-75.209	35	712.197	0.123	-0.066	0.108	-0.098
st-PER-41	-11.765	-75.515	35	658.383	0.098	-0.082	0.185	-0.047
st-PER-42	-11.663	-76	35	574.809	0.115	-0.045	0.099	-0.031
st-PER-43	-11.44	-75.94	35	576.461	0.117	-0.015	0.047	0.021
st-PER-44	-11.031	-76.08	35	730.11	0.096	-0.031	0.05	-0.031
st-PER-45	-10.734	-76.248	35	821.373	0.102	-0.091	0.074	-0.035
st-PER-46	-12.425	-74.224	35	742.7	0.106	0.089	0.104	-0.017
st-PER-47	-12.576	-74.725	35	1123.614	0.151	-0.052	0.079	0.009
st-PER-48	-12.45	-74.476	35	1196.811	0.136	-0.07	0.113	-0.064
st-PER-49	-12.376	-74.599	35	1114.578	0.124	-0.044	0.155	-0.076
st-PER-50	-12.324	-74.624	35	855.587	0.114	0.001	0.179	-0.035
st-PER-51	-11.676	-74.9	35	674.659	0.115	-0.018	0.14	-0.012
st-PER-52	-11.792	-74.809	35	657.815	0.092	-0.078	0.194	-0.057
st-PER-53	-11.899	-74.826	35	540.052	0.113	-0.064	0.108	-0.121
st-PER-54	-11.735	-74.835	35	772.359	0.123	-0.056	0.145	0.019
st-PER-55	-12.119	-75.443	35	712.1	0.137	-0.034	0.096	-0.074
st-PER-56	-11.763	-75.487	35	672.914	0.102	-0.08	0.196	-0.04
st-PER-57	-11.354	-75.983	35	606.811	0.109	-0.023	0.039	0.001
st-PER-58	-10.824	-76.065	35	736.736	0.094	-0.055	0.081	-0.014
st-PER-59	-11.025	-75.976	35	780.918	0.101	0.041	-0.017	-0.032
st-PER-60	-10.994	-76.275	35	742.999	0.106	-0.052	0.061	-0.039
st-PER-61	-13.1	-74.233	35	562.719	0.129	0.027	0.077	0.006
st-PER-62	-12.754	-74.39	35	789.297	0.148	0.026	0.07	0.031
st-PER-63	-12.235	-74.895	35	952.673	0.114	-0.077	0.148	-0.034
st-PER-64	-12.425	-74.574	35	1401.824	0.128	-0.114	0.123	-0.041
st-PER-65	-12.475	-74.224	35	1025.104	0.152	0.015	0.083	-0.02
st-PER-66	-11.674	-74.7	35	534.255	0.113	-0.036	0.075	-0.035
st-PER-67	-11.674	-74.826	35	848.004	0.179	0.019	0.114	-0.004
st-PER-68	-12.176	-74.876	35	516.51	0.192	-0.03	0.054	-0.006
st-PER-69	-12.15	-75.324	35	575.947	0.159	-0.065	0.061	-0.045
st-PER-70	-12.125	-75.525	35	679.25	0.123	-0.028	0.073	-0.045
st-PER-71	-12.176	-75.276	35	578.639	0.18	-0.022	0.062	0.001
st-PER-72	-11.841	-75.976	35	659.764	0.1	-0.056	0.206	-0.035
st-PER-73	-11.625	-76.099	35	580.199	0.121	-0.024	0.078	-0.034
st-PER-74	-11.183	-76.075	35	702.48	0.099	-0.049	0.03	-0.032
st-PER-75	-10.614	-76.376	35	770.59	0.11	-0.065	0.126	-0.046
st-PER-76	-12.987	-74.312	35	595.233	0.137	0.053	0.077	0.011
st-PER-77	-12.655	-74.435	35	834.598	0.143	0.025	0.077	0.023
st-PER-78	-12.524	-74.35	35	1307.708	0.137	-0.119	0.09	-0.021
st-PER-79	-12.475	-74.174	35	1102.124	0.15	-0.008	0.07	-0.022

(continúa en la página siguiente)



Tabla E.1: Continuación de la tabla

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-80	-12.576	-74.675	35	1227.149	0.141	-0.097	0.104	0.012
st-PER-81	-11.976	-74.858	35	760.06	0.125	-0.117	0.132	-0.08
st-PER-82	-11.926	-74.476	35	748.699	0.116	-0.105	0.164	-0.052
st-PER-83	-11.976	-74.912	35	694.516	0.098	-0.065	0.137	0
st-PER-84	-12.085	-74.895	35	634.797	0.13	-0.061	0.095	-0.065
st-PER-85	-12.311	-75.325	35	890.954	0.107	-0.032	0.195	-0.044
st-PER-86	-11.674	-75.425	35	837.561	0.116	-0.052	0.138	0.017
st-PER-87	-11.724	-76.063	35	684.31	0.119	-0.031	0.134	-0.003
st-PER-88	-11.16	-75.976	35	647.144	0.098	-0.042	0.004	0.016
st-PER-89	-10.949	-76.469	35	704.376	0.109	-0.043	0.073	-0.034
st-PER-90	-10.815	-76.585	35	634.713	0.116	-0.019	0.111	-0.039
st-PER-91	-12.775	-74.342	35	725.308	0.138	0.037	0.073	0.016
st-PER-92	-13.026	-74.451	35	598.373	0.138	0.069	0.064	-0.003
st-PER-93	-12.776	-74.675	35	749.61	0.144	-0.008	0.062	0.065
st-PER-94	-12.191	-74.325	35	659.955	0.111	-0.013	0.125	-0.038
st-PER-95	-12.299	-74.575	35	946.135	0.11	-0.026	0.179	-0.06
st-PER-96	-12.675	-74.599	35	940.224	0.153	-0.021	0.07	0.088
st-PER-97	-12.408	-74.974	35	752.375	0.102	0.041	0.187	0.011
st-PER-98	-12.353	-75.03	35	740.928	0.101	0.028	0.179	-0.021
st-PER-99	-12.025	-74.641	35	691.755	0.109	-0.051	0.103	-0.037
st-PER-100	-11.91	-75.115	35	746.473	0.117	-0.068	0.168	-0.038
st-PER-101	-11.758	-75.525	35	803.816	0.156	0.011	0.145	-0.013
st-PER-102	-11.95	-75.65	35	706.469	0.094	-0.11	0.182	-0.004
st-PER-103	-11.825	-76.074	35	753.017	0.097	-0.117	0.119	0.012
st-PER-104	-11.751	-75.813	35	618.116	0.105	-0.047	0.18	-0.076
st-PER-105	-10.818	-76.439	35	719.956	0.107	-0.073	0.107	-0.045
st-PER-106	-13.1	-74.265	35	576.696	0.131	0.02	0.07	0.001
st-PER-107	-12.975	-74.318	35	588.045	0.14	0.072	0.078	0.015
st-PER-108	-12.399	-74.425	35	781.392	0.112	0.027	0.167	-0.029
st-PER-109	-12.508	-74.734	35	742.925	0.115	0.075	0.125	-0.04
st-PER-110	-12.25	-74.599	35	901.785	0.107	-0.04	0.178	-0.031
st-PER-111	-12.425	-74.955	35	927.037	0.122	-0.004	0.137	-0.049
st-PER-112	-12.55	-74.3	35	940.343	0.155	0.02	0.074	0.017
st-PER-113	-12.425	-74.337	35	722.88	0.103	0.087	0.167	0.008
st-PER-114	-12.012	-74.987	35	691.439	0.106	-0.062	0.093	-0.04
st-PER-115	-11.885	-75.485	35	805.715	0.11	-0.097	0.144	0.012
st-PER-116	-11.608	-75.658	35	698.047	0.129	0.001	0.134	0.021
st-PER-117	-11.92	-75.8	35	692.45	0.093	-0.092	0.19	-0.017
st-PER-118	-11.666	-75.956	35	621.668	0.11	-0.041	0.141	-0.034
st-PER-119	-11.063	-76.42	35	726.678	0.108	-0.02	0.07	-0.024
st-PER-120	-11.151	-76.413	35	738.379	0.108	-0.026	0.092	-0.04
st-PER-121	-13.216	-74.076	35	609.421	0.119	-0.019	0.099	-0.026
st-PER-122	-12.95	-74.513	35	649.349	0.143	0.061	0.059	0.017
st-PER-123	-12.753	-74.575	35	792.68	0.152	0.015	0.062	0.051
st-PER-124	-12.675	-74.55	35	766.019	0.13	0.016	0.114	0.017
st-PER-125	-12.625	-74.425	35	932.473	0.172	0.043	0.037	0.031
st-PER-126	-12.425	-74.649	35	1160.084	0.136	-0.05	0.119	-0.043

(continúa en la página siguiente)



Tabla E.1: Continuación de la tabla

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-127	-12.225	-74.826	35	883.76	0.115	-0.066	0.156	-0.042
st-PER-128	-12.658	-75.243	35	803.28	0.139	0.059	0.066	-0.015
st-PER-129	-11.989	-74.675	35	708.725	0.11	-0.069	0.12	-0.066
st-PER-130	-11.892	-75.041	35	697.214	0.102	-0.099	0.178	-0.048
st-PER-131	-11.751	-75.325	35	633.502	0.108	-0.042	0.198	-0.064
st-PER-132	-12.033	-75.692	35	687.878	0.109	-0.058	0.111	-0.032
st-PER-133	-11.848	-75.948	35	755.596	0.128	-0.031	0.176	-0.009
st-PER-134	-11.551	-76.124	35	713.998	0.124	-0.014	0.125	0.002
st-PER-135	-11.118	-76.196	35	753.355	0.105	-0.027	0.039	-0.031
st-PER-136	-12.815	-74.506	35	699.436	0.134	0.045	0.068	0.008
st-PER-137	-12.85	-74.375	35	681.803	0.148	0.05	0.067	0.03
st-PER-138	-12.685	-74.264	35	780.543	0.139	0.022	0.073	0.039
st-PER-139	-12.455	-74.595	35	708.444	0.105	0.02	0.167	-0.004
st-PER-140	-12.475	-75.025	35	724.728	0.105	0.07	0.16	-0.021
st-PER-141	-12.599	-74.712	35	900.264	0.157	0.019	0.063	0.041
st-PER-142	-12.675	-74.887	35	764.621	0.128	0.04	0.107	0.008
st-PER-143	-11.859	-74.675	35	677.366	0.095	-0.061	0.179	-0.049
st-PER-144	-11.888	-74.737	35	740.912	0.104	-0.115	0.185	-0.023
st-PER-145	-12.041	-75.05	35	683.288	0.109	-0.03	0.094	-0.03
st-PER-146	-12.124	-75.374	35	583.192	0.144	-0.091	0.084	-0.073
st-PER-147	-12.192	-75.517	35	695.159	0.158	0.002	0.077	-0.027
st-PER-148	-11.709	-76.107	35	615.851	0.109	-0.057	0.144	-0.02
st-PER-149	-11.444	-76.068	35	656.51	0.117	-0.042	0.063	-0.036
st-PER-150	-11.225	-76.174	35	738.985	0.106	-0.028	0.033	-0.025
st-PER-151	-13.219	-74.044	35	642.66	0.119	-0.022	0.088	-0.023
st-PER-152	-12.925	-74.15	35	647.726	0.141	0.012	0.094	0.056
st-PER-153	-12.915	-74.165	35	656.622	0.139	0.002	0.088	0.05
st-PER-154	-12.945	-74.705	35	725.841	0.137	0.019	0.08	0.012
st-PER-155	-12.757	-74.843	35	774.85	0.144	0.016	0.071	0.043
st-PER-156	-12.818	-74.896	35	801.415	0.16	0.041	0.084	0.007
st-PER-157	-12.689	-75.125	35	823.343	0.145	0.035	0.067	0.023
st-PER-158	-12.575	-75.191	35	711.059	0.101	0.057	0.155	-0.005
st-PER-159	-12.724	-75.45	35	813.392	0.155	0.031	0.063	0.058
st-PER-160	-12.625	-75.175	35	1014.255	0.155	-0.036	0.079	0.053
st-PER-161	-12.425	-75.276	35	691.983	0.096	0.088	0.186	0.062
st-PER-162	-12.374	-75.525	35	817.55	0.112	0.036	0.153	-0.055
st-PER-163	-11.976	-75.976	35	746.286	0.104	-0.154	0.164	-0.028
st-PER-164	-11.559	-76.208	35	742.391	0.127	-0.011	0.165	0.01
st-PER-165	-11.359	-76.158	35	712.764	0.113	-0.042	0.05	-0.017
st-PER-166	-13.275	-74.025	35	660.882	0.119	-0.024	0.107	-0.034
st-PER-167	-13.338	-74.15	35	618.208	0.119	-0.024	0.098	-0.026
st-PER-168	-13.05	-74.583	35	645.995	0.136	0.032	0.056	0.017
st-PER-169	-12.967	-74.784	35	763.829	0.144	-0.02	0.055	0.027
st-PER-170	-12.15	-74.75	35	627.092	0.162	-0.048	0.057	-0.035
st-PER-171	-12.925	-74.901	35	812.782	0.161	-0.005	0.054	0.026
st-PER-172	-11.718	-74.789	35	679.121	0.102	-0.052	0.174	-0.038
st-PER-173	-12.604	-74.91	35	910.893	0.16	0.037	0.052	0.004

(continúa en la página siguiente)



Tabla E.1: Continuación de la tabla

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-174	-12.225	-74.575	35	933.753	0.116	-0.064	0.143	-0.028
st-PER-175	-12.625	-75.226	35	975.646	0.162	0.006	0.067	0.04
st-PER-176	-12.325	-75.496	35	961.918	0.113	-0.023	0.171	-0.054
st-PER-177	-12.376	-75.324	35	1200.738	0.118	-0.076	0.162	-0.091
st-PER-178	-12.524	-75.525	35	1173.143	0.137	-0.05	0.109	-0.056
st-PER-179	-11.465	-76.225	35	730.708	0.12	-0.048	0.101	-0.009
st-PER-180	-11.225	-76.255	35	775.366	0.111	-0.013	0.038	-0.036
st-PER-181	-13.38	-74.339	35	576.269	0.125	-0.015	0.102	-0.004
st-PER-182	-13.301	-74.3	35	578.105	0.124	-0.002	0.08	-0.02
st-PER-183	-13.266	-74.345	35	562.744	0.129	0.034	0.079	-0.02
st-PER-184	-13.084	-74.641	35	659.437	0.137	0.009	0.056	0.028
st-PER-185	-12.875	-75.037	35	773.001	0.152	-0.048	0.048	0.035
st-PER-186	-12.75	-75.15	35	827.138	0.148	0.016	0.067	0.052
st-PER-187	-12.858	-75.274	35	708.622	0.147	0.012	0.068	0.057
st-PER-188	-12.993	-74.809	35	759.599	0.141	-0.039	0.055	0.021
st-PER-189	-12.617	-75.391	35	891.748	0.163	0.046	0.057	0.008
st-PER-190	-12.625	-75.324	35	899.322	0.167	0.04	0.051	0.033
st-PER-191	-12.256	-75.119	35	957.765	0.112	-0.079	0.161	-0.037
st-PER-192	-12.624	-75.525	35	1046.492	0.154	-0.026	0.062	0.035
st-PER-193	-12.458	-75.441	35	999.114	0.135	0.02	0.133	-0.065
st-PER-194	-11.255	-76.295	35	769.959	0.114	0.003	0.049	-0.029
st-PER-195	-13.325	-74.425	35	595.097	0.128	0.005	0.092	0.001
st-PER-196	-13.267	-74.509	35	611.652	0.132	0.001	0.075	0.008
st-PER-197	-13.426	-74.725	35	616.782	0.132	-0.089	0.176	0.072
st-PER-198	-13.099	-74.788	35	726.109	0.133	-0.066	0.053	0.01
st-PER-199	-13.211	-74.768	35	693.899	0.129	-0.103	0.079	0.046
st-PER-200	-13.033	-74.957	35	809.53	0.166	-0.058	-0.006	-0.007
st-PER-201	-12.858	-75.251	35	701.334	0.147	0.025	0.078	0.06
st-PER-202	-12.737	-75.447	35	759.397	0.13	0.029	0.115	0.011
st-PER-203	-12.492	-75.225	35	1119.607	0.136	-0.054	0.11	-0.044
st-PER-204	-12.526	-74.775	35	1026.435	0.142	-0.001	0.106	-0.034
st-PER-205	-12.475	-75.175	35	1250.706	0.14	-0.073	0.088	-0.05
st-PER-206	-12.475	-75.1	35	1368.21	0.131	-0.128	0.106	-0.04
st-PER-207	-12.524	-75.374	35	1371.535	0.131	-0.138	0.1	-0.045
st-PER-208	-11.451	-76.25	35	746.319	0.122	-0.036	0.097	-0.01
st-PER-209	-11.225	-76.412	35	741.395	0.113	0.005	0.104	-0.041
st-PER-210	-13.426	-74.537	35	611.972	0.134	-0.019	0.124	0.038
st-PER-211	-13.184	-74.575	35	626.137	0.134	0.013	0.055	-0.005
st-PER-212	-13.275	-74.652	35	648.043	0.129	-0.061	0.067	0.017
st-PER-213	-13.389	-74.675	35	634.831	0.133	-0.086	0.121	0.072
st-PER-214	-13.176	-74.924	35	728.906	0.141	-0.113	0.063	0.016
st-PER-215	-13.026	-75.037	35	813.87	0.181	-0.06	-0.017	0.007
st-PER-216	-12.933	-75.108	35	799.719	0.183	-0.013	0.024	0.008
st-PER-217	-12.611	-75.325	35	725.032	0.107	0.052	0.133	-0.02
st-PER-218	-12.475	-75.592	35	862.578	0.127	0.048	0.13	-0.076
st-PER-219	-12.55	-74.55	35	1170.318	0.142	-0.091	0.084	0.008
st-PER-220	-12.524	-74.249	35	1437.901	0.132	-0.159	0.091	-0.038

(continúa en la página siguiente)



Tabla E.1: Continuación de la tabla

id estacion	Latitud	Longitud	n	l_1	L_{CV}	$L_{Skewness}$	$L_{Kurtosis}$	t_5
st-PER-221	-12.475	-75.024	35	1439.792	0.125	-0.152	0.142	-0.034
st-PER-222	-12.475	-74.974	35	1474.804	0.124	-0.156	0.153	-0.031
st-PER-223	-11.384	-76.344	35	750.279	0.123	0.001	0.088	-0.016

Fuente: Elaboración Propia

Tabla E.2: Discordancia D_i asociado a las regiones homogéneas

No	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
1	0.244	0.732	0.66	0.904	0.817	0.205	0.601	0.7	1.313	0.6
2	0.792	0.582	0.944	1.63	0.308	1.936	0.442	0.866	0.496	0.442
3	0.481	1.133	1.172	0.925	0.48	0.793	0.475	0.416	0.282	0.781
4	2.134	2.146	1.5	0.258	1.007	0.998	1.542	0.24	0.046	0.709
5	0.585	0.876	0.088	1.116	0.44	0.946	0.435	1.959	0.665	0.563
6	0.88	2.291	0.631	1.273	0.655	1.038	0.544	0.314	0.91	0.356
7	2.025	0.745	1.161	1.648	0.746	0.68	0.621	0.922	0.877	1.576
8	2.274	0.952	0.902	0.485	0.611	0.784	1.146	2.012	1.075	1.397
9	0.365	1.446	1.693	0.968	2.207	1.602	1.324	2.748	0.876	0.716
10	0.635	1.049	1.375	0.042	1.104	0.578	1.243	0.44	1.11	0.806
11	0.585	0.226	0.86	0.813	1.225	0.987	2.055	0.931	0.924	2.578
12	-	1.16	0.715	0.578	0.874	1.842	0.821	0.727	2.026	1.561
13	-	0.7	1.653	0.348	1.132	0.705	0.514	0.85	1.082	1.139
14	-	1.298	0.647	0.386	1.281	1.074	2.238	0.519	1.677	0.79
15	-	0.235	-	1.224	1.684	1.235	-	0.657	1.642	1.574
16	-	1.658	-	1.796	1.286	0.452	-	0.6	-	0.063
17	-	1.439	-	1.51	1.212	0.785	-	0.645	-	1.348
18	-	0.803	-	1.585	0.917	1.362	-	1.066	-	-
19	-	0.052	-	0.91	1.33	-	-	1.065	-	-
20	-	0.927	-	1.691	1.375	-	-	1.477	-	-
21	-	0.813	-	0.777	1.39	-	-	1.847	-	-
22	-	0.421	-	0.373	2.212	-	-	-	-	-
23	-	1.051	-	1.788	0.192	-	-	-	-	-
24	-	1.375	-	1.069	0.626	-	-	-	-	-
25	-	0.992	-	0.905	0.998	-	-	-	-	-
26	-	1.008	-	-	1.362	-	-	-	-	-
27	-	0.907	-	-	0.06	-	-	-	-	-
28	-	0.707	-	-	0.467	-	-	-	-	-
29	-	0.209	-	-	-	-	-	-	-	-
30	-	1.534	-	-	-	-	-	-	-	-
31	-	0.765	-	-	-	-	-	-	-	-
32	-	0.291	-	-	-	-	-	-	-	-
33	-	0.59	-	-	-	-	-	-	-	-
34	-	1.898	-	-	-	-	-	-	-	-
35	-	0.679	-	-	-	-	-	-	-	-
36	-	2.447	-	-	-	-	-	-	-	-
37	-	0.746	-	-	-	-	-	-	-	-
38	-	1.508	-	-	-	-	-	-	-	-
39	-	0.553	-	-	-	-	-	-	-	-
40	-	1.509	-	-	-	-	-	-	-	-
41	-	2.034	-	-	-	-	-	-	-	-
42	-	0.736	-	-	-	-	-	-	-	-
43	-	1.765	-	-	-	-	-	-	-	-
44	-	0.062	-	-	-	-	-	-	-	-

(continúa en la página siguiente)

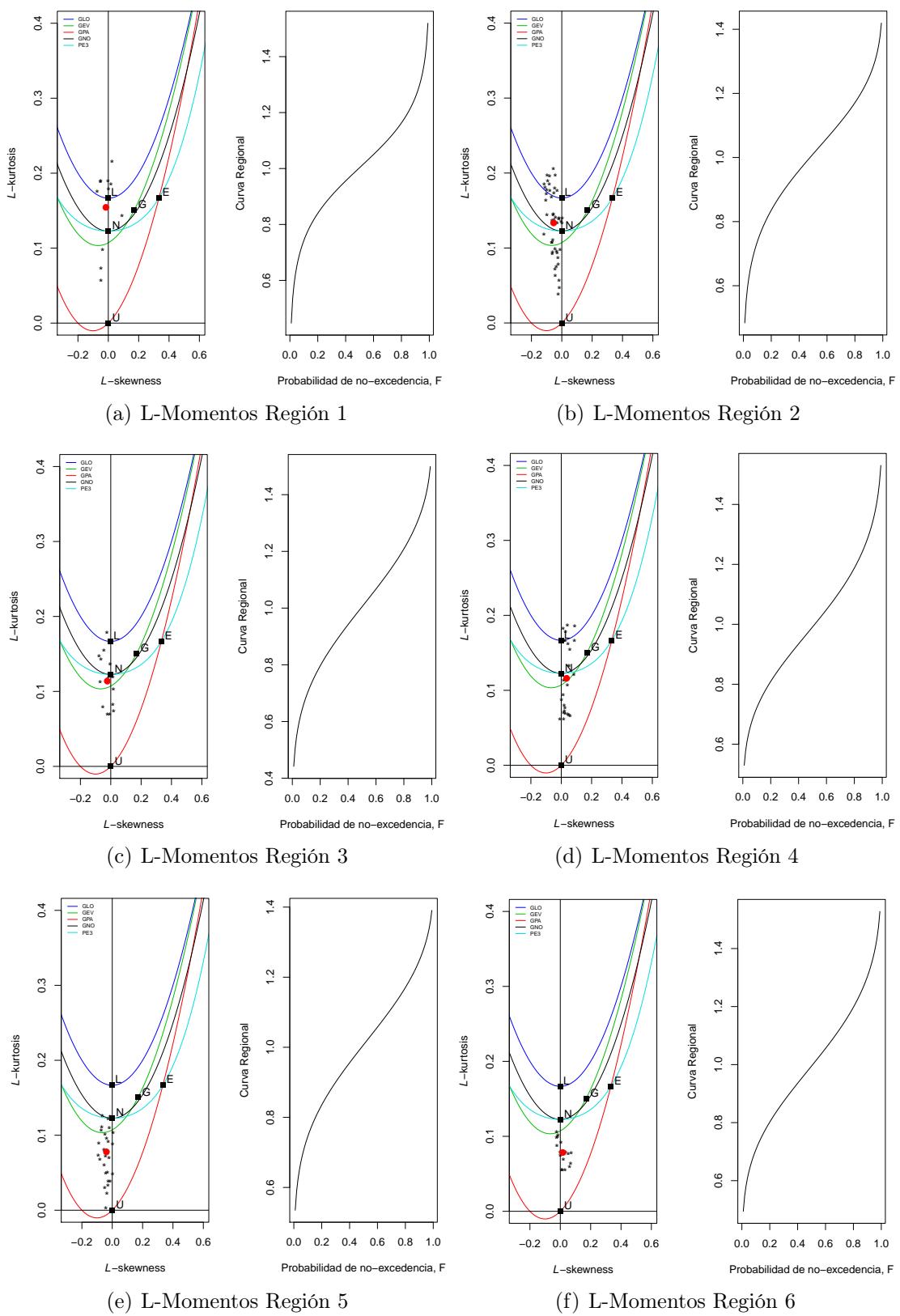


Tabla E.2: Continuación de la tabla

No	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
45	-	0.247	-	-	-	-	-	-	-	-
46	-	0.705	-	-	-	-	-	-	-	-
Mín	0.244	0.052	0.088	0.042	0.06	0.205	0.435	0.24	0.046	0.063
Máx	2.274	2.447	1.693	1.796	2.212	1.936	2.238	2.748	2.026	2.578

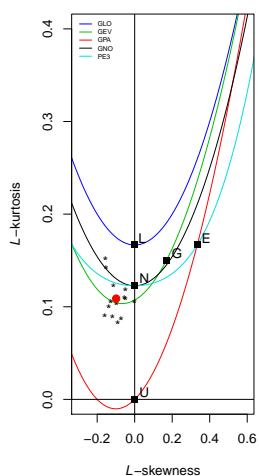
Fuente: Elaboración Propia

Figura E.1: L Momentos para las 10 regiones clasificadas

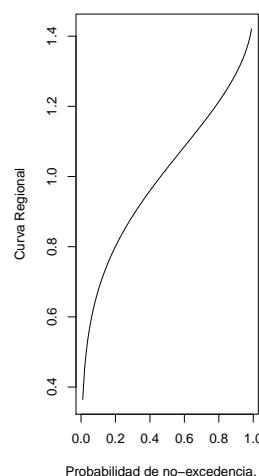


...continúa

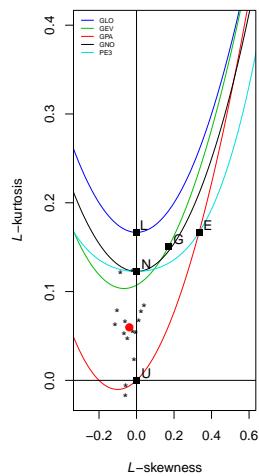
...Continúa figura E.1



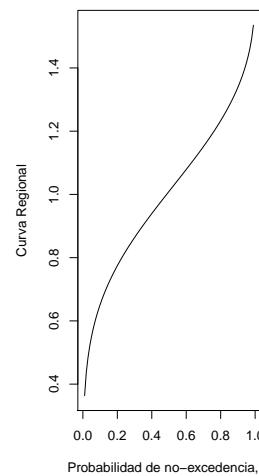
(g) L-Momentos Región 7



(h) L-Momentos Región 8



(i) L-Momentos Región 9



(j) L-Momentos Región 10

Fuente: Software ICI-RAFT