

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas



PROYECTO INTEGRADOR

“Modelación de una Red Neuronal Artificial para Predicción de Lluvias, usando Temperaturas Superficiales del Mar.”

Previo a la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

Javier Leonidas Suárez Pesantes

Guayaquil – Ecuador

Año: 2018

AGRADECIMIENTOS

Dedicado a Dios principalmente por permitirme alcanzar una meta más en mi vida, a mis padres Javier Suárez y Mirna Pesantes por su ayuda en este camino, a mi hermana y abuela por su motivación diaria y resto de familiares que han contribuido y alentado en todos los pasos que eh dado, a mis amigos y compañeros por motivar y ayudar con luchas y éxitos.

A mi tutor Ph.D. Francisco Vera por haber compartido sus valiosos conocimientos, al Ing. Gabriel Mantilla mentor y maestro que me permitió el desarrollo de este proyecto y al INAMHI por su apoyo en las herramientas y datos que proporcionaron confiando en mí.

DEDICATORIA

A Dios por darme la oportunidad de cumplir un sueño tan grande como este, a mis abuelos maternos Víctor Quevedo y Sofia Candelejo por su ayuda fuera y dentro de este mundo, a mi abuelo paterno Enrique Suárez, gracias por creer en mi hasta esta edad y por sus constantes consejos.

A mis tías Elena, Lilia y Graciela Suárez por su ayuda en todos los niveles de estudios.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Yo Javier Leonidas Suárez Pesantes doy mi consentimiento para que la Escuela Superior Politécnica del Litoral realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Javier Suárez Pesantes

EVALUADORES



Sandra Lorena García Bustos

PROFESOR MATERIA INTEGRADORA



Francisco Xavier Vera Alcívar

PROFESOR TUTOR

V

V

RESUMEN

En el presente estudio se busca ajustar de la mejor manera creando, transformando y comparando modelos de una forma en que se pueda diseñar para optimizar las precipitaciones en la estación meteorológica Guayaquil- Aeropuerto. Teniendo en cuenta modelos anteriores sin beneficio para las predicciones, se planteó crear un modelo con mucha capacidad de computación que sea capaz de optimizar las predicciones minimizando errores. Considerando Modelos de Regresión Lineal, Gamma, Transferencia lineal y Redes Neuronales mediante muchos enfoques distintos (transformaciones, adecuaciones e iteraciones) para lo cual se utilizó: Precipitaciones medias mensuales Enero 1915 – Mayo 2018 y Temperaturas Superficiales del Mar respecto a la Isolina con mayor correlación en las fechas antes indicadas.

Al comparar modelos descritos anteriormente, se obtuvo que la aplicación de la Red Neuronal utilizando Retardos significativos retornó estimaciones cercanas a la realidad hasta 30 metros alrededor de la estación meteorológica y con un error menor en comparación con todos los modelos aplicados, pero que se puede mejorar integrando variables que expliquen más el comportamiento de la precipitación de manera que la red se alimente de estas y pueda lograr predicciones más impecables.

Los datos utilizados se realizaron con ayuda del Software R Studio.

Palabras clave: Redes Neuronales, Precipitación, Temperatura, Iteraciones.

ABSTRACT

In this study we seek to adjust the best way to create, transform and compare models in a way that can have to optimize precipitation at the Guayaquil-Aeropuerto weather station. Considering the previous models for predictions, you can create a model with a lot of computing capacity that is able to optimize precipitation and minimize errors. Models of Linear Regression, Gamma, Linear Transfer and Neural Networks among many others (transformations, adjustments and iterations) for what they are used: Monthly average precipitation January 1915 - May 2018 and Sea Surface Temperatures with respect to Isoline with the highest correlation.

When comparing models described above, it was obtained that the application of the Neural Network using significant delays returned estimates close to reality up to 30 meters around the meteorological station and with a minor error in comparison with all applied models, but that can be improved integrating variables that explain more the behavior of the precipitation so that the network feeds on these and can achieve more impeccable predictions.

The data used was made with the help of the R Studio Software.

Keywords: Neural Networks, Precipitation, Temperature, Iterations.

INDICE GENERAL

RESUMEN.....	VI
ABSTRACT	VII
INDICE GENERAL	VIII
ABREVIATURAS.....	X
SIMBOLOGÍA	XI
LISTA DE FIGURAS.....	XII
LISTA DE TABLA	XIII
SECCIÓN 1	1
1. INTRODUCCIÓN	1
1.1 Descripción del problema	2
1.2 Justificación del problema	2
1.3 Alcance.....	2
1.4 Objetivos	3
1.4.1 Objetivo General	3
1.4.2 Objetivos Específicos.....	3
1.5 Marco teórico.....	3
1.5.1 Precipitación atmosférica	4
1.5.2 Imagen ráster	5
1.5.3 Temperatura superficial del mar.....	5
1.5.4 Transferencia lineal.....	6
1.5.5 Redes neuronales	6
SECCIÓN 2	9
2. METODOLOGÍA	9
2.1 Distancia de Mahalanobis	9
2.2 Correlación de Pearson.....	9
2.3 Regresión Lineal Simple	10
2.4 Modelo Gamma Log.....	10
2.5 Modelo de Transferencia Lineal	11
2.6 Red Neuronal (Feed - Forward)	12
SECCIÓN 3	14

3. RESULTADOS Y ANÁLISIS	14
3.1 Estadística Descriptiva	14
3.1.1 Precipitación Acumulada Mensual	14
3.1.2 Temperatura Superficial del Mar	15
3.2 Estadística Inferencial	16
3.2.1 Distribución de Precipitación	16
3.2.2 Valores Sospechosos	18
3.3 Estadística Especializada	20
3.3.1 Correlación Precipitación vs Temperaturas	20
3.4 Modelación	22
3.4.1 Regresión Lineal	22
3.4.2 Modelo Gamma Log	25
3.4.3 Modelo de Transferencia Lineal	26
3.4.4 Redes Neuronales	30
3.5 Discusión y Análisis Final de Modelos Propuestos	39
SECCIÓN 4	41
4. CONCLUSIONES Y RECOMENDACIONES	41
4.1 Conclusiones	41
4.2 Recomendaciones	42
BIBLIOGRAFÍA	43
ANEXOS	45
GLOSARIO	45
Valores de la Red Neuronal Artificial Modelo 2	46

ABREVIATURAS

INAMHI	Instituto Nacional de Meteorología e Hidrológica del Ecuador.
DGAC	Dirección General de Aviación Civil del Ecuador.
NOAA	National Oceanic and Atmospheric Administration.
RNA	Red Neuronal Artificial.
RL	Regresión Lineal.
MTL	Modelo Transferencia Lineal.
SST	Temperatura Superficial del Mar.
SCE	Sumatoria Cuadrática del Error.

SIMBOLOGÍA

Mm	Milímetros.
C ⁰	Grados centígrados.

LISTA DE FIGURAS

Figura 1.1 Ejemplo de Pluviómetro	4
Figura 1.2: Ejemplo de Imagen Raster.....	5
Figura 1.3: Ejemplo de Red Neuronal Biológica vs Red Neuronal Artificial.....	7
Figura 1.4: Ejemplo de Red Neuronal de una capa oculta	8
Figura 3.1: Diagrama de Paneles Precipitación Acumulada Mensual, INAMHI, Guayaquil-Aeropuerto.....	15
Figura 3.2: Histograma Temperatura Superficial del Mar Mensual, NOAA.....	15
Figura 3.3: Histograma Precipitación Mensual, INAMHI, Guayaquil Aeropuerto ...	16
Figura 3.4: Gráfico Quantile - Quantile, INAMHI, Guayaquil- Aeropuerto.....	17
Figura 3.5: Gráfico Valores Sospechosos, INAMHI, Guayaquil-Aeropuerto	19
Figura 3.6: Temperatura Superficial del Mar, en el Dominio de la Isolina con correlación 0.67	21
Figura 3.7: Serie Temporal Temperatura Superficial del Mar, en el dominio	21
Figura 3.8: Serie Temporal Regresión Lineal.....	23
Figura 3.9: Autocorrelaciones de Residuos Regresión Lineal	24
Figura 3.10: Serie Temporal Modelo Gamma Log	25
Figura 3.11: Cross - correlación Precipitación filtrada vs los residuos del modelo auto-regresivo de orden 5 para la Temperatura	27
Figura 3.12: Serie Temporal Modelo de Transferencia Lineal	29
Figura 3.13: Diseño de Red Neuronal Modelo 1	32
Figura 3.14: Serie Temporal Red Neuronal Modelo 1	33
Figura 3.15: Diseño de Red Neuronal Modelo 2.....	35
Figura 3.16:Autocorrelaciones de Residuos RNA Modelo 2.....	36
Figura 3.17: Serie Temporal Red Neuronal Modelo 2	37
Figura 3.18: Curva Roc Red Neuronal Modelo 2.....	38
Figura 3.19: Gráfico Predicciones de los Mejores Modelos	39

LISTA DE TABLA

Tabla 3.1: Estadísticas de precipitación por meses	14
Tabla 3.2: Pronósticos Regresión Lineal	23
Tabla 3.3: Pronósticos Modelo Gamma Log	26
Tabla 3.4: Pronósticos Modelo de Transferencia Lineal	29
Tabla 3.5: Pronósticos Red Neuronal Modelo 1	34
Tabla 3.6: Pronósticos Red Neuronal Modelo 2	38
Tabla 3.7: Sumatoria cuadrática de los errores	39

SECCIÓN 1

1. INTRODUCCIÓN

A lo largo de los años se ha intentado predecir el clima utilizando modelos matemáticos y estadísticos que permitan proyectar con el mayor grado de exactitud variables como las temperaturas y las precipitaciones con un margen de error bastante pequeño que permitan tomar decisiones que resulten en un buen modelo de gestión pública. (Wang, 1996 - 1997).

La importancia de la predicción de las precipitaciones incumbe a sucesos que tiene esta para resguardar y crear planes de contingencia por efectos de cambios bruscos de clima y fenómenos como la corriente de “El Niño” y con esto disminuir problemas en zonas de riesgo.

Para poder proyectar las precipitaciones, es necesario comprender un poco del aspecto físico de las mismas. La precipitación atmosférica se da generalmente por el calentamiento del aire húmedo en donde existe elevación de masas de aire en vertical lo que produce cúmulos que empujan el agua ya condensada hacia el suelo, para medir el agua que cae en un determinado lugar se utiliza un instrumento llamado pluviómetro que es un cilindro circular calibrado en milímetros (mm), cada milímetro de agua que recolecta el pluviómetro equivale a decir $\frac{1 \text{ litro de agua}}{m^2}$, dicho de otra manera los mm que registra el pluviómetro representan la altura en mm que alcanzaría la capa de agua sobre el suelo. (Hernández Leal, 2016)

Como el pluviómetro solo puede medir lo que cae en sitio específico en este caso en la estación meteorológica Guayaquil – Aeropuerto se utilizan métodos matemáticos (Regresión Lineal, Regresión no Lineal, Modelos ARIMA y por supuesto Redes Neuronales) para determinar cuanta lluvia cayo en cierta zona. (INAMHI, 2015)

1.1 Descripción del problema

Las constantes exigencias en tiempos actuales donde la información llega casi de inmediato generan la necesidad de precisión. En específico, las predicciones de precipitaciones que son de mucha importancia para las prevenciones a poblaciones con exposición a riesgos o fenómenos de parte de esta, considerando que en el Ecuador no se cuenta con radares meteorológicos que midan y estimen con mayor precisión las predicciones del clima existe la necesidad de estudiar estos problemas y usar recursos menores para establecer predicciones mucho más efectivas que ayuden a incrementar la preparación ante posibles fenómenos que provocarían las precipitaciones.

1.2 Justificación del problema

Este trabajo se ha elaborado con el fin de comprobar que se puede ajustar un modelo con Redes Neuronales para predecir las precipitaciones de una manera más efectiva a métodos distintos que han generado poco beneficio como en los pasados fenómenos (“El Niño”) de los años 1982 y 1997; para que, desde un punto de vista estadístico, se puedan crear planes de emergencia que eviten catástrofes en nuestro país.

1.3 Alcance

El presente estudio se establece en la estación meteorológica Guayaquil – Aeropuerto perteneciente a la DGAC con datos proporcionados por INAMHI y se estima en la costa del Ecuador continental, utilizando datos medidos de precipitación media mensual (mm) y las temperaturas superficiales del mar (SST) que está en periodos mensuales desarrollado por NOAA, de Enero 1915 - Mayo 2018.

En este trabajo proyectará de forma óptima la precipitación en la estación antes mencionada y con una estimación de 30 metros a la redonda, no se espera predecir para otra zona o para un país entero, lo que es incierto, pero si tener una idea de que tan probable es la precipitación en cierto punto específico.

1.4 Objetivos

1.4.1 Objetivo General

Diseñar un modelo de Redes Neuronales para predecir las precipitaciones atmosféricas en los primeros 5 meses del año 2018, analizando sus errores, comparando con diferentes modelos y estimando en la estación meteorológica Guayaquil – Aeropuerto.

1.4.2 Objetivos Específicos

- ✓ Minimizar el error de las Predicciones de precipitaciones de Enero - Mayo 2018.
- ✓ Predecir los primeros 5 meses del año 2018.
- ✓ Comparar diferentes modelos y su interacción con Redes Neuronales.

1.5 Marco teórico

En la siguiente sección se explica la definición de variables, recursos y metodologías a implementar para la aplicación de la Red Neuronal Artificial óptima para la predicción de precipitaciones (lluvias).

1.5.1 Precipitación atmosférica

Se conoce a la precipitación como la capa o placa de agua que es reunida en los objetos meteorológicos como pluviómetros y otros, se mide en milímetros de agua (mm) o litros por metro cuadrado ($\frac{\text{listros}}{\text{m}^2}$), las precipitaciones usadas están medidas en mm de manera mensual desde Enero 1915 – Mayo 2018 de la estación meteorológica Guayaquil - Aeropuerto con código M0056 longitud y latitud (-79.88333, -2.15333). (Troncoso, 2010)



Figura 1.1 Ejemplo de Pluviómetro

FUENTE: INAMHI, actualización de llluvias intensas 2015

1.5.1.1 Variación de la precipitación

Se analiza por estaciones del año o por año para verificar ciclos de periodos secos y húmedos por los que pasan ciertas partes del planeta en relación con la costa del ecuador en donde se puede evidenciar fenómenos como el niño y la niña. (Troncoso, 2010)

1.5.2 Imagen ráster

Es una estructura de datos compuesta por arreglos que simbolizan una rejilla de píxeles que se conoce como matriz y se puede visualizar o escanear en algún dispositivo de representación. En el presente estudio se utiliza una imagen raster de temperaturas superficiales del mar en un rango de 160 este, 30 norte, 60 oeste y 30 sur. (Wu, 1999)

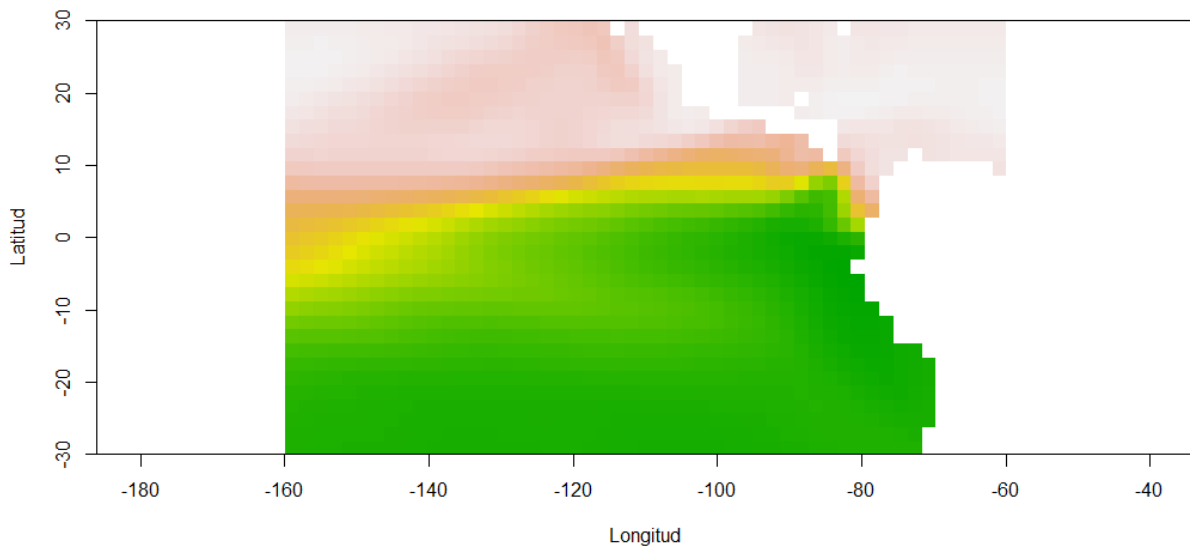


Figura 1.2: Ejemplo de Imagen Raster.

FUENTE: Elaboración propia

1.5.3 Temperatura superficial del mar

La temperatura superficial del mar es una de las insignias más importantes para la simulación del clima a lo largo de los años, designada por sensores remotos y obtenida entre una capa de agua de (0-200 metros), siendo clave en el cambio del clima provee la capacidad de evaluar patrones que puedan afectar ecosistemas o fenómenos como "El niño".

Se mide en (C^0) y se registra en los ficheros e imágenes ráster antes mencionadas en la que se usaran datos de reanálisis de forma mensual desde Enero 1915 – Mayo 2018. (] Boyin Huang, 1915-2018)

1.5.4 Transferencia lineal

En el modelo de transferencia lineal establece que la variable independiente esté cros-correlacionada con la variable a predecir o dependiente y con al menos una parte de datos históricos en el tiempo según la medida que estos lleven, pueden ser días, meses, años o de comparabilidad en el tiempo entre ambas variables; surge así la posibilidad de verificar la relación temporal entre las variables mencionadas, donde emerge un indicador de perturbación sin ejecutarse como ruido blanco y que hace imperioso la entrada de una estructura ARMA para definir dicho indicador.

Un ejemplo claro es la utilización del modelo para predecir el consumo de energía eléctrica con un resultado en la predicción mucho más impecable que con modelos rudimentarios o ya conocidos. (Javier Contreras, 2003) o del cálculo de la inflación esperada en Venezuela verificando los efectos externos que la afectan directamente. (José Guerra, 1997)

1.5.5 Redes neuronales

Se atribuye el nombre red neuronal a un sistema conectado entre múltiples neuronas, biológicamente a esta conexión se la conoce como sinapsis que es la manera en que una red neuronal aprende según las experiencias vividas que al final se llama conocimiento. En este estudio se pretende simular el concepto de red neuronal en lo que se denomina Red Neuronal Artificial que es un conjunto de nodos (neuronas) que se conectan entre sí, para estimar una información determinada y trasladarla al exterior. (Pedro Larrañaga, 2016)

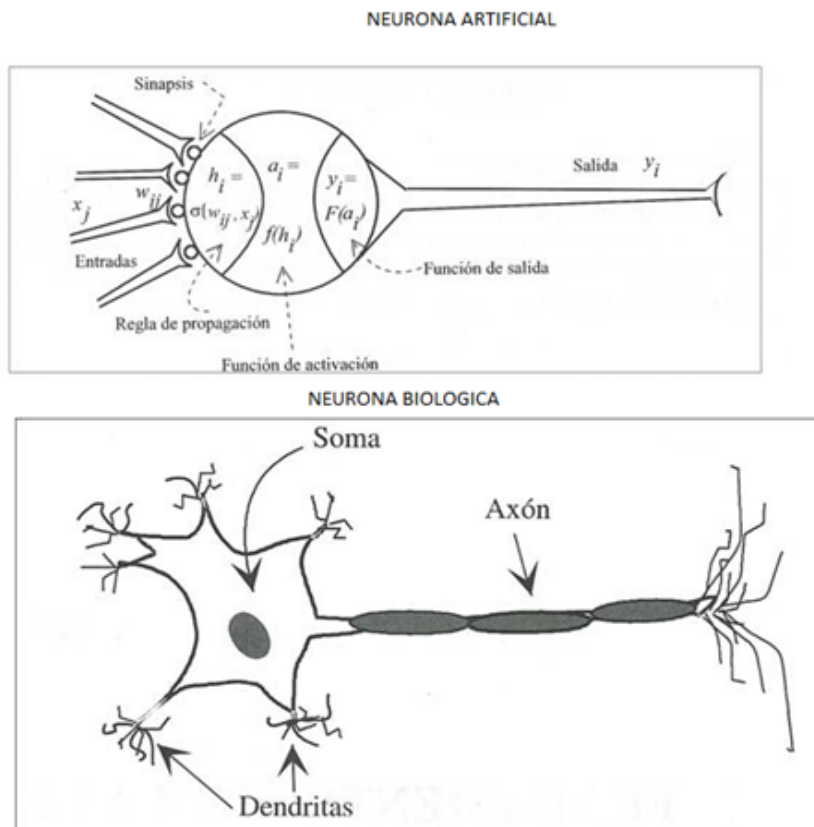


Figura 1.3: Ejemplo de Red Neuronal Biológica vs Red Neuronal Artificial

FUENTE: Pedro Larrañaga, Iñaki Inza, Abdel malik Moujahid Departamento de Ciencias de la Computación e Inteligencia Artificial Universidad del Paiz Vasco-Euskal Herriko Unibertsitatea

Teniendo en cuenta los conceptos anteriores, se puede definir de una manera más formal que una Red Neuronal Artificial comprenden a nodos de entrada (input nodes) los cuales recaudan información, nodos de salida (output nodes) los cuales difunden la información pretendida y nodos ocultos (hidden nodes) que solo intercambian información con otros nodos y son los que más fuertemente están interconectados.

Los nodos ocultos adquieren los datos de entrada que se van transformando en un término conocido como aprendizaje, existen varios tipos de algoritmos para el funcionamiento de los nodos ocultos en este estudio se utiliza Feed-Forward. (Ripley W. N., 2002)

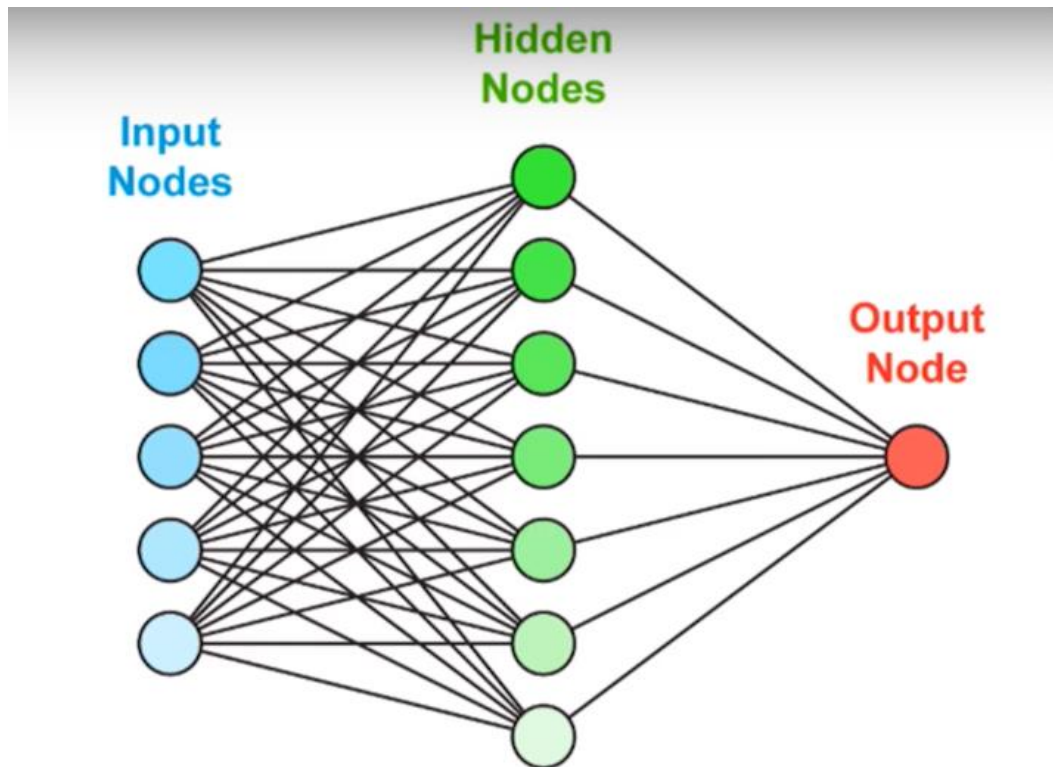


Figura 1.4: Ejemplo de Red Neuronal de una capa oculta

FUENTE: Elaboración propia

1.5.5.1 Red neuronal Feed-Forward

Generalmente conocida como modelos de regresión no lineales, pero con una cantidad de parámetros que permite una flexibilidad para estimar cualquier función suave, utilizada en muchos casos para series temporales y con relación ya confirmada entre las variables, se puede utilizar una cantidad de datos que realice una buena predicción. (Ripley W. N., 2002)

Una vez estructurada la red neuronal se establece los nodos de aprendizaje y entrenamiento donde por medio de patrones objetivos, los valores en los nodos se ajustan de manera iterativa hasta que se muestran respuestas satisfactorias, es decir en base a lo mencionado la red estará en capacidad de crear, destruir y modificar nodos para dar respuestas óptimas para el modo de entrenamiento y validación. (Hernández Leal, 2016)

SECCIÓN 2

2. METODOLOGÍA

2.1 Distancia de Mahalanobis

Es una medida de distancia y es utilizada para verificar la similitud entre dos variables aleatorias multidimensionales y se calcula de la siguiente forma:

$$d_M^2(i, j) = (x_i - x_j)' S^{-1} (x_i - x_j)$$

(2.1)

S es la matriz de covarianzas de la matriz de datos X adecuada como medida de discrepancia entre datos.

2.2 Correlación de Pearson

La correlación de Pearson es muy utilizada para medir la covariación de variables cuantitativas relacionadas fuertemente en forma lineal.

$$r_{xy} = \frac{\sum Z_x Z_y}{n}$$

Donde:

$$Z_x = \frac{X - \bar{X}}{S_X}$$
$$Z_y = \frac{Y - \bar{Y}}{S_Y}$$

(2.2)

r_{xy} = Correlación de Pearson

n = Número total de datos

Z_x, Z_y = Datos estandarizados variables X, Y

\bar{X}, \bar{Y} = Medias de variables X, Y

S_x, S_y = desviación estándar de variables X, Y

2.3 Regresión Lineal Simple

La regresión lineal es una función de relación entre la variable dependiente Y con la variable independiente X , con la cual se puede verificar causas, efectos y variación de la variable dependiente o de interés. El modelo es el siguiente:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \text{ donde } t: 1, 2, 3 \dots$$

(2.3)

Y_t = Variable dependiente

X_t = Variables independientes

β_0 = Intercepto o valor esperado de Y cuando X_t es cero

β_1 = Coeficiente que mide el cambio en Y por cada cambio de unidad en X_t

ε_t = Errores de las observaciones debido a variables no controladas

2.4 Modelo Gamma Log

Se busca que $U_i = E[Y_i]$ en términos de la predicción teniendo p covariables

$$\beta' X_i = \beta_0 + \beta_1 X_i^1 + \dots + \beta_p X_i^p$$

β' = Vector de parámetros

X = Matriz del modelo

Si una función monótona y diferenciable definiendo relación entre U_i y la predicción $\beta' X_i$ sería de la siguiente forma:

$$g(U_i) = \beta' X_i$$

Esto permite modelar una función según su parámetro así, $\theta = \beta' X$ entonces se puede aplicar la función logarítmica de la siguiente forma:

$$\log(U_i)$$

Distribución Gamma (λ, ν)

$$E(Y) = \frac{\lambda}{\nu}$$

Modelo final:

$$\log(\lambda) - \log(U_i) = \beta_0 + \beta_1 X_i$$

$$\log(U_i) = \log(\lambda) - \beta_0 - \beta_1 X_i$$

(2.4)

2.5 Modelo de Transferencia Lineal

Se utiliza el modelo de transferencia para explicar el comportamiento de sistemas en general que dependen de la Cross-Correlación de $X(variable\ independiente)$ con $Y(variable\ dependiente)$ (Ripley K.-S. C., 2018) de la siguiente forma matemática:

$$Y_t = \mu + \alpha(B) * X_t + \gamma(B) * \varepsilon_t$$

$$\alpha(B) = C * \frac{\delta_0 + \delta_1 * B^1 + \dots + \delta_{s-1} * B^{s-1}}{1 - w_1 * B^1 - \dots - w_r * B^r} * B^d$$

$$\gamma(B) = \frac{1 + \theta_1 * B^1 + \dots + \theta_q * B^q}{1 - \phi_1 * B^1 - \dots - \phi_p * B^p}$$

(2.5)

Donde:

$B = X_{t-i}$

$\delta_s, w_r =$ Polinomios en el operador de retardos B

$d =$ Numero de periodos que transcurren antes de que X afecte a Y

$s, r =$ Orden de los polinomios δ, w

$\varepsilon_t =$ Terminio de perturbación que no se comporta como ruido blanco

$\mu =$ Terminio constante

$C =$ Constante de escala para el efecto directo de X sobre Y

$\theta_q, \phi_p =$ Polinomios en el operador de retardos B , especificaciones ARMA para la perturbación ε_t (variable aleatoria idénticamente distribuida, con media 0, y desviación estándar $\sigma^2 I$)

$q, p =$ Orden de los polinomios θ, ϕ

$Y_t =$ Variable de respuesta

$X_t =$ Variable predictora

2.6 Red Neuronal (Feed - Forward)

Se utilizan para predecir en base a la experiencia o historia de una gran cantidad o volúmenes de datos, teniendo en cuenta variables de entrada (inputs), variables ocultas (hidden) y variables de salida (outputs). El modelo se puede representar de la siguiente forma:

$$y_k = \phi_0 \left(\alpha_k + \sum_h w_{hk} \phi_h \left(\alpha_h + \sum_i w_{ih} x_i \right) \right)$$

(2.6)

$y_k =$ Variable de salida (salidas)

$\phi_h =$ Función fija o de activación para capa oculta

$\alpha_h =$ Constante (sesgo) para capa oculta

$w_{ih} =$ Pesos de capa oculta

$x_i =$ Variable de entrada (entradas)

ϕ_0 = Función fija o de activación para entradas

α_k = Constante (sesgo) para entradas

w_{hk} = Pesos de entradas

La función de activación está definida por:

$$\phi_h(x) = \frac{e^x}{1 + e^x}$$

Las salidas pueden ser lineales, logísticas o de umbral este ultimo de la siguiente forma:

$$\phi_0(x) = I(x > 0)$$

SECCIÓN 3

3. RESULTADOS Y ANÁLISIS

3.1 Estadística Descriptiva

En lo sucesivo de la sección se elabora estadística descriptiva de la precipitación atmosférica medidas en *mm* y la temperatura superficial del mar medidas en C° . Para el efecto se usó registros desde Enero 1915 - Mayo 2018 y con esto preparar la conducta de las variables mencionadas con los modelos a aplicarse entre ellas.

3.1.1 Precipitación Acumulada Mensual

Los datos usados se obtuvieron de la estación meteorológica Guayaquil – Aeropuerto perteneciente a la DGAC con datos proporcionados por INAMHI, organizada y tratada para el análisis actual. Dado a cómo se comporta la precipitación según (Troncoso, 2010) se realiza un análisis descriptivo general con énfasis en los fenómenos y perturbaciones que se observe.

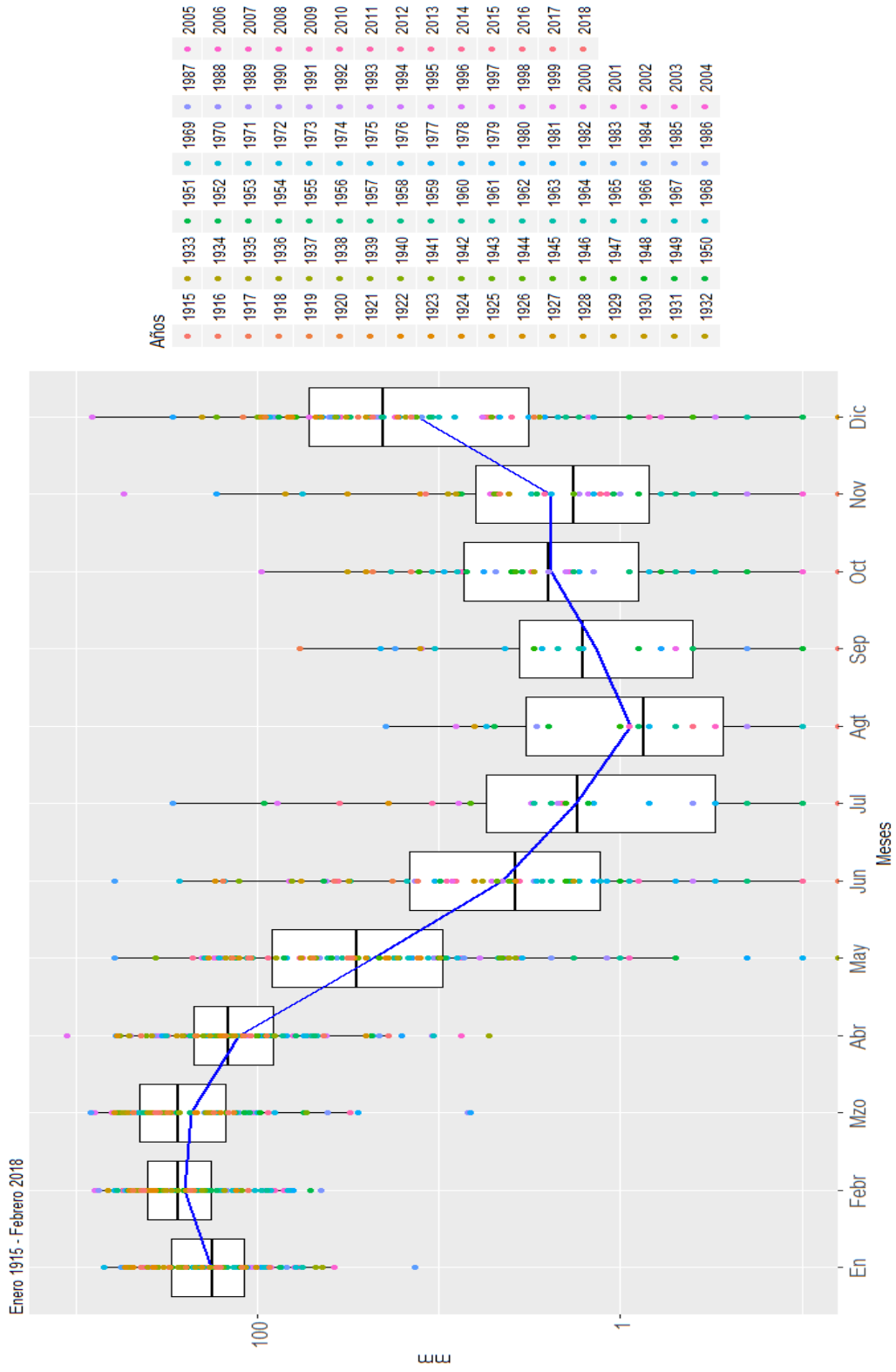


Figura 3.1: Diagrama de Paneles Precipitación Acumulada Mensual, INAMHI, Guayaquil-Aeropuerto
Fuente: Elaboración Propia

En la Figura 3.1 se distingue las precipitaciones de la base de datos correspondiente desde el año 1915 al 2018, se puede apreciar que los meses de Enero, Febrero, Marzo, Abril, mayo y parte de Junio experimentan altos índices de lluvia, la causa de este evento es la ubicación de la zona intertropical con constantes precipitaciones en donde se encuentra la estación meteorológica Guayaquil – Aeropuerto, mientras que en los meses de Julio a Noviembre existe menor probabilidad de lluvia; además se puede observar una reactivación de lluvia en Diciembre.

Tabla 3.1: Estadísticas de precipitación por meses

Meses	Media	Desviación Estándar	Mínimo	Mediana	Máximo
Enero	220,08	138,72	13,5	177,4	701,3
Febrero	299,31	162,81	44,5	274	795,2
Marzo	298,2	182,82	6,7	272,75	832,4
Abril	188,53	171,52	5,3	145,4	1124
Mayo	57,54	85,57	0	25,1	618,4
Junio	21,57	72,04	0	1,3	612,2
Julio	5,49	31,06	0	0	291,5
Agosto	0,54	2,3	0	0	19,7
Septiembre	1,52	6,55	0	0	58,2
Octubre	3,11	10,57	0	0	94,4
Noviembre	10,17	57,08	0	0	549,4
Diciembre	38,39	90,14	0	13,6	818,6

FUENTE: Elaboración propia

En la Tabla 3.1 se distingue el valor promedio de las precipitaciones (media) que siguiendo al análisis realizado en la Figura 3.1 con promedios altos entre los meses de Enero – Abril, intermedios entre Mayo – Julio y bajos de Agosto – Noviembre; además una reactivación en diciembre.

En la desviación estándar que se refiere al valor que difiere de la media, es decir que entre Enero – Abril se registraron lluvias fuertes como débiles en cada mes, siendo abril la más probable en tener días sin lluvia dentro del mes promedio.

En los valores extremos tenemos a los máximos y mínimos importantes ya que provee las máximas y mínimas precipitaciones en los meses, así como la mediana que se refiere al valor intermedio de los datos de precipitación ordenados.

Estos datos serán de gran importancia para la comparación y análisis de modelos.

3.1.2 Temperatura Superficial del Mar

Los datos usados se obtuvieron de NOAA, tratados para el análisis actual. Los ficheros almacenan registros desde 1915 - 2018. Dado que las temperaturas son tan importantes en el comportamiento de los cataclismos y cambios del clima, se realiza un análisis descriptivo en la zona con mayor correlación obtenida en la Figura 3.6 especificando en variaciones y forma de la misma.

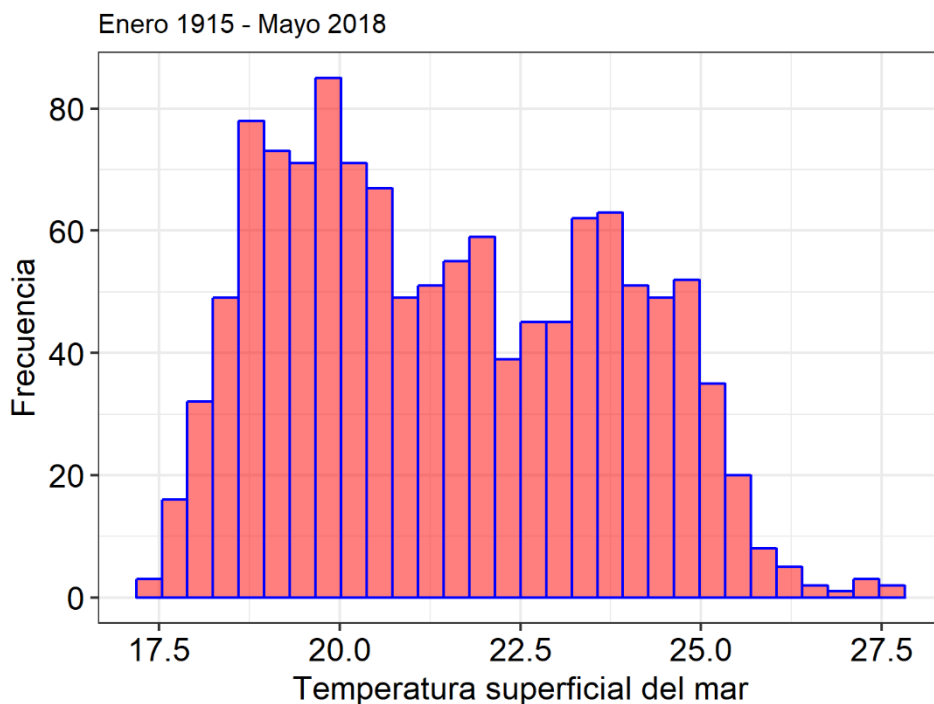


Figura 3.2: Histograma Temperatura Superficial del Mar Mensual, NOAA

FUENTE: Elaboración propia

En la Figura 3.2 se puede distinguir las temperaturas superficiales del mar con una correlación significativa, por otra parte, se visualiza una acumulación de observaciones en el rango de $19C^0$ a $21C^0$ aproximadamente, además se puede notar la oscilación de las temperaturas entre $18C^0$ a $28C^0$. Dado que ya esta probado que las precipitaciones y temperaturas son proporcionales, se espera que a mayor temperatura las lluvias aumenten en el sitio de estudio. ([Boyin Huang, 1915-2018)

3.2 Estadística Inferencial

En lo sucesivo de la sección se elabora estadística inferencial de la precipitación atmosférica y con esto se logra especificar fenómenos y cambios en los que la precipitación incide regularmente y su forma sobre la estación meteorológica.

3.2.1 Distribución de Precipitación

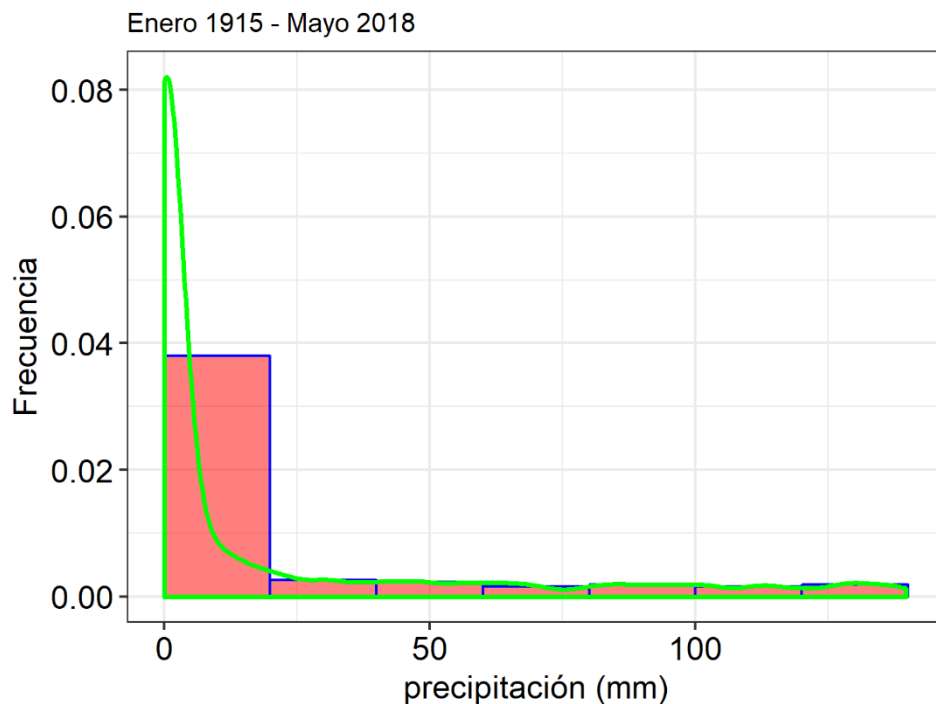


Figura 3.3: Histograma Precipitación Mensual, INAMHI, Guayaquil Aeropuerto

FUENTE: Elaboración propia

En la Figura 3.3 se presenta un histograma de las precipitaciones más representativas que dan lugar a la distribución aproximada en estas, se aprecia una compacta acumulación de 0 a un valor aproximado a 25 mm, visualizando una distribución que se equipara a una Gamma lo cual es importante en los modelos a implementarse en este estudio. Dado que se podría ajustar dicha distribución es eficaz indagar el comportamiento de la precipitación en dicho modelo.

Prueba Shapiro Wilk

$H_0 =$ Precipitación pertenece a una distribución normal

$H_a =$ Precipitación no pertenece a una distribución normal

Según la prueba de Shapiro Wilk con un valor p de $2.2 e^{-16} < 0.05$ existe evidencia estadística para rechazar la hipótesis nula H_0 , lo cual solo indica que los datos de precipitación no se ajustan a una distribución normal.

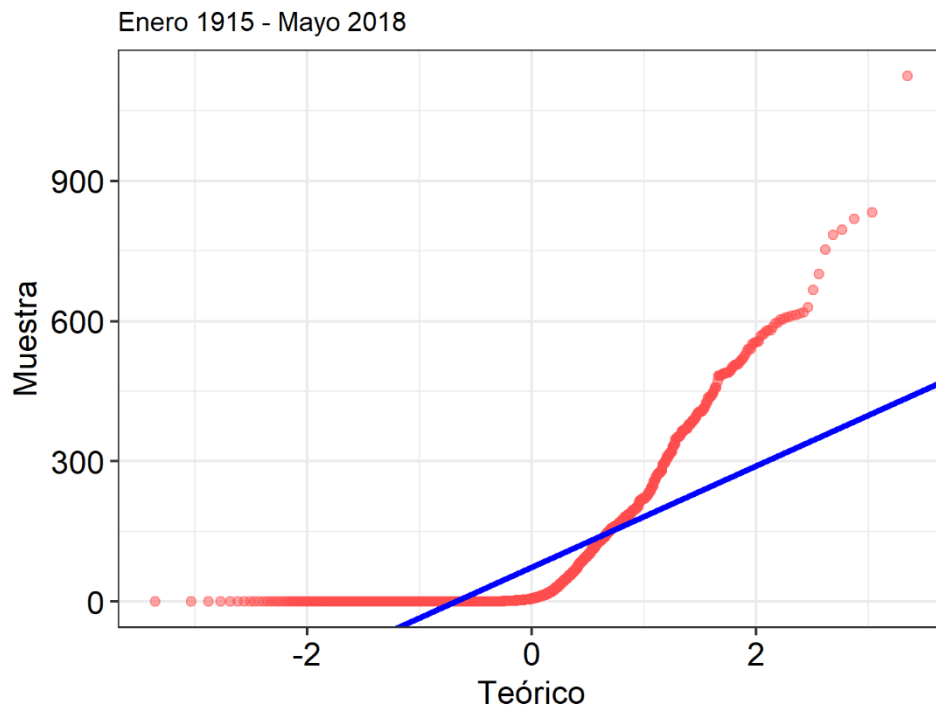


Figura 3.4: Gráfico Quantile - Quantile, INAMHI, Guayaquil- Aeropuerto

FUENTE: Elaboración propia

$H_0 =$ La muestra proviene de una población con distribución normal

$H_0 =$ La muestra no proviene de una población con distribución normal

En la Figura 3.4 se distingue que los datos de la muestra de precipitación no provienen de una distribución normal ya que los datos no se ajustan a la línea referencial.

3.2.2 Valores Sospechosos

Un problema a menudo es el uso de los datos y arreglo de estos, frecuentemente los valores sospechosos son datos que no son congruentes con el resto, en general son valores extremos o muy bajos, dado que esto es un indicador que interviene mucho en optimizar las predicciones ya que pueden crear errores muy grandes como potenciar la predicción según sea el caso.

Distancia de mahalanobis

Para esto se utilizó la distancia de mahalanobis para verificar similitud entre los diferentes datos de precipitación, dado que mahalanobis es una longitud entre punto y una distribución D se puede escribir de la siguiente forma:

$$d_M^2(i, j) = (\text{Precipitación}_i - \text{precipitación}_j)' S^{-1} (\text{Precipitación}_i - \text{precipitación}_j)$$

$$S^{-1} = \text{covarianza de precipitación}$$

(3.1)

utilizando también una distribución X^2 (ji) cuadrado con 3 grados de libertad y un umbral de D mayor a 12 se puede obtener valores que están fuera de lo normal en una serie de datos. (Ruefer, 2016)

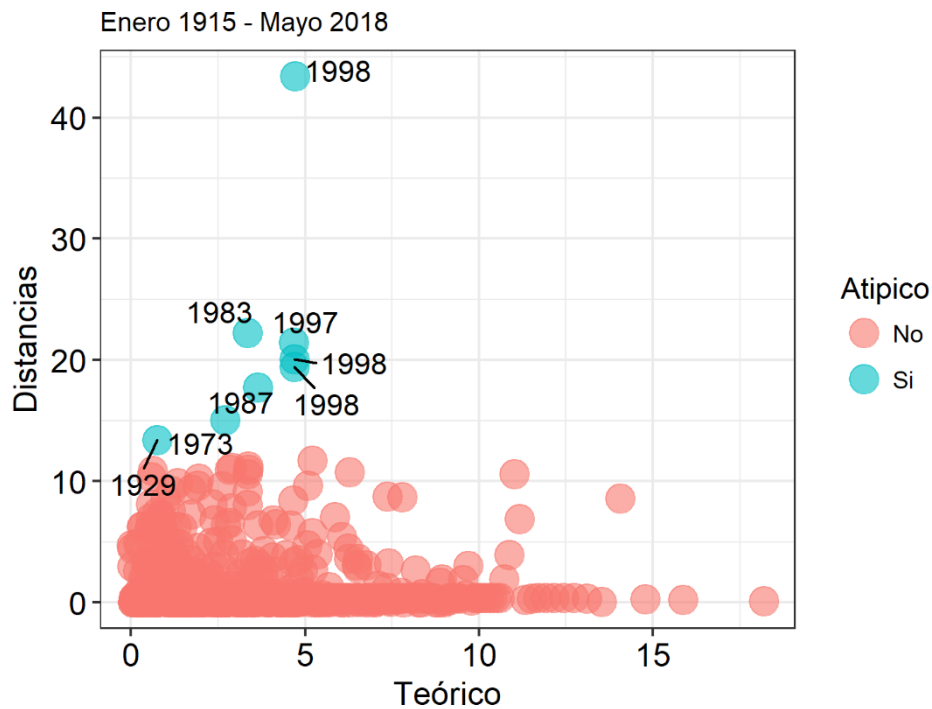


Figura 3.5: Gráfico Valores Sospechosos, INAMHI, Guayaquil-Aeropuerto

FUENTE: Elaboración propia

En la Figura 3.5 se puede distinguir que existen 8 valores sospechosos (atípicos) los cuales tienen una explicación ya que los valores de 1983, 1997 y 1998 son conocidos en Ecuador por los problemas que ocasionó el fenómeno “El Niño”. Considerados fenómenos ENOS declarados por la NOAA (NOAA, n.d.) en los años 1972-1973, 1982-1983 y 1986-1987; siendo los más intensos 1982-1983 y 1997-1998 registrados hasta la fecha (Margarita, 2005), mientras que, en los años 1929 existe un desfase en el mes de febrero que es normal esperar y donde la precipitación es mayor. Todos estos valores se representan como atípicos en los episodios registrados en estos años, que por motivos investigativos es importante mantenerlos en la serie, y conocer si el modelo estadístico es capaz de sobre estimar un evento como el fenómeno del Niño.

3.3 Estadística Especializada

En lo sucesivo luego de un análisis general de las variables a utilizar, se procede con la metodología descrita en el capítulo 2 para demostrar los comportamientos conjuntos de precipitación y temperatura, basado en la evidencia estadística se resuelve adjuntar los siguientes análisis:

3.3.1 Correlación Precipitación vs Temperaturas

En este ensayo se pretende conocer la correlación de Pearson que según su metodología sirve para medir la covariación entre variables estrechamente relacionadas de forma lineal, para el efecto se realiza ajustes multidimensionales desde Enero 1915 – Mayo 2018. Los múltiples ajustes figuran pixeles como grillas donde se contienen valores de la temperatura, latitudes y longitudes, teniendo en cuenta esto se procede a crear una serie de tiempo con los valores de temperatura de igual manera procediendo a adjuntar los datos referenciales de precipitación, obteniendo dos series de preámbulo para el cálculo de la correlación de Pearson. De esta forma se puede construir la Figura 3.6 que simboliza el cálculo de las correlaciones en el rango especificado. (J. Cedeño, 2018)

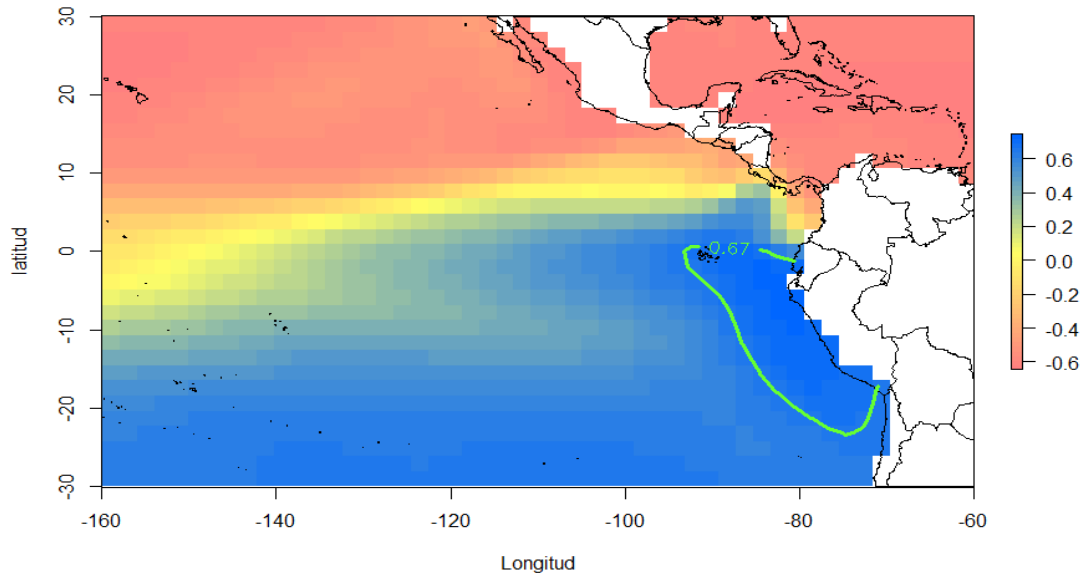


Figura 3.6: Temperatura Superficial del Mar, en el Dominio de la Isolina con correlación 0.67

FUENTE: Elaboración propia

En la Figura 3.6 se puede distinguir la correlación mínima, media y máxima en todo el rango mencionado, para escoger la mayor correlación posible entre precipitación y temperatura se obtiene la curva de nivel donde la correlación en este caso 0.67 (percentil 95) donde obtendremos valores promedios espaciales para la temperatura que es la variable predictora que se utilizó para los modelos.

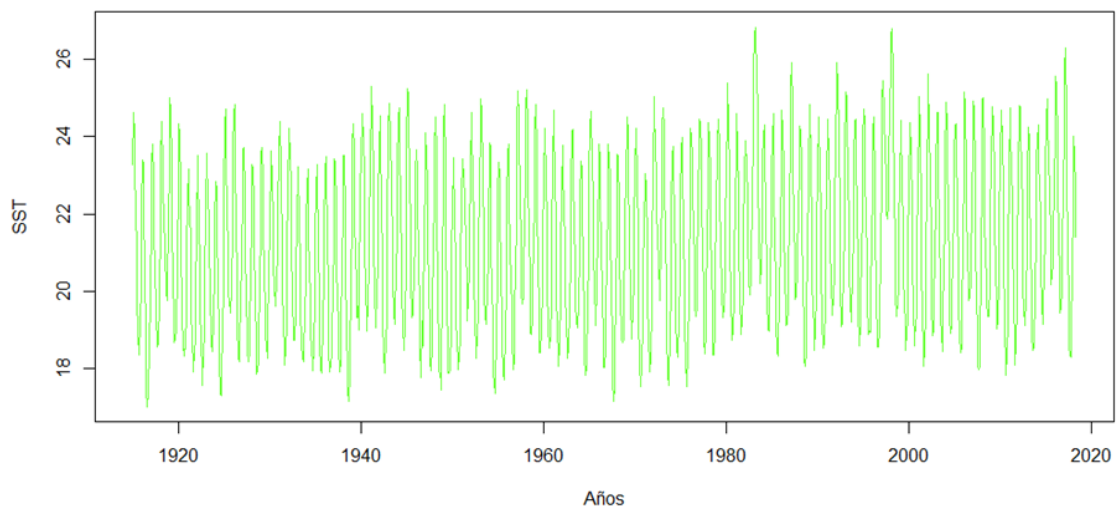


Figura 3.7: Serie Temporal Temperatura Superficial del Mar, en el dominio

FUENTE: Elaboración propia

En la Figura 3.7 se puede distinguir la serie temporal de temperaturas (SST) con la máxima correlación en la curva de nivel antes mencionada como se especifica en la Figura 3.2 en su distribución con temperaturas entre $18C^0$ a $28C^0$, observando además que en los años 1983 y 1998 se encuentran los valores máximos de temperatura que produjeron una precipitación alta verificada en la Figura 3.6.

3.4 Modelación

En este apartado se explica los modelos preliminares para el correcto diseño de la red neuronal que es el objetivo ya descrito del contenido, en donde se puede notar mejor el comportamiento de las variables.

3.4.1 Regresión Lineal

$$\begin{aligned} \text{Precipitación}^{\frac{1}{2}} &= \beta_0 + \beta_1 \text{Temperatura} + \varepsilon_t \\ \text{Precipitación}^{\frac{1}{2}} &= -53,78 + 2,83 \text{Temperatura} + \varepsilon_t \end{aligned} \quad (3.2)$$

En este caso se realizó una regresión lineal dado la proporcionalidad confirmada entre precipitación y temperatura, pero con una variante que es la transformación de precipitación y utilizando a la temperatura como variable independiente, muestra la variación de la precipitación en función de la temperatura que permitió realizar las proyecciones representadas en el Gráfico 3.8.

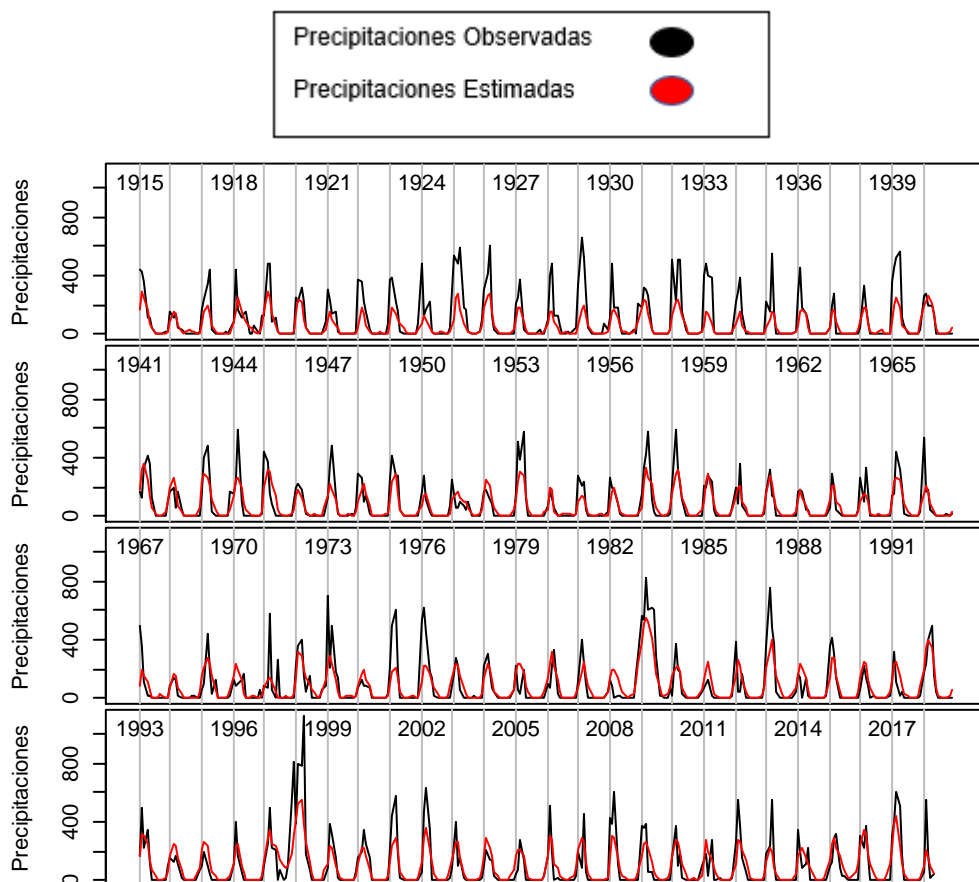


Figura 3.8: Serie Temporal Regresión Lineal

FUENTE: Elaboración propia

En la Figura 3.8 se puede distinguir los valores observados y estimados lo cual visualmente nos indica que en general existe una subestimación de la precipitación, ya que es un modelo en donde no se tiene en cuenta el tiempo se puede pasar a un modelo con una estructura mejorada.

Tabla 3.2: Pronósticos Regresión Lineal

Meses	Precipitación Estimada	Precipitación Observada
Enero	99,5	59,2
Febrero	205,7	553,9
Marzo	167,7	177,7
Abril	110,3	18,9
Mayo	48,0	46,3

FUENTE: Elaboración propia

En la Tabla 3.2 se puede notar una subestimación en febrero y marzo, como una sobreestimación de Enero, Abril y Mayo, lo que indica que es una proyección optimista en algunos casos pero que se puede mejorar.

Análisis de residuos

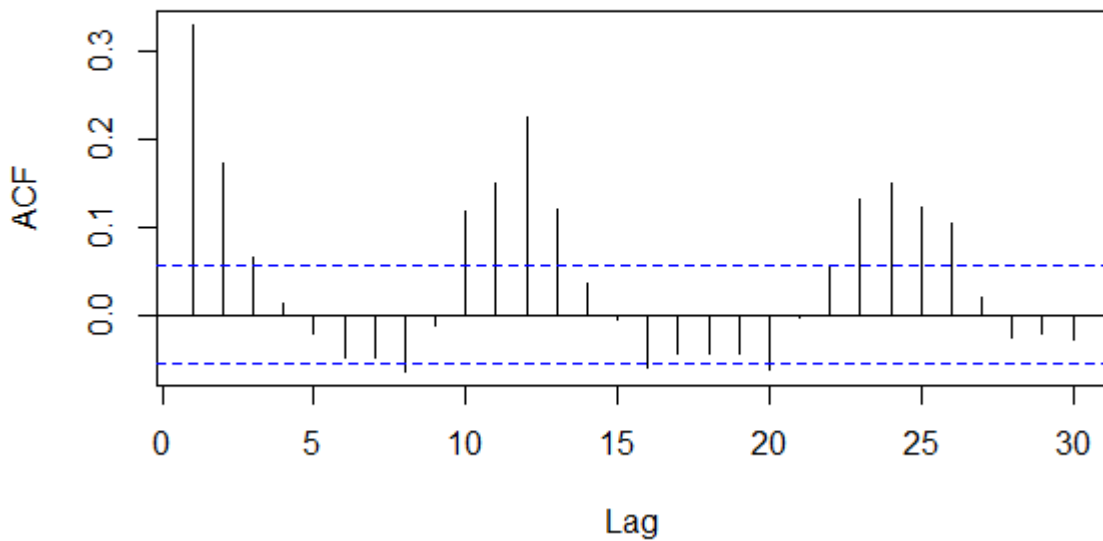


Figura 3.9: Autocorrelaciones de Residuos Regresión Lineal

FUENTE: Elaboración propia

En la Figura 3.9 se detalla las autocorrelaciones de los residuos en donde se puede observar que varias espigas sobresalen de los intervalos rotulados en color azul, así como una notable forma cíclica que supone el uso de un componente autorregresivo estacional, en este caso la espiga 12 se puede interpretar como la inclusión en el modelo de un retardo en la precipitación que se tendrá en cuenta para la implementación en la red neuronal a diseñar.

3.4.2 Modelo Gamma Log

$$\log(E(\text{Precipitación})) = \log \lambda - \beta_0 - \beta_1 \text{Temperatura}$$
$$\log(E(\text{Precipitación})) = 14,80 - 0,87 \text{Temperatura}$$

(3.3)

En este modelo se utiliza el logaritmo en los valores esperados de precipitación teniendo en cuenta la distribución Gamma que se efectúa por la forma de la distribución en la Figura 3.3; además de la elasticidad de la distribución de precipitación que es beneficioso para el modelo mencionado. Un ejemplo claro son las ventajas mencionadas de la inclusión del modelo en precipitaciones que inician intensamente y bajan progresivamente. (Henriquez)

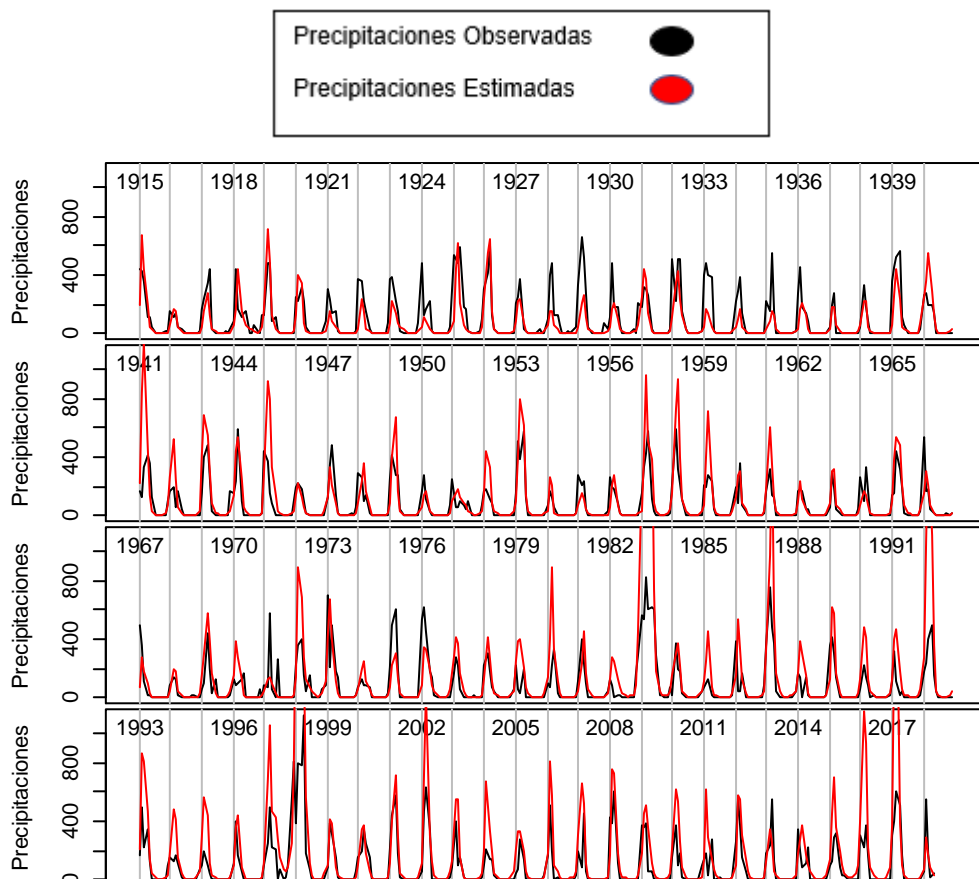


Figura 3.10: Serie Temporal Modelo Gamma Log

FUENTE: Elaboración propia

En la Figura 3.10 se puede distinguir una sobreestimación en general en tal caso se puede verificar que este modelo es muy optimista ya que espera mayor precipitación en cuanto a los valores estimados.

Tabla 3.3: Pronósticos Modelo Gamma Log

Meses	Precipitación Estimada	Precipitación Observada
Enero	79,2	59,2
Febrero	305,9	553,9
Marzo	199,0	177,7
Abril	93,2	18,9
Mayo	30,8	46,3

FUENTE: Elaboración propia

En la Tabla 3.3 se puede distinguir que en los meses de Enero, Marzo y Abril se sobreestima y en Febrero y Mayo se subestima de acuerdo a esto el Modelo Gamma Log es menos efectivo que el Modelo de Regresión Lineal.

3.4.3 Modelo de Transferencia Lineal

Según lo expuesto en el marco teórico de este trabajo, es necesario que la precipitación este Cross-Correlacionada con la temperatura, para el efecto se realiza un Gráfico CCF que impulsa a la obtención del modelo ARMA.

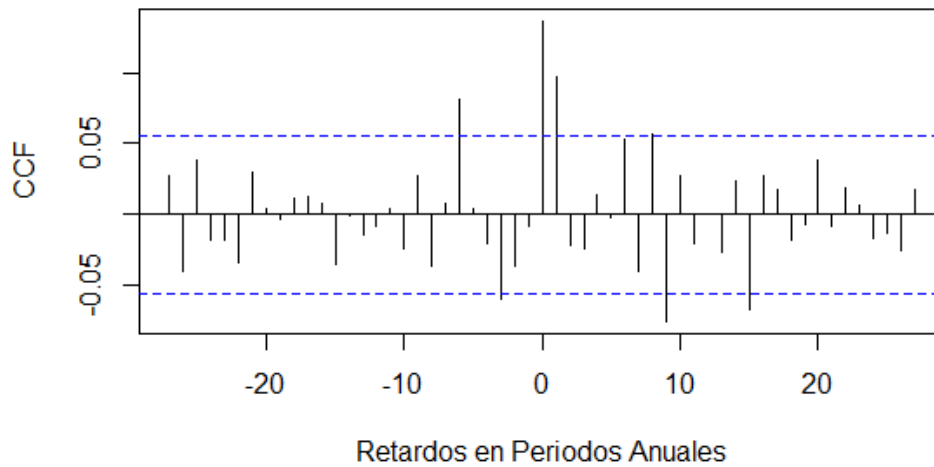


Figura 3.11: Cross - correlación Precipitación filtrada vs los residuos del modelo auto-regresivo de orden 5 para la Temperatura

FUENTE: Elaboración propia

En la Figura 3.11 se puede demostrar el retado a partir del 0 a izquierda es decir la espiga -3 sobresale de los intervalos de confianza, teniendo un retardo significativo de hasta 3 meses de la temperatura con respecto a la precipitación, en el dominio explicado anteriormente. Según la metodología Box-Cox, se analizó este sistema dinámico, según la ayuda de la tabla 11.1 del libro Time Series Analysis Forecasting and Control.

Esto marca una huella en la Precipitación que crea un movimiento en la temperatura con esto se estructuró un el modelo con (r, s, d) como (0,1,3). (Javier Contreras, 2003).

Con lo analizado utilizando temperatura como variable dependiente se propone un modelo AR (5) para que los residuos sean de ruido blanco, en donde se empleó la función ARIMAX (Ripley K.-S. C., 2018).

El modelo de transferencia lineal está compuesto de la siguiente forma según la metodología:

$$Y_t \sim \beta_0 + \delta_0 B^3 X_t + \delta_1 B^1 X_t + \frac{1}{1 - (\phi_1 B^1 + \phi_2 B^2 + \phi_3 B^3 + \phi_4 B^4 + \phi_5 B^5)} * \varepsilon_t$$

$$\begin{aligned} Y_t(1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5) \\ = \beta_0^* + \delta_0 B^3 X_t (1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5) \\ + \delta_1 B^1 X_t (1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5) + \varepsilon_t \end{aligned}$$



$$\begin{aligned} Y_t = \beta_0^* + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \delta_1 X_{t-1} - \delta_1 X_{t-2} + (\delta_0 \\ - \delta_1 \phi_2) X_{t-3} + (-\delta_0 \phi_1 - \delta_1 \phi_3) X_{t-4} + (-\delta_0 \phi_2 - \delta_1 \phi_4) X_{t-5} + (-\delta_0 \phi_3 \\ - \delta_1 \phi_5) X_{t-6} - \delta_0 \phi_4 X_{t-7} - \delta_0 \phi_5 X_{t-8} + \varepsilon_t \end{aligned}$$

$$\begin{aligned} Y_t = 467,36 + 0,47 Y_{t-1} - 0,03 Y_{t-2} - 0,13 Y_{t-3} + 0,05 Y_{t-4} + 0,14 Y_{t-5} \\ - 76,41 X_{t-1} + 35,91 X_{t-2} + 56,54 X_{t-3} - 35,58 X_{t-4} + 5,59 X_{t-5} \\ + 18,35 X_{t-6} - 2,94 X_{t-7} - 8,24 X_{t-8} + \varepsilon_t \end{aligned}$$

(3.4)

Y_t : Precipitación en el mes t

X_{t-i}, Y_{t-i} : Temperatura y precipitación del mes anterior i

Precipitaciones Observadas	
Precipitaciones Estimadas	

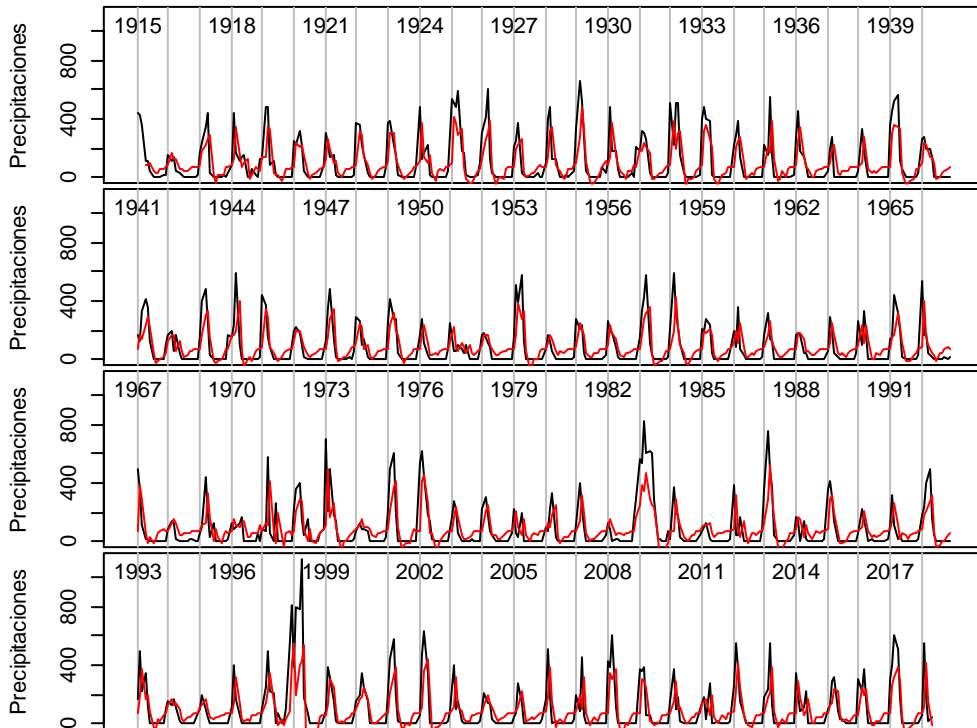


Figura 3.12: Serie Temporal Modelo de Transferencia Lineal

FUENTE: Elaboración propia

En la Figura 3.12 se puede distinguir que las estimaciones no están cercanas a la realidad y provee un resultado cauteloso teniendo en cuenta que en este modelo se toman los retos y el tiempo, se espera sea superior a los dos modelos analizados anteriormente que según lo observado no alcanza a las estimaciones a lo largo de 1915 – 2018.

Tabla 3.4: Pronósticos Modelo de Transferencia Lineal

Meses	Precipitación Estimada	Precipitación Observada
Enero	138,5	59,2
Febrero	242,9	553,9
Marzo	253,0	177,7
Abril	229,4	18,9
Mayo	164,2	46,3

FUENTE: Elaboración propia

En la tabla 3.4 se puede observar lo mencionado antes, que es un modelo muy reservado en las estimaciones dado que en el mes de marzo por

ejemplo existe un error alto así mismo en abril y en contraste con los dos modelos anteriores el Modelo de Regresión lineal sigue siendo el mejor.

3.4.4 Redes Neuronales

En este ensayo se estableció dos modelos de redes neuronales que resultaron interesantes en la fase de conformación de estas, de las cuales se escogerá la mejor para la Prueba y recomendación final.

$$Y_t = \text{Precipitación atmosférica en el mes } t$$
$$X_t = \text{Temperatura superficial del mar en el mes } t$$

Teniendo en cuenta la variable Y_t (variable dependiente) y X_t (variable independiente), según la metodología de redes neuronales se propone a 10 nodos como entrada dado a un análisis sobre aquellas de la siguiente manera:

$$X_{t-1}, Y_{t-1}: \text{Temperatura y precipitación del mes anterior}$$
$$X_{t-2}, Y_{t-2}: \text{Temperatura y precipitación de 2 meses anteriores}$$
$$X_{t-12}, Y_{t-12}: \text{Temperatura y precipitación de 12 meses anteriores}$$
$$X_{t-13}, Y_{t-13}: \text{Temperatura y precipitación de 13 meses anteriores}$$
$$X_{t-24}, Y_{t-24}: \text{Temperatura y precipitación de 24 meses anteriores}$$

es decir, los retardos de cada variable ya que en la práctica lo mejor es trabajar con la historia.

Se propone una capa oculta con 20 nodos para la sinapsis que estarán denotados de la siguiente manera H_1, H_2, \dots, H_{20} y una capa de salida con un solo nodo, puesto que es necesario la estimación de una sola variable en este caso Y_t es decir Precipitaciones. (Ripley W. N., 2002)

3.4.4.1 Modelo 1

Para el modelo 1 se utilizó una transformación en los retardos y la variable real de precipitaciones para intentar establecer datos entre 0 y 1 los cuales adopta mejor una red neuronal dado que la función de activación es logística se esperaría mejores resultados (Ripley W. N., 2002), siendo la transformación de la siguiente forma:

$$m_o(\hat{Y}_t) = \frac{\alpha m_o(Y_t)}{\alpha m_o(Y_t) + 1} = 0.5$$
$$\alpha = \frac{1}{m_o(Y_t)}$$

(3.5)

m_o = mediana de las precipitaciones mayores a cero

Y_t = Precipitaciones

$$\alpha = \frac{1}{\frac{\frac{1}{77} Y_{t-i}}{\frac{1}{77} Y_{t-i} + 1}} ; \text{ donde } i = 1, 2, 12, 13 \text{ y } 24$$

(3.6)

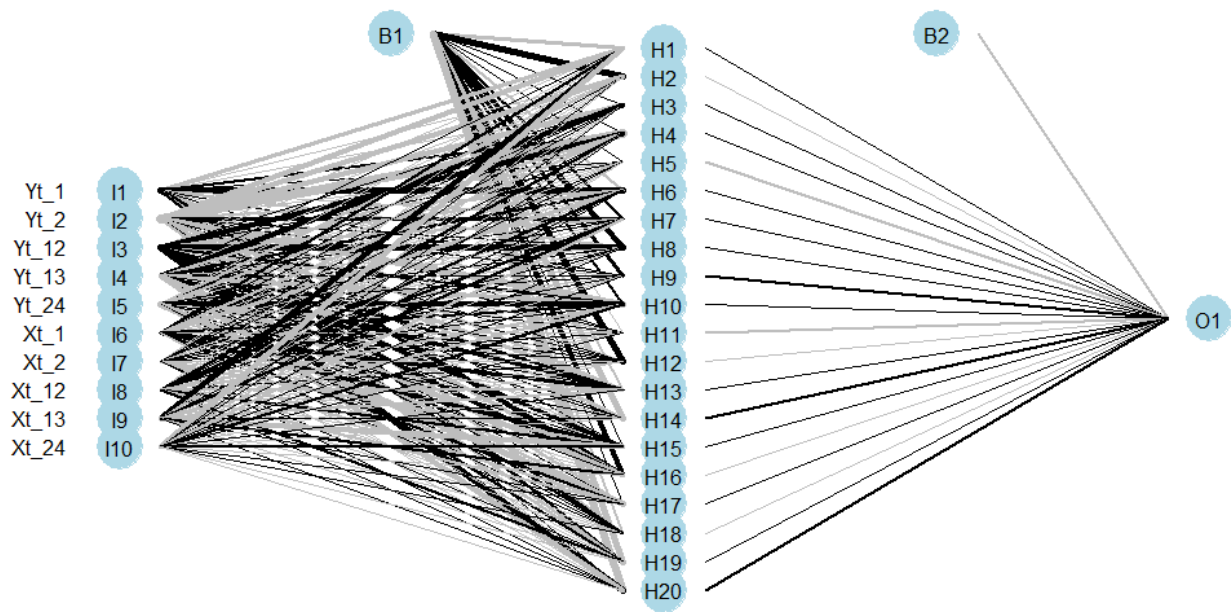


Figura 3.13: Diseño de Red Neuronal Modelo 1

FUENTE: Elaboración propia

En la Figura 3.13 se puede distinguir el diseño de la red neuronal con 3 capas, la capa de entrada con 10 nodos, la capa oculta con 20 nodos y la capa de salida con un solo nodo que representa la estimación de las precipitaciones.

A continuación, se detalla el Modelo matemático para la Red Neuronal del modelo 1:

$$\begin{aligned}
 y_t = \phi_0(\alpha + w_1\phi_1(\alpha_1 + w_{1,1}x_{t-1} + w_{2,1}y_{t-1} + w_{3,1}x_{t-2} \\
 + w_{4,1}y_{t-2} + w_{5,1}x_{t-12} + w_{6,1}y_{t-12} + w_{7,1}x_{t-13} \\
 + w_{8,1}y_{t-13} + w_{9,1}x_{t-24} + w_{10,1}y_{t-24}) \\
 + w_2\phi_2(\alpha_2 + w_{1,2}x_{t-1} + w_{2,2}y_{t-1} + w_{3,2}x_{t-2} \\
 + w_{4,2}y_{t-2} + w_{5,2}x_{t-12} + w_{6,2}y_{t-12} + w_{7,2}x_{t-13} \\
 + w_{8,2}y_{t-13} + w_{9,2}x_{t-24} + w_{10,2}y_{t-24}) + \dots \\
 + w_{20}\phi_{20}(\alpha_{20} + w_{1,20}x_{t-1} + w_{2,20}y_{t-1} \\
 + w_{3,20}x_{t-2} + w_{4,20}y_{t-2} + w_{5,20}x_{t-12} \\
 + w_{6,20}y_{t-12} + w_{7,20}x_{t-13} + w_{8,20}y_{t-13} \\
 + w_{9,20}x_{t-24} + w_{10,20}y_{t-24})
 \end{aligned}
 \tag{3.7}$$

Para el modelado de la red neuronal se utilizó los parámetros y funciones descritos en (Ripley W. N., 2002), así como la función nnet en donde se utilizó fase de entrenamiento (Enero 1915 – Diciembre 2017) y validación (5 primeros meses del 2018).

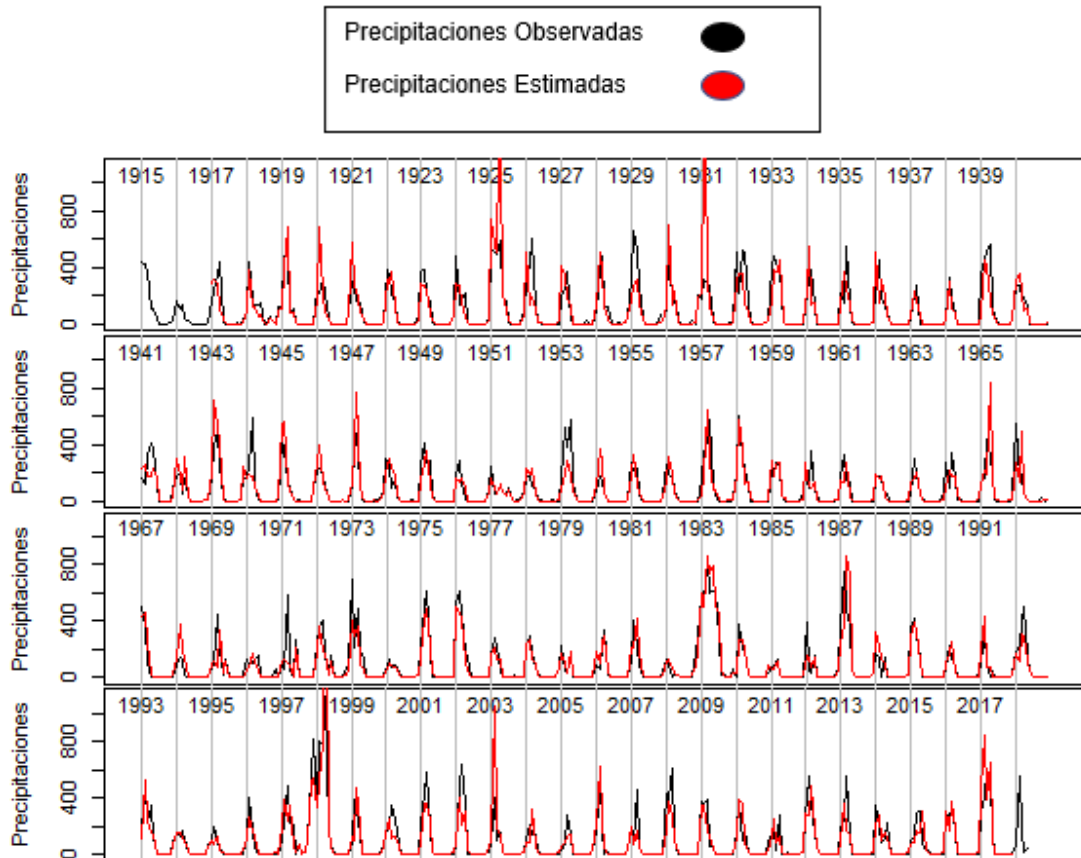


Figura 3.14: Serie Temporal Red Neuronal Modelo 1

FUENTE: Elaboración propia

En la Figura 3.14 se puede notar sobreestimaciones en algunos años, pero en general mantiene una tendencia casi cercana a las observaciones para lo cual se analizó las estimaciones de los cinco últimos meses.

Tabla 3.5: Pronósticos Red Neuronal Modelo 1

Meses	Precipitación Estimada	Precipitación Observada
Enero	248,2	59,2
Febrero	698,8	553,9
Marzo	231,6	177,7
Abril	70,2	18,9
Mayo	3,1	46,3

FUENTE: propia

En la Tabla 3.5 se puede apreciar una notable sobreestimación, aunque se usa los retardos no supera al Modelo de Regresión Lineal, pero es mucho mejor que los modelos Gamma Log y de Transferencia Lineal.

3.4.4.2 Modelo 2

Para el modelo 2 se utilizó una transformación en los retardos y la variable real de precipitación para intentar replicar la transformación usada en el Modelo de Regresión Lineal con la única diferencia que se aplicara en las variables de entrada para los retardos de precipitaciones de la red neuronal de esta manera quedaría un diseño que se aprecia en la Figura 3.15.

$$\sqrt{(Y_{t-i})}; \text{ donde } i = 1, 2, 12, 13 \text{ y } 24$$

(3.8)

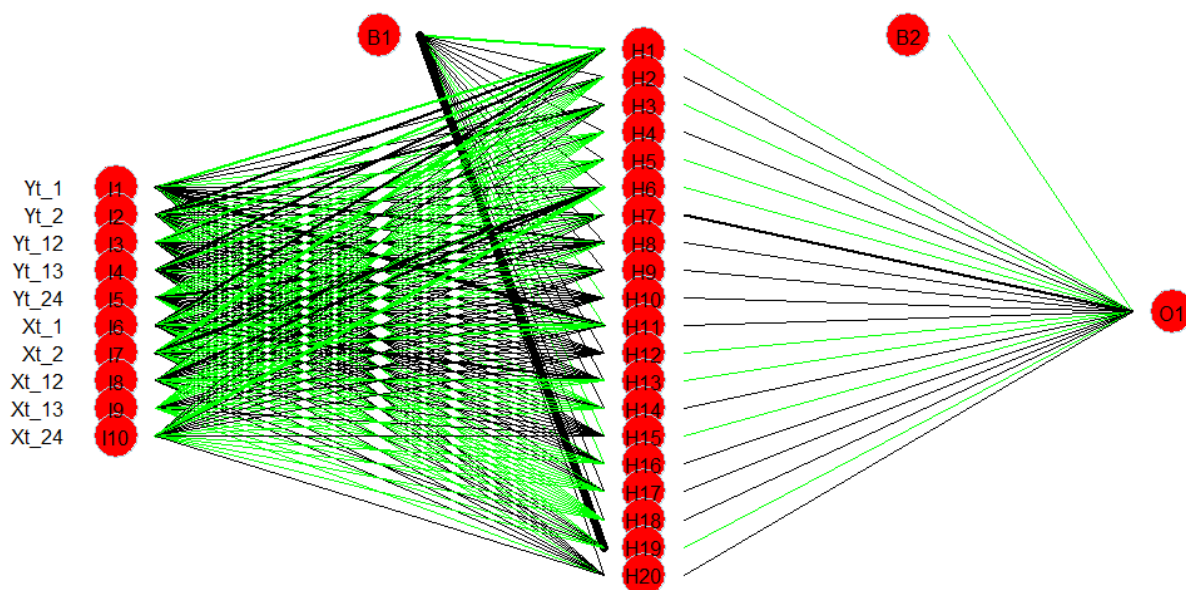


Figura 3.15: Diseño de Red Neuronal Modelo 2

FUENTE: Elaboración propia

En la Figura 3.15 se puede distinguir el diseño de la red neuronal con 3 capas, la capa de entrada con 10 nodos, la capa oculta con 20 nodos y la capa de salida con un solo nodo que representa la estimación de las precipitaciones.

A continuación, se detalla el Modelo matemático para la Red Neuronal del modelo 2:

$$\begin{aligned}
 y_t = & \phi_0(\alpha + w_1\phi_1(\alpha_1 + w_{1,1}x_{t-1} + w_{2,1}y_{t-1} + w_{3,1}x_{t-2} + w_{4,1}y_{t-2} \\
 & + w_{5,1}x_{t-12} + w_{6,1}y_{t-12} + w_{7,1}x_{t-13} + w_{8,1}y_{t-13} \\
 & + w_{9,1}x_{t-24} + w_{10,1}y_{t-24}) \\
 & + w_2\phi_2(\alpha_2 + w_{1,2}x_{t-1} + w_{2,2}y_{t-1} + w_{3,2}x_{t-2} + w_{4,2}y_{t-2} \\
 & + w_{5,2}x_{t-12} + w_{6,2}y_{t-12} + w_{7,2}x_{t-13} + w_{8,2}y_{t-13} \\
 & + w_{9,2}x_{t-24} + w_{10,2}y_{t-24}) + \dots \\
 & + w_{20}\phi_{20}(\alpha_{20} + w_{1,20}x_{t-1} + w_{2,20}y_{t-1} + w_{3,20}x_{t-2} \\
 & + w_{4,20}y_{t-2} + w_{5,20}x_{t-12} + w_{6,20}y_{t-12} + w_{7,20}x_{t-13} \\
 & + w_{8,20}y_{t-13} + w_{9,20}x_{t-24} + w_{10,20}y_{t-24})
 \end{aligned}
 \tag{ 3.9}$$

$$\begin{aligned}
y_t = & \phi_0(-0,01 \\
& - 1,60 \phi_1(-64,28 - 52,19 x_{t-1} \\
& + 46,21 y_{t-1} - 59,78 x_{t-2} + 81,71 y_{t-2} \\
& - 28,48 x_{t-12} + 75,09 y_{t-12} - 97,08 x_{t-13} \\
& + 18,27 y_{t-13} - 44,62 x_{t-24} + 45,29 y_{t-24}) \\
& + 0,42 \phi_2(0 + 0,09 x_{t-1} - 0,05 y_{t-1} + 0,10 x_{t-2} \\
& + 0,07 y_{t-2} - 0,21 x_{t-12} - 0,15 y_{t-12} - 0,24 x_{t-13} \\
& - 0,09 y_{t-13} - 0,13 x_{t-24} - 0,25 y_{t-24}) + \dots \\
& + 0,01 \phi_{20}(0,08 + 0,11 x_{t-1} + 0,02 y_{t-1} \\
& + 0,11 x_{t-2} + 0,08 y_{t-2} - 0,01 x_{t-12} + 0,08 y_{t-12} \\
& + 0,14 x_{t-13} + 0,20 y_{t-13} - 0,02 x_{t-24} \\
& + 0,23 y_{t-24})
\end{aligned}$$

(3.10)

Para el modelado de la red neuronal se utilizó los parámetros y funciones descritos en (Ripley W. N., 2002), así como la función nnet en donde se utilizó fase de entrenamiento (Enero 1915 – Diciembre 2017) y validación (5 primeros meses del 2018).

Análisis de autocorrelaciones

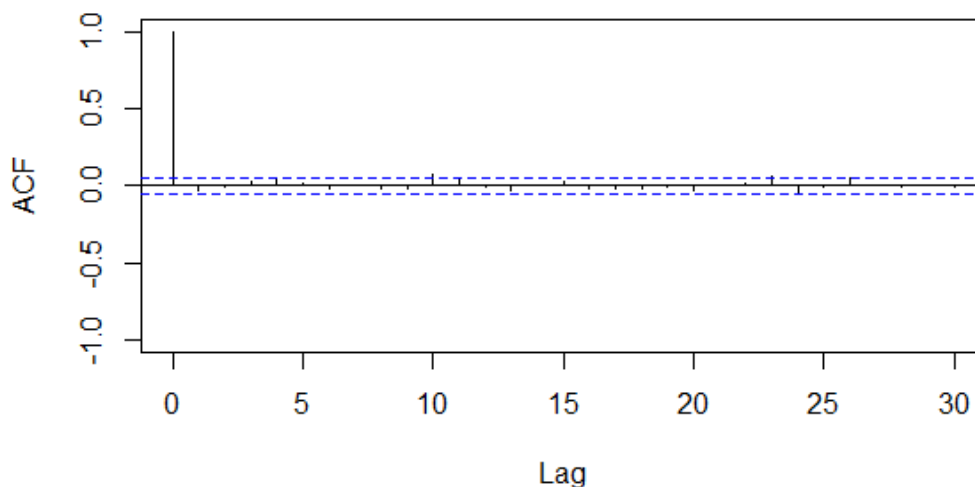


Figura 3.16: Autocorrelaciones de Residuos RNA Modelo 2
FUENTE: Elaboración propia

En la Figura 3.16 se puede distinguir que no existen autocorrelaciones y como se acoto en la Figura 3.11 existe la necesidad de incluir retardos que se implementaron en el modelo de la Red neuronal para la optimización.

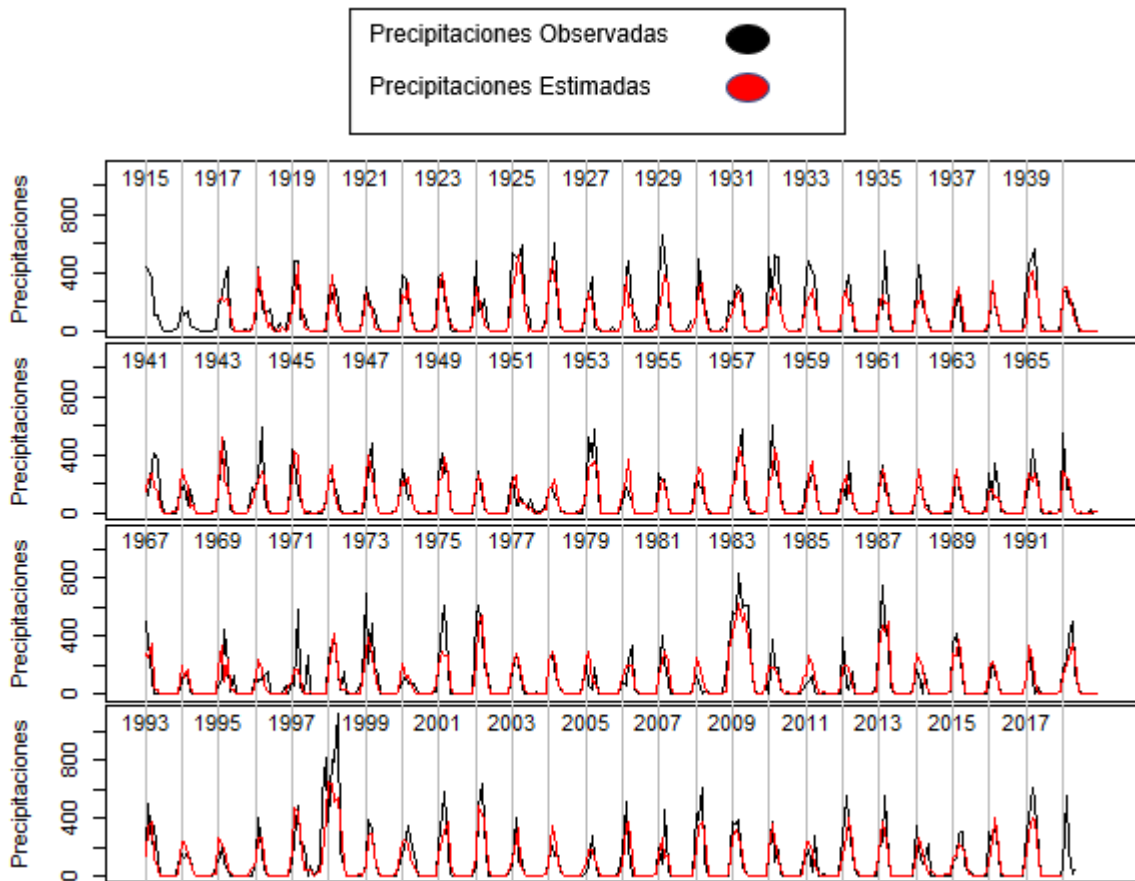


Figura 3.17: Serie Temporal Red Neuronal Modelo 2

FUENTE: Elaboración propia

En la figura 3.17 se puede distinguir una mejor adecuación de la red neuronal en base a las variables de entradas que se le proporcionó, dado que se nota una estimación correcta en la mayoría de los años se espera una estimación cercana a la realidad.

Tabla 3.6: Pronósticos Red Neuronal Modelo 2

Meses	Precipitación Estimada	Precipitación Observada
Enero	152,0	59,2
Febrero	441,9	553,9
Marzo	276,4	177,7
Abril	44,2	18,9
Mayo	0,2	46,3

FUENTE: Elaboración Propia

En la Tabla 3.6 se puede observar uno de los mejores pronósticos que la red neuronal puede conseguir, aunque no exactas, pero si cercanas con una semilla (6930) se pueden plantear muchas más estimaciones con diferentes semillas que optimizaran las estimaciones para estudios futuros.

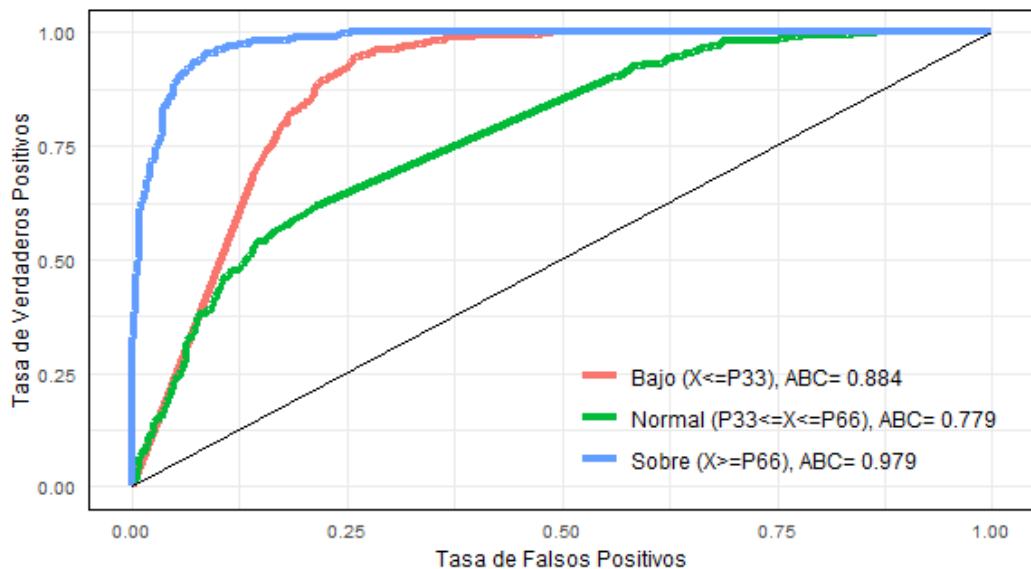


Figura 3.18: Curva Roc Red Neuronal Modelo 2

FUENTE: Elaboración propia

En la Figura 3.18 se puede notar la precisión del modelo que es muy bueno para estimar sobre la normal es decir que en los meses de Enero – Mayo 2018 existiría probablemente lluvias intensas en un

97%, en contraste con lo dicho anteriormente el modelo también es bueno para estimar bajo la normal es decir en estos meses existiría probablemente menos intensidad de lluvias en un 88%.

3.5 Discusión y Análisis Final de Modelos Propuestos

Tabla 3.7: Sumatoria cuadrática de los errores

Modelos	SCE
Regresión Lineal	12'586 026
Gamma Log	63'368 405
Transferencia Lineal	15'247 428
RNA Modelo 1	34'226 086
RNA Modelo 2	8'481 833

FUENTE: Elaboración Propia

En la Tabla 3.7 se puede distinguir los diferentes modelos aplicados y la sumatoria cuadrática del error que fue la medida para la discriminación de los modelos dado que en cada uno se usan diferentes parámetros se concluyó utilizarla, de esta manera se puede tener la fidelidad de que RNA Modelo 2 estimara pronósticos óptimos para la futura toma de decisiones.

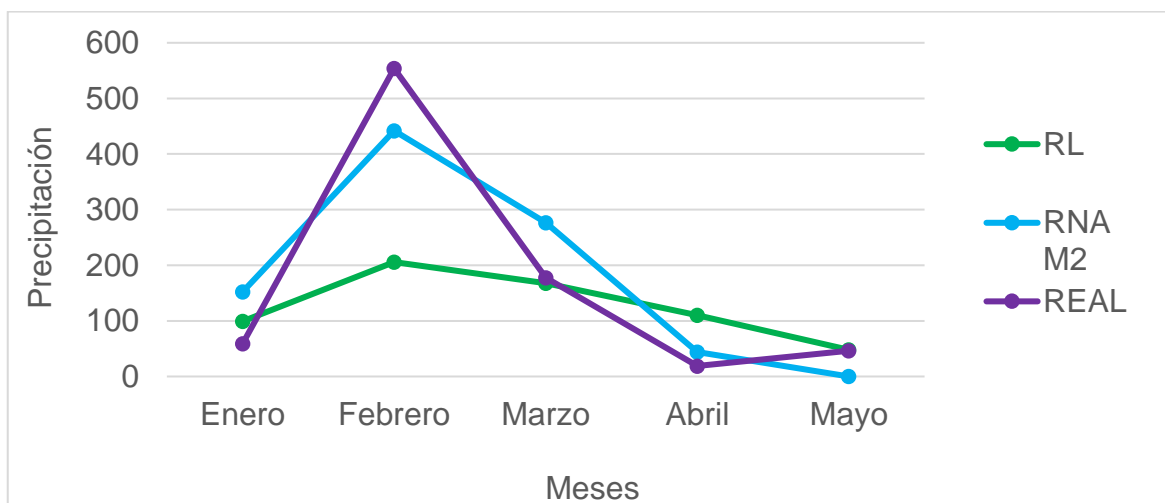


Figura 3.19: Gráfico Predicciones de los Mejores Modelos

FUENTE: Elaboración Propia

En la Figura 3.19 se puede distinguir los dos mejores modelos obtenidos en los análisis previos, se nota claramente que la Red Neuronal Modelo 2 después de diferentes iteraciones trata de adecuarse a la realidad por lo cual es sin duda el mejor modelo obtenido que gracias a la Regresión lineal también mostrado aquí, se pudo establecer la transformación final que llegó a estabilizar la predicción.

SECCIÓN 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- ✓ A través de los análisis generales de las precipitaciones en la estación meteorológica Guayaquil- Aeropuerto y de las Temperaturas superficiales del mar obtenidos de NOAA, se puede concluir que significativamente existe una relación positiva es decir proporcional entre ambas variables.
- ✓ En base a la distribución de precipitaciones se puede concluir que existen meses en donde hay menor precipitación y que en los primeros 5-6 meses del año existen lluvias abundantes.
- ✓ Existen temperaturas entre 18 a 28 grados en la correlación de la curva de nivel esto debido a la posición de la estación meteorológica en la Zona intertropical.
- ✓ Se coincide en los análisis con los eventos ocurridos en los años 1982 y 1997 en donde se observan fenómenos que si se comparan con la historia se deben a fenómenos como “El Niño”.
- ✓ En base a la Minimización de los errores entre modelos aplicados se pudo lograr en la Red Neuronal Modelo 2 la mínima SCE que sugiere el mejor modelo posible utilizando datos históricos.
- ✓ En la comparación de modelos es importante mencionar que la Regresión Lineal con transformación \sqrt{y} fue determinante para la aplicación de la Red Neuronal que se planteó y la implicación de los residuos para su implementación

4.2 Recomendaciones

- ✓ Se recomienda Mejorar el código utilizado en la modelación de las Redes Neuronales para obtener una mejor estimación o en su defecto ejecutar el procesamiento de la red hasta obtener una semilla mucho mejor que la planteada en el mejor modelo.
- ✓ Se recomienda realizar un análisis de puntos de quiebre para determinar distintos patrones que puedan mejorar tanto la serie de datos y por consiguiente las predicciones.
- ✓ Se recomienda utilizar variables Humedad y Vientos para optimizar las estimaciones de la precipitación en trabajos futuros ya que esas variables tienen alta correlación e influyen en el comportamiento de las lluvias y fenómenos.

BIBLIOGRAFÍA

-] Boyin Huang, V. F.-M. (Enero- Mayo de 1915-2018). Satellite images were collected in NetCDF format and converted to GeoTiff files for analyses. Several satellites and sensors were used to obtain the variables and to complete the period of analysis. *Satellite images were collected in NetCDF format and converted to GeoTiff files for analyses. Using remote sensors from satellite and in situ information from January 1854 to May 2018* . NOAA.
- Henriquez, F. J. (s.f.). *www.webnode.com*. Obtenido de Ciencias Actuariales y Financieras: <http://files.freddy-jose-henriquez.webnode.com.ve/>
- Hernández Leal, E. J. (2016). Generación de pronósticos para la precipitación diaria en una serie de tiempo de datos meteorológicos. *Ingenio Magno*, 144 - 155.
- INAMHI. (2015). *Determinación de Ecuaciones para el Calculo de Intensidades Máximas de Precipitación*. Quito: Subproceso de Estudios e Investigaciones Hidrológicas.
- J. Cedeño, G. M. (2018). Poster Categorical predictability of precipitation in the Ecuadorian coast and Galapagos islands using Support Vector Machines. *V INTERNATIONAL CONFERENCE ON EL NINO SOUTHERN OSCILLATION*.
- Javier Contreras, R. E. (03 de Agosto de 2003). ARIMA Models to predict Next-Day Electricity Prices. *IEEE TRANSACTIONS ON POWER SYSTEMS*.
- José Guerra, G. S. (1997). Modelos de series de tiempo para predecir la inflación en Venezuela. ., Venezuela.
- Pedro Larrañaga, I. I. (15 de Septiembre de 2016). Redes neuronales. España.
- Ripley, K.-S. C. (2018). Time Series Analysis.
- Ripley, W. N. (15 de Marzo de 2002). Modern Applied Statistics with S.

Ruefer, S. (08 de Diciembre de 2016). *R noticias y tutoriales aportados por (750) R bloggers*. Obtenido de <https://www.r-bloggers.com/outlier-detection-with-mahalanobis-distance/>

Troncoso, J. b. (Febrero de 2010). Atlas climatológico del Ecuador. Quito, Ecuador.

Wang, M. T. (1996 - 1997). *Precipitation Variability and Its Relation to Pacific Sea Surface Temperature*. Illinois: Department of Atmospheric Sciences, University of Illinois at Urbana–Champaign, Urbana, Illinois.

Wu, X. W. (1999). Estimation of Sea Surface Temperatures Using GOES-8/9 Radiance Measurements. *Bulletin of the American Meteorological Society* 80, no. 6 .

ANEXOS

GLOSARIO

Precipitación: Descenso del agua contenida en las nubes debido a la condensación de vapor.

Atípico: Observaciones o datos que son diferentes al resto de la serie.

Distribución estadística: Función que establece a cada evento en la variable aleatoria la probabilidad que este ocurra.

Estación meteorológica: Base estratégica que registra y mide diferentes variables meteorológicas.

Zona intertropical: Franja latitudinal del planeta tierra en donde los rayos solares chocan verticalmente.

Meteorología: Estudio de la atmosfera y su estado en el tiempo.

Milímetros: Medida de representación que detalla el agua acumulada que se precipita hacia el suelo.

Proyección: Estimaciones en las que se establecen supuestos para anticipar un suceso.

Predicción: anticipación de un suceso de interés.

Predicción Estadística: Proyección en situación de incertidumbre utilizando supuestos de la manera más realista para tratar de representar lo que va a ocurrir.

ENOS: Oscilación sur fenómeno “El Niño” que es una guía climática que implica en cambios de la temperatura del mar.

Valores de la Red Neuronal Artificial Modelo 2

b->h1	-6,43E+01	i6->h4	-1,27E-01	i4->h8	1,38E+01	i2->h12	1,41E-01	b->h16	-1,03E-01	i9->h19	-3,50E+01
i1->h1	-5,22E+01	i7->h4	-1,95E-01	i5->h8	4,69E+01	i3->h12	9,75E-02	i1->h16	-1,01E-02	i10->h19	-1,66E+01
i2->h1	4,62E+01	i8->h4	-1,45E-01	i6->h8	9,31E+00	i4->h12	8,79E-02	i2->h16	1,16E-01	b->h20	8,50E-02
i3->h1	-5,98E+01	i9->h4	-6,27E-02	i7->h8	-1,96E+01	i5->h12	3,43E-01	i3->h16	1,25E-01	i1->h20	1,13E-01
i4->h1	8,17E+01	i10->h4	-2,29E-01	i8->h8	1,94E+01	i6->h12	3,24E-01	i4->h16	9,65E-02	i2->h20	1,84E-02
i5->h1	-2,85E+01	b->h5	2,65E-02	i9->h8	-1,42E+01	i7->h12	4,26E-01	i5->h16	9,87E-02	i3->h20	1,11E-01
i6->h1	7,51E+01	i1->h5	-2,31E-02	i10->h8	-2,26E+01	i8->h12	2,82E-01	i6->h16	-3,46E-01	i4->h20	8,18E-02
i7->h1	-9,71E+01	i2->h5	-7,89E-02	b->h9	8,53E-02	i9->h12	2,70E-01	i7->h16	-2,89E-01	i5->h20	-8,54E-03
i8->h1	1,83E+01	i3->h5	-2,01E-02	i1->h9	1,47E-03	i10->h12	4,56E-01	i8->h16	-2,15E-01	i6->h20	8,02E-02
i9->h1	-4,46E+01	i4->h5	1,11E-01	i2->h9	6,49E-02	b->h13	4,04E+01	i9->h16	-1,69E-01	i7->h20	1,39E-01
i10->h1	4,53E+01	i5->h5	-2,30E-02	i3->h9	4,79E-02	i1->h13	3,44E+01	i10->h16	-1,94E-01	i8->h20	2,02E-01
b->h2	4,62E-04	i6->h5	-2,44E-01	i4->h9	-1,75E-02	i2->h13	1,62E+01	b->h17	2,77E-03	i9->h20	-1,95E-02
i1->h2	9,17E-02	i7->h5	-7,83E-02	i5->h9	-2,58E-02	i3->h13	5,39E+00	i1->h17	5,25E-02	i10->h20	2,26E-01
i2->h2	-4,59E-02	i8->h5	-1,46E-01	i6->h9	7,18E-02	i4->h13	2,06E+01	i2->h17	-9,96E-02	b->o	-1,15E-02
i3->h2	1,00E-01	i9->h5	-8,80E-02	i7->h9	-8,71E-02	i5->h13	4,66E+00	i3->h17	1,19E-01	h1->o	-1,60E+00
i4->h2	6,67E-02	i10->h5	-1,30E-01	i8->h9	7,61E-02	i6->h13	5,75E+01	i4->h17	9,55E-02	i4->o	4,21E-01
i5->h2	-2,12E-01	b->h6	2,97E+01	i9->h9	8,93E-02	i7->h13	1,31E+01	i5->h17	-6,32E-02	h3->o	-2,09E+00
i6->h2	-1,50E-01	i1->h6	1,31E+00	i10->h9	9,78E-02	i8->h13	-6,06E+01	i6->h17	-3,17E-01	h4->o	8,97E-02
i7->h2	-2,38E-01	i2->h6	2,43E+01	b->h10	-7,18E-02	i9->h13	-3,42E+01	i7->h17	-2,45E-01	h5->o	-1,29E-01
i8->h2	-8,87E-02	i3->h6	-4,13E+01	i1->h10	1,96E-01	i10->h13	-1,24E+01	i8->h17	-2,40E-01	h6->o	-3,59E+00
i9->h2	-1,33E-01	i4->h6	2,17E+01	i2->h10	2,37E-01	b->h14	5,73E-02	i9->h17	-1,50E-01	h7->o	4,93E+01
i10->h2	-2,51E-01	i5->h6	-1,08E+01	i3->h10	2,60E-01	i1->h14	-1,25E-02	i10->h17	-1,22E-01	h8->o	3,92E+00
b->h3	2,98E+01	i6->h6	-1,01E+01	i4->h10	1,92E-01	i2->h14	-9,17E-02	b->h18	2,90E-03	h9->o	9,93E-02
i1->h3	4,50E+00	i7->h6	9,97E+01	i5->h10	2,34E-01	i3->h14	-9,77E-02	i1->h18	-4,02E-01	h10->o	1,58E+00
i2->h3	6,54E+01	i8->h6	5,14E+01	i6->h10	2,75E-01	i4->h14	4,26E-03	i2->h18	-9,86E-02	h11->o	2,75E+00
i3->h3	-2,13E+00	i9->h6	-4,32E+00	i7->h10	3,35E-01	i5->h14	6,99E-03	i3->h18	-4,42E-01	h12->o	-5,29E-01
i4->h3	-5,06E+01	i10->h6	-1,02E+02	i8->h10	2,33E-01	i6->h14	-1,75E-01	i4->h18	-2,79E-01	h13->o	-2,93E+00
i5->h3	-4,49E+00	b->h7	-3,71E+00	i9->h10	3,43E-01	i7->h14	-9,05E-02	i5->h18	-1,75E-01	h14->o	2,63E-02
i6->h3	1,91E+01	i1->h7	1,35E-02	i10->h10	3,53E-01	i8->h14	-6,55E-02	i6->h18	-4,68E-01	h15->o	-3,09E-02
i7->h3	4,22E+01	i2->h7	2,81E-02	b->h11	-1,32E+01	i9->h14	-1,79E-01	i7->h18	-4,43E-01	h16->o	5,16E-01
i8->h3	-8,80E+00	i3->h7	-9,06E-03	i1->h11	1,05E+02	i10->h14	-1,84E-01	i8->h18	-3,86E-01	h17->o	1,51E-01
i9->h3	-3,90E+01	i4->h7	1,42E-02	i2->h11	-1,64E+01	b->h15	-5,63E-02	i9->h18	-4,47E-01	h18->o	1,68E-01
i10->h3	1,22E+00	i5->h7	-1,10E-02	i3->h11	4,11E+01	i1->h15	9,51E-02	i10->h18	-5,00E-01	h19->o	-3,28E+00
b->h4	1,00E-01	i6->h7	4,29E-01	i4->h11	-3,35E+01	i2->h15	1,14E-01	b->h19	2,75E+02	h20->o	5,98E-03
i1->h4	-9,83E-02	i7->h7	-2,72E-01	i5->h11	2,80E+01	i3->h15	8,98E-02	i1->h19	-1,84E+00		
i2->h4	-1,99E-02	i8->h7	6,55E-02	i6->h11	-5,76E+01	i4->h15	1,32E-01	i2->h19	-5,65E+00		
i3->h4	-3,32E-04	i9->h7	-1,32E-01	i7->h11	-6,02E+01	i5->h15	1,52E-01	i3->h19	8,37E-01		
i4->h4	3,20E-03	i10->h7	9,92E-04	i8->h11	-1,94E+00	i6->h15	1,74E-01	i4->h19	3,07E+00		
i5->h4	-1,90E-03	b->h8	1,19E+01	i9->h11	1,10E+01	i7->h15	1,07E-01	i5->h19	-2,33E+00		
i6->h4	-1,27E-01	i1->h8	6,08E+01	i10->h11	4,24E+01	i8->h15	2,33E-01	i6->h19	3,24E+01		
i7->h4	-1,95E-01	i2->h8	-3,30E+01	b->h12	6,16E-02	i9->h15	5,84E-02	i7->h19	-2,22E+01		
i8->h4	-1,45E-01	i3->h8	3,79E+01	i1->h12	2,72E-01	i10->h15	1,42E-01	i8->h19	2,50E+01		