



**UNIVERSIDAD TÉCNICA AMBATO
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA
CARRERA DE INGENIERÍA CIVIL**

**PROYECTO TÉCNICO PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO CIVIL**

TEMA:

**“HOMOGENEIZACIÓN DE SERIES PLUVIOMÉTRICAS DE LA CUENCA
DEL RÍO PASTAZA, UTILIZANDO MÓDULOS DE INTELIGENCIA
ARTIFICIAL.”**

AUTOR: Alexander Adalberto Rogel Rojas

TUTOR: Ing. Mg. Galo Wilfrido Núñez Aldás

AMBATO - ECUADOR

Septiembre – 2022


CERTIFICACIÓN

En mi calidad de Tutor del Proyecto Técnico, previo a la obtención del Título de Ingeniero Civil, con el tema: **“HOMOGENEIZACIÓN DE SERIES PLUVIOMÉTRICAS DE LA CUENCA DEL RÍO PASTAZA, UTILIZANDO MÓDULOS DE INTELIGENCIA ARTIFICIAL”**, elaborado por el Sr. **Alexander Adalberto Rogel Rojas**, portador de la cédula de ciudadanía: C.I. 172323292-0, estudiante de la Carrera de Ingeniería Civil, de la Facultad de Ingeniería Civil y Mecánica.

Certifico:

- Que el presente proyecto técnico es original de su autor.
- Ha sido revisado cada uno de sus capítulos componentes.
- Esta concluido en su totalidad.

Ambato, septiembre 2022



Ing. Mg. Galo Wilfrido Núñez Aldás

TUTOR

AUTORÍA DE LA INVESTIGACION

Yo, **Alexander Adalberto Rogel Rojas**, con C.I. 172323292-0 declaro que todas las actividades y contenidos expuestos en el presente Proyecto Técnico con el tema **“HOMOGENEIZACIÓN DE SERIES PLUVIOMÉTRICAS DE LA CUENCA DEL RÍO PASTAZA, UTILIZANDO MÓDULOS DE INTELIGENCIA ARTIFICIAL”**, así como también los análisis estadísticos, gráficos, conclusiones y recomendaciones son de mi exclusiva responsabilidad como autor del proyecto, a excepción de las referencias bibliográficas citadas en el mismo.

Ambato, septiembre 2022



Alexander Adalberto Rogel Rojas

C.I. 172323292-0

AUTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que haga de este Proyecto Técnico o parte de él, un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos en línea patrimoniales de mi Proyecto Técnico, con fines de difusión pública, además apruebo la reproducción de este documento dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor.

Ambato, septiembre 2022



Alexander Adalberto Rogel Rojas

C.I. 172323292-0

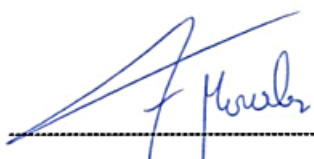
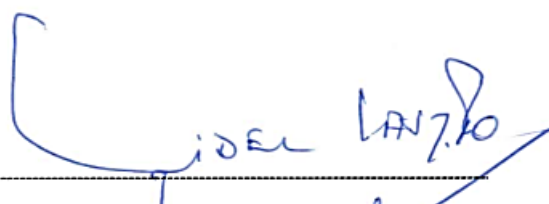
AUTOR

APROBACIÓN DEL TRIBUNAL DE GRADO

Los miembros del Tribunal de Grado aprueban el informe del Proyecto Técnico, realizado por el estudiante Alexander Adalberto Rogel Rojas de la Carrera de Ingeniería Civil bajo el tema: **“HOMOGENEIZACIÓN DE SERIES PLUVIOMÉTRICAS DE LA CUENCA DEL RÍO PASTAZA, UTILIZANDO MÓDULOS DE INTELIGENCIA ARTIFICIAL”**.

Ambato, septiembre 2022

Para constancia firman:

	
-----	-----
Ing. Mg. Fabian Rodrigo Morales Fiallos	Ing. Mg. Fidel Alberto Castro Solórzano
MIEMBRO CALIFICADOR	MIEMBRO CALIFICADOR

DEDICATORIA

El presente trabajo está dedicado a mi hermana Nayeli Rogel, quien desde que abrió los ojos inspiró valentía y fortaleza para enfrentar mis miedos, me acompañó en cada sueño depositando su confianza y amor. Gracias por guardar mi historia y direccionar mi vida hasta convertirme en la persona que soy.

A mi pequeña hermana Valeria Rojas, quien con su inocencia y ternura inspiró mis mejores virtudes y me enseñó a amar incondicionalmente a mi familia.

A mi padre Justo Rogel, quien con su gran amor confió en mí y con su esfuerzo me ha brindado todas las herramientas necesarias para alcanzar mi meta, por enseñarme a jamás rendirme y darlo todo para conseguir mis sueños, por llamarme cada día y estar pendiente de cómo me siento.

A mi madre Alexandra Rojas, quien me ha acompañado la mayor parte de mi vida, por enseñarme que en los momentos difíciles podemos sacar nuestra mejor versión, por aceptarme como soy y confiar en mí dándome valor cuando me he querido rendir.

A mi tía María Rojas, quien supo que este día llegaría hace mucho tiempo atrás, por estar en mi infancia y haber sido un pilar fundamental en mi ideología de vida, por arrullarme en tus brazos cuando huía de mi casa y a pesar de tu pronta partida, tu memoria está presente en todo lo que hago.

A mis abuelitos paternos Antoliano Gonzaga y Digna Sarango, con quienes crecí parte de mi infancia y me cuidaron con amor, por hacerme un amante del café e inculcarme una de mis pasiones más preciadas, la guitarra.

A mis abuelitos maternos Nelson Rojas y Ayda León, quienes han celebrado mis triunfos y me han llenado de amor en cada visita, por haberme permitido vivir con ellos durante mi etapa final de la carrera.

Alexander Rogel

AGRADECIMIENTO

A la facultad de Ingeniería Civil y Mecánica de la Universidad Técnica de Ambato, durante mi estadía como estudiante he gozado de gratos momentos mientras recibía la instrucción que permitió mi formación profesional.

Al ingeniero Andrés Hidalgo, quien gracias a su paciencia y conocimiento fue pilar fundamental del desarrollo de esta investigación.

A mi tutor Ing. Galo Núñez, con su tiempo, paciencia y experiencia permitió que este trabajo llegara a su culminación.

A mi hermano de otra mamá César, quién me acompañó y alentó en los momentos más difíciles durante mi preparación profesional, por nunca negarme una conversación y socorrer cuando más necesitaba a alguien, por ser la gran persona que es y con quién estaré en deuda toda mi vida.

A mi mejor amigo Bryan, quién me acompañó durante todo este proceso y siempre supo que decirme para levantarme, por ayudarme en mis momentos difíciles y compartir los mejores momentos en mi vida universitaria.

A mis grandes amigas Andrea, Myrka y Cynthia, con quienes exploré el lado más sincero de la amistad, quienes confiaron en mi durante todo este proceso.

A la persona con el corazón más grande que conozco, mi gran amigo Álvaro Ortiz, quien me dio la mano y abrió sus puertas para formarme tanto personalmente como profesionalmente, una de las mentes más hábiles que conozco y mi mentor en la ingeniería.

A mi amigo Daniel, quien con su paciencia y conocimiento ayudó en la revisión de este trabajo.

ÍNDICE GENERAL DE CONTENIDOS

CERTIFICACIÓN	ii
AUTORÍA DE LA INVESTIGACION	iii
DERECHOS DE AUTOR	iv
APROBACIÓN DEL TRIBUNAL DE GRADO	v
DEDICATORIA	vii
AGRADECIMIENTO	viii
RESUMEN	xviii
ABSTRACT	xix
CAPÍTULO I	1
1.1 Antecedentes Investigativos.....	1
1.1.1 Antecedentes.....	1
1.1.2 Justificación.....	3
1.1.3 Fundamentación teórica.....	4
1.1.3.1 Ciclo del agua.....	4
1.1.3.2 Ecuación del balance hídrico.....	6
1.1.3.3 Sistema hidrológico.....	7
1.1.3.4 Cuenca hidrográfica.....	10
1.1.3.4.1 Características geomorfológicas de la cuenca.....	12
1.1.3.4.2 Sistema de Drenaje.....	13
1.1.3.4.3 Relieve.....	15
1.1.3.5 Precipitación.....	18
1.1.3.5.1 Tipos de precipitación.....	19
1.1.3.5.2 Variabilidad temporal y espacial.....	21
1.1.3.5.3 Mediciones Hidrológicas.....	21
1.1.3.5.4 Registros pluviométricos.....	22
1.1.3.5.5 Verificación de registros pluviométricos y meteorológicos.....	24
1.1.3.5.5.1 Causas de errores en las series pluviométricas.....	24
1.1.3.5.5.2 Problemas de las series.....	25
1.1.3.5.5.3 Causas de la heterogeneidad en series climáticas.....	25
1.1.3.5.5.4 Homogeneidad de datos pluviométricos.....	26

1.1.3.6	Inteligencia artificial (IA)	26
1.1.3.6.1	Ramas de la inteligencia artificial	27
1.1.3.6.2	Machine Learning (ML) o Aprendizaje Automatizado	27
1.1.3.6.3	Partes del Machine Learning	28
1.1.3.6.4	Deep learning o Aprendizaje profundo	29
1.1.3.6.5	Algoritmos principales del Deep Learning	29
1.1.3.6.6	Red neural artificial (RNA)	30
1.1.3.6.7	Ventajas de las redes neuronales artificiales	32
1.1.3.7	Lenguajes de programación usados para la inteligencia artificial	33
1.1.3.8	Lenguaje de programación Phyton	33
1.1.3.8.1	Librerías de Python	34
1.1.4	Hipótesis	35
1.1.4.1	Hipótesis alternativa	35
1.1.4.2	Hipótesis nula	35
1.2	Objetivos	36
1.2.1	Objetivo General	36
1.2.2	Objetivos Específicos	36
CAPÍTULO II		37
2.1	Materiales	37
2.2	Métodos	37
2.2.1	Nivel o tipo de investigación	37
2.2.2	Población y muestra	38
2.2.3	Operacionalización de las variables	39
2.2.3.1	Variable independiente	39
2.2.3.2	Variable dependiente	40
2.2.4	Plan de recolección de la información	41
2.2.5	Plan de procesamiento y análisis de la información	46
CAPÍTULO III		47
3.1	Recolección y tratamiento de datos	47
3.2	Diseño del software de estimación basado en redes neuronales artificiales	51
3.2.1	Arquitectura del código de estimación	51

3.2.2	Arquitectura de la red neuronal artificial.....	51
3.3	Descripción de la codificación	52
3.2.2.1	Importación de librerías	53
3.2.2.2	Importación de datos de entrada	53
3.2.2.3	Fase de entrenamiento.....	54
3.2.2.4	Obtención y exportación de datos faltantes de series pluviales	56
3.3	Evaluación y calibración del código	119
3.4	Homogeneización de la Cuenca del Río Pastaza	124
3.4.1	Estimación de datos de la estación meteorológica Totorillas	126
3.4.2	Estimación de datos de la estación meteorológica Cebadas	127
3.4.3	Estimación de datos de la estación meteorológica Alao	128
3.4.4	Estimación de datos de la estación meteorológica Guaslán.....	130
3.4.5	Estimación de datos de la estación meteorológica Urbina.....	131
3.4.6	Estimación de datos de la estación meteorológica Tisaleo	132
3.4.7	Estimación de datos de la estación meteorológica Pillaro	134
3.4.8	Estimación de datos de la estación meteorológica Pastocalle.....	135
3.4.9	Estimación de datos de la estación meteorológica Patate	136
3.4.10	Estimación de datos de la estación meteorológica Cotopilalo	138
3.5	Homogeneización de la Cuenca del Río Pastaza por método manual	139
3.5.1	Estimación de datos de la estación meteorológica Totorillas	140
3.5.2	Estimación de datos de la estación meteorológica Cebadas	141
3.5.3	Estimación de datos de la estación meteorológica Alao	142
3.5.4	Estimación de datos de la estación meteorológica Guaslán.....	143
3.5.5	Estimación de datos de la estación meteorológica Urbina.....	144
3.5.6	Estimación de datos de la estación meteorológica Tisaleo	145
3.5.7	Estimación de datos de la estación meteorológica Píllaro	147
3.5.8	Estimación de datos de la estación meteorológica Pastocalle.....	149
3.5.9	Estimación de datos de la estación meteorológica Patate	150
3.5.10	Estimación de datos de la estación meteorológica Cotopilalo	151
3.6	Análisis de resultados.....	153
3.6.1	Coefficiente de determinación (R^2).....	155
3.6.2	Prueba de rachas	156
3.6.3	Correlación de rangos de Spearman	160

3.7 Verificación de la hipótesis.....	164
CAPÍTULO IV.....	165
4.1 Conclusiones.....	165
4.2 Recomendaciones.....	166
Referencias bibliográficas.....	167
Anexos.....	170
Información levantada de las cuencas del Ecuador disponibles en el año 2009.....	170
Código.....	188

ÍNDICE DE TABLAS

Tabla 1.	Cantidades estimadas de agua en el mundo.....	5
Tabla 2.	Lenguajes de programación con soporte en inteligencia artificial	33
Tabla 3.	Operacionalización de la variable independiente.	39
Tabla 4.	Operacionalización de la variable Dependiente.....	40
Tabla 5.	Datos pluviométricos de la cuenca del rio Pastaza.	42
Tabla 6.	Estaciones disponibles en el año 2009 de la cuenca del Río Pastaza.	47
Tabla 7.	Listado de estaciones que no se poseen datos pluviométrica.	120
Tabla 8.	Resumen de estaciones meteorológicas de la cuenca del río Pastaza....	124
Tabla 9.	Resultados de la estación Totorillas	126
Tabla 10.	Resultados de la estación Cebadas.....	127
Tabla 11.	Resultados de la estación Alao.....	128
Tabla 12.	Resultados de la estación Guaslán.	130
Tabla 13.	Resultados de la estación Urbina	131
Tabla 14.	Resultados de la estación Tisaleo.....	132
Tabla 15.	Resultados de la estación Píllaro.....	134
Tabla 16.	Gráfica estimada de la estación Píllaro.	135
Tabla 17.	Resultados de la estación Pastocalle.	135
Tabla 18.	Resultados de la estación Patate.....	136
Tabla 19.	Resultados obtenidos de la estación Cotopilalo.....	138
Tabla 20.	Resultados con método manual de la estación Totorillas.	140
Tabla 21.	Resultados con método manual de la estación Cebadas.	142
Tabla 22.	Resultados con método manual de la estación Alao.	143
Tabla 23.	Resultados con método manual de la estación Guaslán.....	144
Tabla 24.	Resultados con método manual de la estación Urbina.....	145
Tabla 25.	Resultados con método manual de la estación Tisaleo.....	146

Tabla 26.	Resultados con método manual de la estación Píllaro.	148
Tabla 27.	Resultados con método manual de la estación Pastocalle.....	149
Tabla 28.	Resultados con método manual de la estación Patate.	151
Tabla 29.	Resultados con método manual de la estación Cotopilalo.	152
Tabla 30.	Resumen de resultados de homogeneización.	153
Tabla 31.	Coefficiente de determinación R^2	155
Tabla 32.	Datos mensuales de la estación totorillas.....	156
Tabla 33.	Cálculo N1 y N2.....	157
Tabla 34.	Valores críticos para el número de rachas.....	159
Tabla 35.	Resumen de resultados de Prueba de Rachas.....	160
Tabla 36.	Estaciones disponibles en el año 2009 de la cuenca del Río Arenillas.	170
Tabla 37.	Estaciones disponibles en el año 2009 de la cuenca del Río Balao. ..	170
Tabla 38.	Estaciones disponibles en el año 2009 de la cuenca del Río Cañar. ..	171
Tabla 39.	Estaciones disponibles en el año 2009 de la cuenca del Río Carchi..	171
Tabla 40.	Estaciones disponibles en el año 2009 de la cuenca del Río Cayapas.	171
Tabla 41.	Estaciones disponibles en el año 2009 de la cuenca del Río Catamayo.	172
Tabla 42.	Estaciones disponibles en el año 2009 de la cuenca del Río Chone. .	173
Tabla 43.	Estaciones disponibles en el año 2009 de la cuenca del Río Esmeraldas.	174
Tabla 44.	Estaciones disponibles en el año 2009 de la cuenca del Río Guayas.	176
Tabla 45.	Estaciones disponibles en el año 2009 de la cuenca del Río Jama. ...	179
Tabla 46.	Estaciones disponibles en el año 2009 de la cuenca del Río Jipijapa.	179
Tabla 47.	Estaciones disponibles en el año 2009 de la cuenca del Río Jubones.	180
Tabla 48.	Estaciones disponibles en el año 2009 de la cuenca del Río Mira.....	181

Tabla 49.	Estaciones disponibles en el año 2009 de la cuenca del Río Muisne.	182
Tabla 50.	Estaciones disponibles en el año 2009 de la cuenca del Río Napo....	183
Tabla 51.	Estaciones disponibles en el año 2009 de la cuenca del Río Portoviejo.	184
Tabla 52.	Estaciones disponibles en el año 2009 de la cuenca del Río Puyango.	184
Tabla 53.	Estaciones disponibles en el año 2009 de la cuenca del Río Rosa.....	185
Tabla 54.	Estaciones disponibles en el año 2009 de la cuenca del Río Santiago.	185
Tabla 55.	Estaciones disponibles en el año 2009 de la cuenca del Río Taura. ..	187
Tabla 56.	Estaciones disponibles en el año 2009 de la cuenca del Río Zapotal.	187

ÍNDICE DE FIGURAS

Figura 1.	El ciclo hidrológico.....	6
Figura 2.	Representación conceptual del sistema hidrológico global.	8
Figura 3.	Clasificación de modelos hidrológicos.	10
Figura 4.	Cuenca hidrográfica y partes que la componen.	11
Figura 5.	Orden de las corrientes de una Cuenca.	13
Figura 6.	Extensión media de la esorrentía superficial.....	15
Figura 7.	Sinuosidad de la corriente.....	15
Figura 8.	Método de las cuadrículas.	16
Figura 9.	Curva hipsométrica.	17
Figura 10.	Pendiente de la corriente de una cuenca hidrográfica.....	18
Figura 11.	Precipitación de convección.....	19
Figura 12.	Precipitación orográfica.	20
Figura 13.	Precipitación ciclónica.	20
Figura 14.	Pluviómetro.....	22
Figura 15.	Pluviógrafo.....	23
Figura 16.	Ejemplo de hietograma.	24
Figura 17.	Diagrama que muestra como las diferentes partes de un modelo de inteligencia artificial, se relacionan entre si dentro de disciplinas diferentes.....	30
Figura 18.	Algunos ejemplos del flujo que tienen las redes neuronales artificiales.	31
Figura 19.	Diagrama de una neurona artificial.	32
Figura 20.	Metodología para la elaboración del código de estimación de datos...	45
Figura 21.	Ubicación de las estaciones meteorológicas en la cuenca Río Pastaza.	50
Figura 22.	Diagrama de flujo del programa de estimación de datos.	51

Figura 23.	Perceptrón simple.....	52
Figura 24.	Perceptrón multicapa.....	52
Figura 25.	Algoritmo para la importación de datos de entrada.	54
Figura 26.	Algoritmo del entrenamiento y calibración de la RNA.....	55
Figura 27.	Algoritmo para determinar el modelo predictivo.	56
Figura 28.	Algoritmo para completar datos ausentes.	119
Figura 29.	Convergencia con 100 iteraciones del modelo predictivo	122
Figura 30.	Convergencia con 500 iteraciones del modelo predictivo.	122
Figura 31.	Convergencia con 1000 iteraciones del modelo predictivo.	123
Figura 32.	Convergencia con 5000 iteraciones del modelo predictivo.	123
Figura 33.	Gráfica estimada de la estación Totorillas.	127
Figura 34.	Gráfica estimada de la estación Cebadas.	128
Figura 35.	Gráfica estimada de la estación Alao.	129
Figura 36.	Gráfica estimada de la estación Guaslán.....	131
Figura 37.	Gráfica estimada de la estación Urbina.....	132
Figura 38.	Gráfica estimada de la estación Tisaleo.	133
Figura 39.	Gráfica estimada de la estación Pastocalle.....	136
Figura 40.	Gráfica estimada de la estación Patate.	137
Figura 41.	Gráfica estimada de la estación Cotopilalo.....	139
Figura 42.	Disposición espacial para la completación por razones de distancia.	139
Figura 43.	Determinación del número de rachas.	158

RESUMEN

En la presente tesis se investigó un enfoque de Machine Learning en el campo de la hidrología, para ello se indagó los problemas de heterogeneidad en los anuarios de las estaciones meteorológicas publicados por el Instituto Nacional de Meteorología e Hidrología (INAMHI), en donde se encontró la ausencia de valores en las series pluviométricas.

Para completar los datos faltantes generalmente se hace uso de un método convencional que engloba características generales de una serie pluviométrica, por esta razón en el presente trabajo se hace uso de un algoritmo en función de una red neuronal artificial entrenada para emular el comportamiento de los datos de precipitación registrados a través de la librería scikit-learn en un software libre de programación.

El rendimiento de la red neuronal artificial se evaluó en función de la estimación de los datos pluviométricos faltantes de las estaciones meteorológicas ubicadas en la cuenca del río Pastaza, en donde se analizó el error cuadrático obtenido con el algoritmo y con el método convencional. Para validar el funcionamiento del software creado, se obtuvo la correlación por rangos de los valores calculados de ambos métodos, lo que permitió determinar la eficacia que el software basado en machine learning tiene.

Palabras clave: Precipitación, Estaciones meteorológicas, Inteligencia artificial, Machine learning, Red neuronal artificial.

ABSTRACT

In the present work a Machine Learning approach was investigated in the field of hydrology. For this purpose, were examined the existing problems of heterogeneity in the yearbooks of meteorological stations published by the National Institute of Meteorology and Hydrology (INAMHI), where the absence of values in the pluviometric series was found.

In order to complete the missing data, it was used a conventional method which interpolates pluviometric data between stations in function of their distances. For this reason, in the present work was used an algorithm that contains an artificial neural network (ANN) trained to emulate the behavior of the precipitation data recorded through the scikit-learn library in a free programming software.

The performance of the artificial neural network was evaluated based on the estimation of the missing pluviometric data from the meteorological stations located in the Pastaza river watershed, where the quadratic error obtained was analyzed both with the algorithm and the conventional method. To validate the running of the created software, it was obtained the correlation by ranges of the calculated values by both methods, which allowed determining the effectiveness of the machine learning based software.

Keywords: Precipitation, weather station, artificial intelligence, Machine learning, artificial neural network

CAPÍTULO I

MARCO TEÓRICO

1.1 Antecedentes Investigativos

1.1.1 Antecedentes

Una investigación realizada en 2020 por especialistas en hidrología de la Universidad Islámica de Azad quienes utilizaron métodos meteorológicos e hidrológicos para la determinación de datos de precipitación faltantes demostró que se puede obtener un registro completo de datos hidrológicos, con el cual se optimiza la realización de un proyecto, ya que reduce el tiempo de estudio y los costes de operación.

Las metodologías que utilizaron fueron: Algoritmo genético, Redes neuronales artificiales y algoritmo de avance/retroceso. El Algoritmo genético parte de crear una población inicial aleatoria y evaluar cada individuo de la población hasta alcanzar el número máximo de generaciones creando funciones de ajuste y utilizando variables de entrada y salida con las cuales se determinaron coeficientes de correlación. Las redes neuronales artificiales utilizan tres capas interconectadas (ingreso de datos, proceso y resultados) para predecir los datos de precipitación faltantes, y el algoritmo de avance/retroceso utiliza iteraciones que se caracterizan por su rápida convergencia para obtener la estimación de datos de precipitación[1].

El estudio “Manejo de problemas de datos faltantes con métodos de muestro” propone un proceso para la estimación de valores faltantes en función de métodos de muestreo para solucionar problemas comunes en la toma de datos debido a un equipo defectuoso, falta de un valor porque se perdió o errores humanos al registrar una respuesta, la técnica que se utiliza es la de imputación de datos faltantes, la cual es una excelente estrategia para completar el valor faltante en los datos con un valor plausible que es la estimación del valor verdadero de la observación.

Los tipos de imputación que utilizan son Imputación media (reemplaza un valor faltante con la media general de la muestra), Imputación de regresión (reemplaza cada valor faltante por un dato pronosticado utilizando una regresión múltiple) e Imputación múltiple (reemplaza cada valor faltante con un conjunto de valores plausibles extraídos

de su distribución predictiva). Con estas técnicas finalmente se obtienen los resultados que demuestran una eficacia hasta del 95% para obtener un registro completo de datos hidrológicos[2].

En la Universidad Nacional de San Agustín se realizó una investigación que muestra la estimación de caudales medios diarios en un tramo de la cuenca Amazonas que se fueron muy similares a los datos reales. Para ello utilizaron un modelo difuso que es la combinación de redes neuronales (que simulan el funcionamiento del cerebro para la solución de problemas) y lógica difusa (que utiliza inferencia para producir resultados).

Este modelo difuso parte de obtener información acerca de caudales de precipitación diaria de las fuentes de datos hidrológicos brindados por la Administración Nacional Oceánica y Atmosférica, y a través algoritmos de aprendizaje estudiaron 32 escenarios para obtener los coeficientes de correlación. De los escenarios estudiados, el mejor resultó ser el escenario 32 que cumple con resultados muy favorables como, por ejemplo, el menor error medio absoluto (0.0366), menor error porcentual absoluto (9.2452%), mayor coeficiente de correlación de Pearson (0.9937) y mayor coeficiente de correlación de concordancia de 0.9937[3].

En la Escuela de la Tecnología de la Información de la Universidad de Murdoch, un estudio realizado en 2012 utilizó redes neuronales artificiales modulares para la estimación de registros de precipitación faltantes, es decir los datos de lluvia mensuales que faltan en 4 estaciones ubicadas en la región noreste de Tailandia. El método propuesto se divide en dos pasos: el primero es el de agrupamiento difuso, el cual utiliza dos redes neuronales que relacionan las precipitaciones registradas y las divide en periodos secos y húmedos según la naturaleza de los datos y el segundo es crear un módulo de agregación cuya función es la determinación de los datos de lluvia que faltan. El estudio seleccionó las tres estaciones de control más cercanas para estimar los datos faltantes en la estación objetivo[4].

1.1.2 Justificación

La cuenca del Rio Pastaza es denominada una de las cuencas de drenaje más importantes del Ecuador, ya que representa el 45% de las producciones de hortalizas a nivel nacional. Nace de la unión del rio Chambo y del rio Patate, los mismos que desembocan en el rio Marañón (afluente principal del rio Amazonas). Posee gran variedad de biodiversidad de fauna y flora, debido a que su territorio este contenido por la región interandina y la región amazónica, lo que le permite gozar de 3 pisos climáticos (subtropical, templado y alto andino). Además, es considerada una de las fuentes de almacenamiento de agua dulce más grandes del mundo, cuyo potencial hídrico tiene gran relevancia para la generación de energía[5].

Dentro de los registros de las estaciones de recolección de datos meteorológicos, que se obtienen en línea del Instituto Nacional de Meteorología e Hidrología del Ecuador (INAMHI), se puede observar que la información está incompleta y parte de esa información es inconsistente. Estos errores pueden ser ocasionados por la toma incorrecta de las medidas, variabilidad del instrumento utilizado o cambios en el entorno, provocando que los datos obtenidos de los instrumentos no sean óptimos para ser utilizados, ya que se pueden obtener conclusiones erróneas[6].

Todo registro de series pluviales debe pasar por un proceso de calidad de homogeneidad, para poder normalizar los datos y que estos puedan ser aplicados para la obtención de un inventario de recursos hídricos, los que a su vez serán útiles para proyectos de riego y abastecimiento de agua, así como investigaciones de cambio de variabilidad climática, en zonas donde existen relieves muy irregulares que provocan fenómenos que cambian con el tiempo.

Existen diversos métodos para tratar la heterogeneidad de los registros pluviales, partiendo desde métodos clásicos que están basados en regresión lineal y cálculo de la normal, así como nuevos métodos que hacen uso de regresión difusa a través de una red neuronal artificial. Esta última, parte del concepto de “Machine Learning” (Aprendizaje automatizado), el cual ha sido aplicado para encontrar valores faltantes en registros pluviales, en donde se ha observado que este método tiene ventaja con respecto a otros[1].

El uso de las redes neuronales artificiales ha impactado en la última década, expandiéndose cada vez más en diversos campos de investigación, siendo una alternativa que ayuda en la optimización y simulación de procesos complejos. Actualmente, existen estudios donde se han comparado el uso de redes neuronales artificiales modulares (RNAM) y redes neuronales artificiales simples (RNAS) en el campo de la hidrología para completar datos mensuales de una estación objetivo, a partir de los registros pluviales de tres estaciones cercanas. Los resultados obtenidos demostraron que las RNAM tienen una ventaja considerable con respecto a las RNAS y su uso puede mejorar el rendimiento de otros métodos convencionales para la estimación de datos faltantes[3].

1.1.3 Fundamentación teórica

1.1.3.1 Ciclo del agua

Es la serie de procesos que se repiten constantemente, siendo el principal motivo de estudio de la hidrología. El agua se concentra dentro de la hidrósfera extendiéndose aproximadamente 15 kilómetros en la atmosfera hasta un kilómetro bajo la corteza terrestre.

Aunque, la definición de ciclo del agua a simple vista resulta simple, es una serie de procesos verdaderamente complejos que se subdividen en muchos más ciclos interrelacionados y se desarrollan a niveles continentales, regionales y locales. Si bien, el volumen de agua permanece constante en el ciclo hidrológico, la distribución no vuelve a ser la misma, debido a que está cambiando de forma continua, ya sea por factores como el crecimiento poblacional que demandan la evolución industrial, y esta según teorías es la causante del calentamiento global[7].

- Reservorios o reservas del agua

El cálculo de la cantidad de agua en la tierra, es una interrogante que se ha debatido desde la segunda mitad del siglo XIX. Cuantitativamente la información es escasa, principalmente al hablarse del océano, esto significa que parte de la información que se posee son estimaciones, puesto que no se conocen los valores de forma precisa.

Alrededor de dos terceras partes de agua dulce de la tierra forman parte del hielo polar y el mayor porcentaje está concentrado en ríos subterráneos encontrados desde los 200 a 600 metros de profundidad (por debajo de estas medidas la mayor parte es salina). Únicamente el 0.006% de agua dulce está contenida en ríos[7].

Tabla 1. Cantidades estimadas de agua en el mundo.

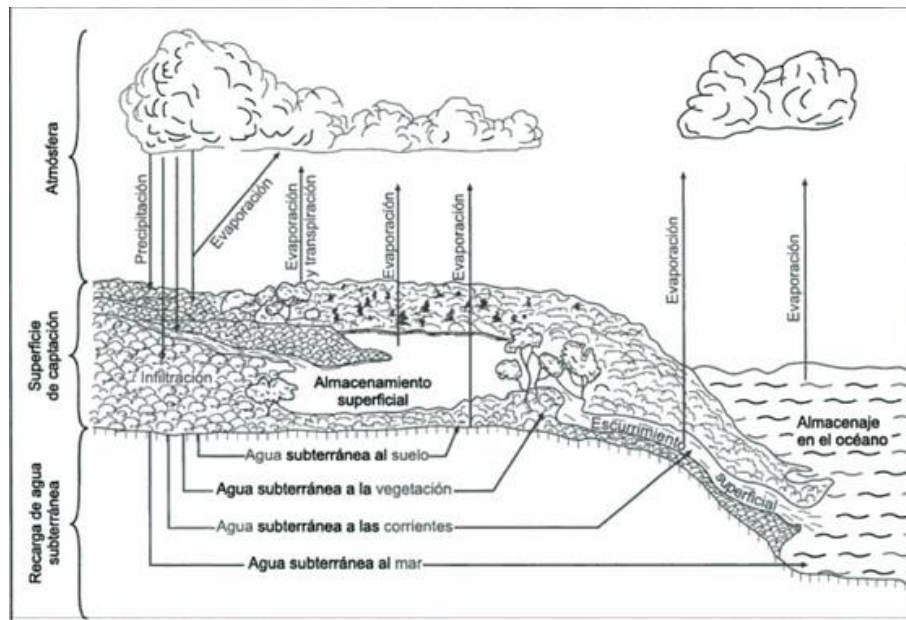
Ítem	Área (10 ⁶ km ²)	Volumen (km ³)	Porcentaje de agua total	Porcentaje de agua dulce
Océanos	361.3	1,338,000,000	96.5	
Agua subterránea				
- Dulce	134.8	10,530,000	0.76	30.1
- Salada	134.8	12,870,000	0.93	
Humedad del suelo	82.0	16,500	0.0012	0.05
Hielo polar	16.0	24,023,500	1.7	68.6
Hielo no polar y nieve	0.3	340,600	0.025	1.0
Lagos				
- Dulces	1.2	91,000	0.007	0.26
- Salinos	0.8	85,400	0.006	
Pantanos	2.7	11,470	0.0008	0.03
Ríos	148.8	2,120	0.0002	0.006
Agua biológica	510.0	1,120	0.0001	0.003
Agua atmosférica	510.0	12,900	0.001	0.04
Agua total	510.0	1,385,984,610	100	
Agua dulce	148.8	35,029,210	2.5	100

Fuente: Balance anual de agua global, UNESCO, 1978[7].

- Procesos del ciclo hidrológico

Dentro de los esquemas que detallan los procesos del ciclo del agua, la evaporación es el medio que lo conduce desde los océanos y superficie terrestre hasta la atmosfera, donde posteriormente el vapor que es transportado se condensará y precipitará hacia la corteza terrestre. El agua precipitada impactará la vegetación, correrá sobre el suelo convirtiéndose en flujo subsuperficial, que se descargará en ríos o lagos como escorrentía superficial o simplemente se infiltrará en el suelo, formando parte de los ríos subterráneos que se exteriorizan en manantiales o se dirigen a ríos para recargar el cauce, su recorrido finaliza en el mar para empezar nuevamente el ciclo hidrológico[7].

Figura 1.El ciclo hidrológico.



Fuente: D. F. C. Aranda, Procesos del ciclo hidrológico[8].

1.1.3.2 Ecuación del balance hídrico

- Definición.

El enfoque principal que se tiene al hacer un análisis hidrológico es el principio de la conservación de la masa, bajo esta consideración se establece que, la diferencia entre

los valores de entrada y los valores de salida del cauce de una cuenca será igual a la variabilidad de almacenamiento[9].

$$I - O = \frac{\Delta S}{\Delta t}$$

I: Entradas

O: Salidas

ΔS: Cambio de almacenamiento

- Componentes de la ecuación.

El componente de entradas (I) se refiere al volumen de agua proveniente de la precipitación, importaciones de agua, escorrentía superficial y aguas subterráneas de otras cuencas. Las salidas (O) harán referencia a los valores que componen la evaporación, transpiración, escorrentía superficial de otras hoyas, exportaciones de agua, aguas subterráneas hacia otras cuencas y la infiltración.

El cambio de almacenamiento (ΔS) hará énfasis al volumen de agua almacenada proveniente de las aguas subterráneas y por el cambio de humedad del suelo[9].

1.1.3.3 Sistema hidrológico

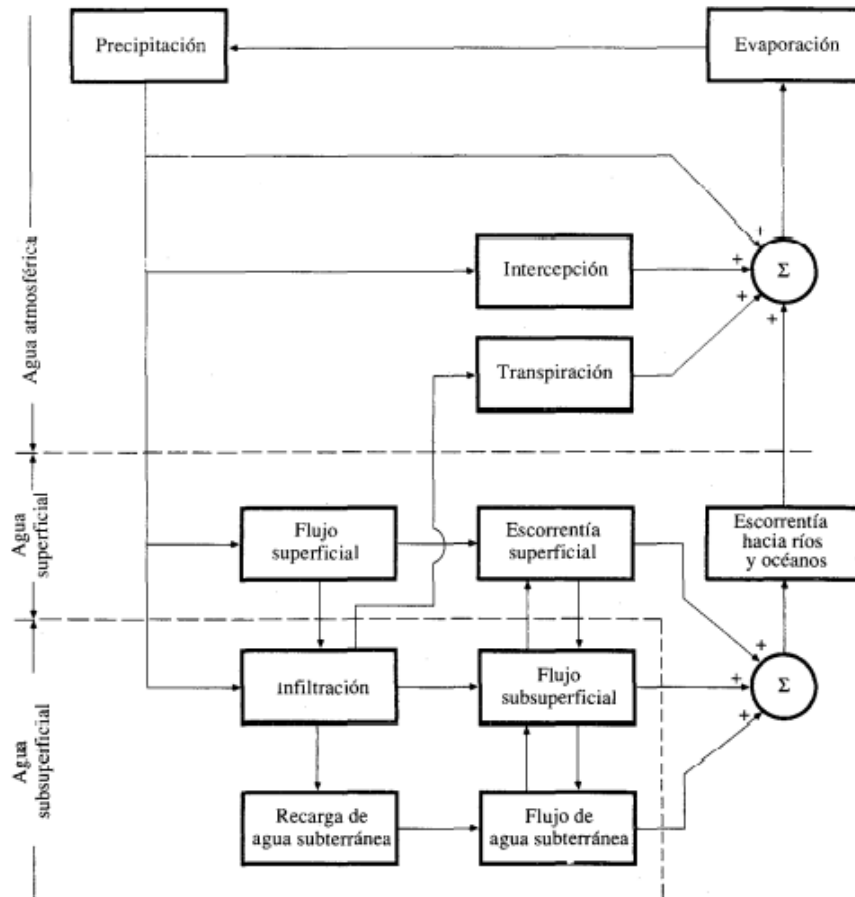
- Definición de Sistema hidrológico y componentes.

Debido a la complejidad que existe para analizar los fenómenos hidrológicos y la falta de un conocimiento absoluto, se lo ha simplificado bajo el concepto de sistema. Sabiendo que, un sistema es un grupo de partes que se encuentran conectadas para formar un todo. Las fases del ciclo hidrológico se consideran como subsistemas para que puedan ser analizadas de forma separada y combinar los resultados en relación a las interacciones que existan entre ellas.

Un sistema hidrológico se define como una estructura o volumen en el espacio, rodeada por una frontera, que acepta agua y otras entradas, opera en ellas internamente y las produce como salidas. La estructura o volumen es el conjunto de todos los

caminos que conforman el flujo, a través de los cuales el agua se transporta hasta el punto de entrada al sistema y continua hasta la salida del mismo. La frontera es la topografía que rodea el sistema y parte de la morfología es la que da la forma a la superficie para generar el sector de entrada en el que el agua entra y durante su paso interactúa dentro del sistema, generando procesos físicos, químicos y biológicos[7].

Figura 2.Representación conceptual del sistema hidrológico global.



Fuente: V. Te Chow, Hidrología aplicada[7].

- Modelo hidrológico

Se considera como la forma más real que se tiene del sistema; las entradas y salidas son variables hidrológicas que pueden ser medidos y valorados, mientras que la estructura es un grupo de ecuaciones cuya función es la de conectar las entradas y salidas[7].

- Clasificación de Modelos

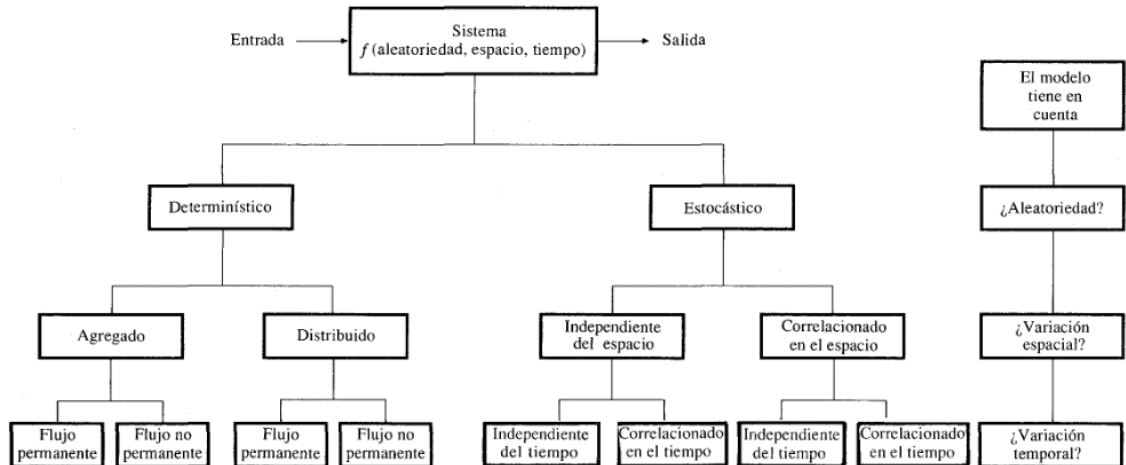
Los modelos pueden clasificarse en modelos físicos y modelos abstractos. Los primeros abarcan los modelos a escala, los cuales hacen una representación del sistema con dimensiones de menor tamaño y los modelos análogos, que usan otro sistema de características físicas que emulan las propiedades de un prototipo.

Los modelos abstractos hacen una representación matemática del modelo, cuyos operadores se describen mediante un conjunto de ecuaciones que buscan relacionar las variables de entrada y salida.

El desarrollo de un modelo que englobe las variables aleatorias dependientes de las 3 dimensiones espaciales y del tiempo es complejo, en consecuencia, se deberá simplificar el modelo despreciando algunas fuentes que causan variabilidad. De acuerdo a la forma en que se realice la simplificación se podrá clasificar a los modelos hidrológicos, existiendo así, tres consideraciones que se deben tomar en cuenta; la aleatoriedad de las variables, la variación en el espacio y la variación en el tiempo.

Los modelos determinísticos hacen pronósticos, por consiguiente, no pueden considerar la aleatoriedad, esto quiere decir que teóricamente la entrada producirá la misma salida. Es el caso contrario a los modelos estocásticos que tienen salidas parcialmente aleatorios por lo que hacen énfasis en predicciones, se clasifican en independientes en el espacio y correlacionados en él, en consideración a la influencia que las variables aleatorias tengan entre ellas en distintos puntos del espacio[7].

Figura 3. Clasificación de modelos hidrológicos.



Fuente: V. Te Chow, Hidrología aplicada[7].

1.1.3.4 Cuenca hidrográfica

- Definición y delimitación de una cuenca hidrográfica.

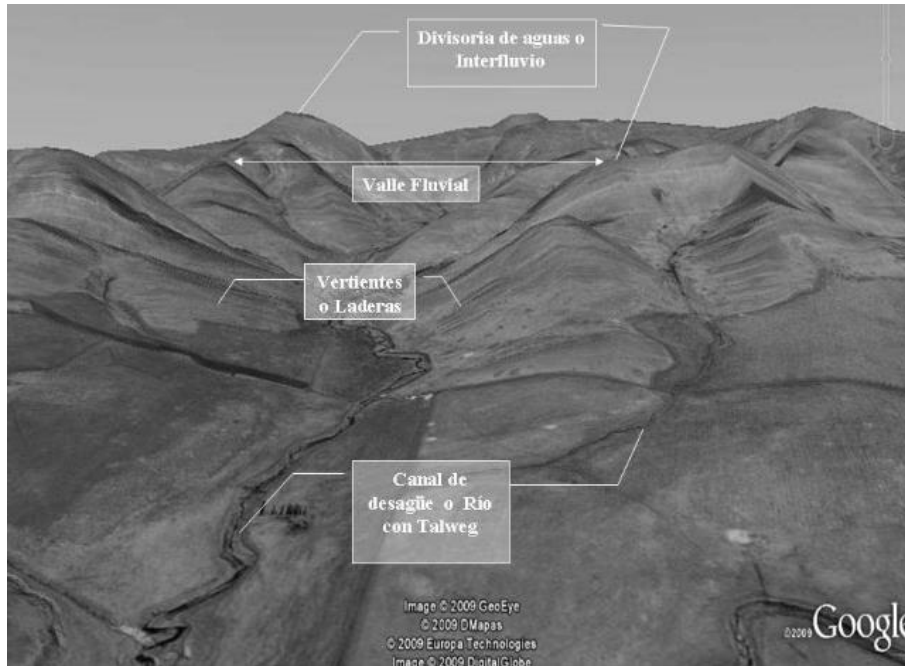
Es el espacio rodeado de las zonas más altas de una montaña, laderas o colinas, donde se discurren de manera natural agua proveniente de precipitaciones, acuíferos o deshielos, creando un drenaje superficial cuyo punto de salida crea un río, y a su vez, este busca desembocarse en un lago o mar.

Las cuencas hidrográficas pueden clasificarse de acuerdo al sistema de drenaje y conducción final. Por las características mencionadas pueden ser arreicas (cuando no desembocan en un río o en el mar, sus aguas desaparecen producto de la evaporación o infiltración), exorreicas (cuando es conducida a un sistema de drenaje más grande como un gran río o mar), criptorreicas (cuando no poseen un sistema de drenaje organizado y se conducen como ríos subterráneos) y endorreicas (cuando llegan a un embalse o lago sin llegar al mar).

La delimitación de una cuenca parte de una proyección en curvas de nivel, siendo el trazo imaginario que la divide de las cuencas contiguas, para analizar la distribución del escurrimiento nacido por las precipitaciones que recorre el sistema y fluye hasta el punto de desembocadura de la cuenca hidrográfica.

La frontera de una cuenca topográfica no siempre tiene la misma proyección horizontal que tiene la cuenca de agua subterránea, por lo que se debería realizar una delimitación que tome en consideración el aporte generado por las aguas subterráneas[10].

Figura 4. Cuenca hidrográfica y partes que la componen.



Fuente F. Gaspari, Elementos metodológicos para el manejo de cuencas hidrográficas[11].

- Líneas divisorias o de cumbres.

Es la línea conceptual que se traza en el contorno de la cuenca hidrográfica, definiendo la división de las adyacentes y permitiendo el transporte o escurrimiento que ocurre durante las precipitaciones, para formar la escorrentía superficial o subsuperficial que es dirigida hasta la salida de la cuenca[12].

- Sistema de Drenaje.

Hace referencia a las direcciones o configuración de los cauces que se conectan entre sí, mismas que permiten la captación del agua desde las zonas altas, de tal manera que se pueda manifestar e interpretar la eficiencia del sistema de drenaje en el escurrimiento resultante. De acuerdo a su forma, se permite obtener una visión sobre las condiciones del suelo y de la superficie de la cuenca[13].

- Punto de Descarga.

Está ubicado en la zona más baja de la cuenca y es el lugar por donde desfoga el caudal final del agua que es recolectada a través de las ramificaciones que alimentan el lecho fluvial de la cuenca.

1.1.3.4.1 Características geomorfológicas de la cuenca

- Área de Drenaje

Es la superficie proyectada sobre un plano horizontal, cuya forma es muy irregular, por consiguiente, no se permite el uso de fórmulas geométricas. Se obtiene luego de que sea trazada la delimitación de la cuenca[13].

- Forma de la Cuenca

Es una de las características que definen de manera importante a la cuenca, puesto que su forma denotara el comportamiento que tenga cuando existan escenarios que provoquen avenidas máximas[14].

- Índice de compacidad o Gravelius.

Es un parámetro adimensional obtenido a partir de la relación entre el perímetro y el valor de la superficie que conforma la cuenca[14].

$$K_c = 0.282 \frac{P}{\sqrt{A}}$$

K_c: Coeficiente de compacidad

P: Perímetro de la cuenca

A: superficie de la cuenca

- Factor de Forma.

Se expresa como la relación entre el ancho promedio y la longitud de la cuenca, es decir:

$$F = \frac{\text{ancho}}{\text{longitud}} = \frac{B}{L}$$

Suponiendo que la distribución de la topografía de la cuenca es rectangular:

$$F = \frac{B \times L}{L \times L} = \frac{A}{L^2}$$

Se debe resaltar que, el factor de forma muestra la posibilidad que existe dentro de una cuenca para concentrar las intensidades de lluvia, puesto que, al ser mayor este factor las probabilidades de una tormenta simultánea en toda la superficie de la cuenca es mayor[13].

1.1.3.4.2 Sistema de Drenaje

- Orden de las corrientes

Es una clasificación que se origina a partir de las bifurcaciones que existen dentro de la cuenca. Para realizar esta clasificación se precisa de un plano que contenga las corrientes que tienen un flujo sin interrupciones, así como las corrientes que son interrumpidas en algún punto de su trayectoria[13].

Figura 5. Orden de las corrientes de una Cuenca.



Fuente: M. V. Béjar, *Hidrología*[13].

- Densidad de drenaje

Es un parámetro que estima la naturaleza del terreno y, además, permite tener una idea sobre la cobertura vegetal de la cuenca. Los valores altos representan zonas con un grado bajo de cobertura vegetal, mientras que, los valores bajos denotan suelos más rígidos con bajo grado de erosión o suelos con alto grado de permeabilidad y cobertura vegetal densa[9].

$$D_d = \frac{L}{A}$$

D_d = densidad de drenaje

L : longitud total de las corrientes perennes o intermitentes

A : área total de la cuenca

- Extensión media de la corriente

Se define como el promedio de la distancia en que el agua de precipitación se escurre sobre la superficie de la cuenca, considerando que este proceso es realizado en línea recta donde la precipitación descendió hasta el punto más cercano al lecho de cualquier corriente de la cuenca.

Para calcular la extensión media de la corriente se debe considerar que la superficie de la cuenca puede representarse como un área de drenaje rectangular, cuya longitud del curso del agua será igual a la longitud total de las corrientes de agua[9].

$$A = 4IL$$

I = Extensión media de la corriente

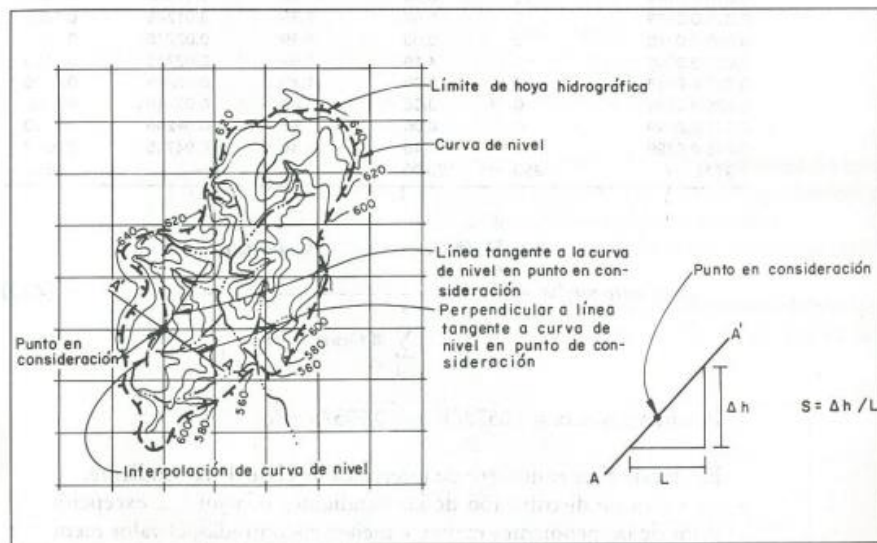
L : longitud total de las corrientes de agua en la cuenca hidrógrafica

A : área de drenaje total de la cuenca

- Método de cuadrículas asociadas o a partir de curvas de nivel

Es el método más complejo para la determinación de los valores que describen las pendientes topográficas de una cuenca, puesto que consiste en la obtención de la distribución en porcentaje de las pendientes topográficas, a través de una muestra estadística de las pendientes normales en función de curvas de nivel de una gran cantidad de puntos dentro de la cuenca que dispondrán el número de cuadrículas a utilizadas.

Figura 8.Método de las cuadrículas.



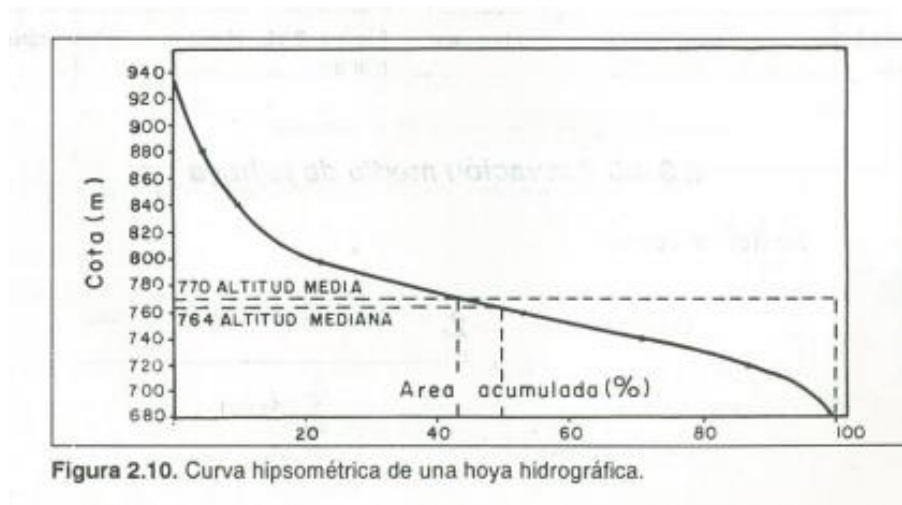
Fuente: Monsalve Sáenz, “Hidrología en la Ingeniería[9].

- Curva Hipsométricas

Se define como la representación gráfica de la topografía de una cuenca, enfocando su estudio en la variabilidad de elevación existente en distintos puntos del terreno de la cuenca, tomando como referencia el nivel medio del mar.

Esta representación define a las elevaciones como parte de las ordenadas y a los valores del área acumulada en las abscisas, relacionando los valores mínimos de elevación encontrados contra el área acumulada correspondiente[9].

Figura 9.Curva hipsométrica.



Fuente: Monsalve Sáenz, “Hidrología en la Ingeniería[9].

- Altura media y mediana de la cuenca

Este parámetro mide la variación en elevación de una cuenca y, es importante porque considera el impacto ejercido sobre la precipitación, sobre las pérdidas de agua por evaporación y transpiración, que tendrán una consecuencia directa en el caudal medio[9].

$$E = \frac{\sum_{i=1}^n (Cota\ media\ intervalo_i \times \text{Área}_i)}{\sum_{i=1}^n (\text{Área}_i)}$$

El valor de “n” hace referencia al número de intervalos.

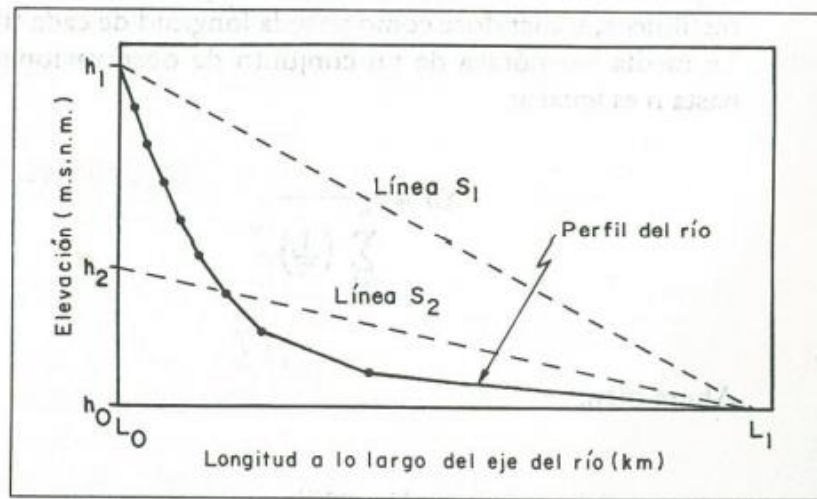
La altura media es la línea que se traza en el valor intermedio de las cotas, mientras que la altura mediana es la línea que se traza en el 50% del área acumulada bajo la curva.

- Pendiente de la corriente principal

La velocidad con la que cursa la escorrentía estará ligada a la pendiente de sus canales fluviales, puesto que, la inclinación del terreno será la principal variable en el análisis de la longitud que recorre el escurrimiento de las corrientes de agua en función del tiempo.

Para este análisis, se debe considerar dos parámetros: Pendiente media S_1 (diferencia total que existe entre la elevación del lecho del río al ser dividido por su longitud entre esos puntos.) y la pendiente media ponderada S_2 (es el valor que se obtiene al dibujar una línea, de tal manera que permita que el área que se encuentre entre la línea y los ejes coordenados sea igual a la comprendida entre la curva del perfil del río y dichos ejes)[9].

Figura 10. Pendiente de la corriente de una cuenca hidrográfica.



Fuente: Monsalve Sáenz, "Hidrología en la Ingeniería"[9].

$$S_1 = \frac{h_1 - h_0}{L_1 - L_0}$$

$$S_2 = \frac{h_2 - h_0}{L_1 - L_0}$$

1.1.3.5 Precipitación

- Definición de precipitación y unidades de medición.

Se define como toda forma de humedad que se origina en las nubes y se deposita en la superficie terrestre.

Su formación sucede por acción del calentamiento y la presencia estímulos de radiación variable, haciendo que el vapor de agua ascienda hasta encontrar la altura

con características de enfriamiento que permita su saturación y, acompañados de la presencia de núcleos de condensación o de congelamiento, provocan que descienda hasta la superficie terrestre, pudiendo presentarse en forma de lluvia, neblina, granizo o nieve. Los núcleos de congelamiento se forman a partir de productos de combustión, óxidos de nitrógeno y partículas de sal diminutas, mientras que los núcleos de congelamiento son conformados por minerales arcillosos como el caolín.

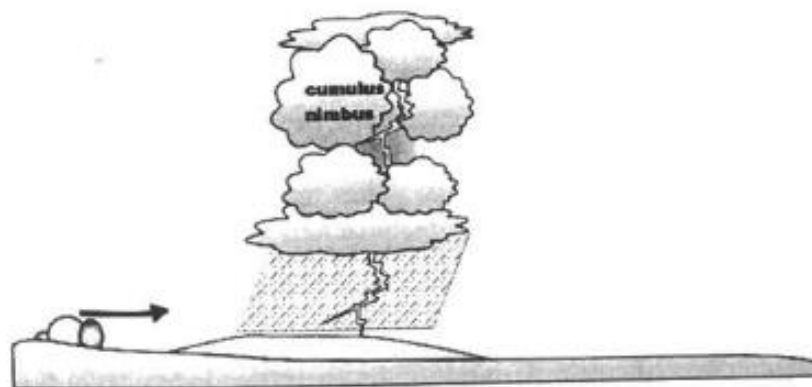
La precipitación se mide en milímetros (mm), o lo que es igual a la relación entre litros y metros cuadrados (L/m²), esto porque la definición del valor corresponde a la altura que se obtiene a partir de la acumulación de precipitación en un metro cuadrado de terreno, en donde el agua contenida no podrá infiltrarse ni evaporarse[15].

1.1.3.5.1 Tipos de precipitación

Se clasifican de acuerdo a las condiciones que producen movimiento vertical del aire:

Las precipitaciones convectivas, suceden cuando la masa de aire que está más cerca de la superficie de la cuenca eleva su temperatura de forma abrupta, por consiguiente, se reduce el valor de su densidad y la masa de aire intenta ascender y de sustituirse por otra masa de aire más densa. Pueden elevarse hasta llegar a grandes alturas, donde existen las condiciones necesarias para entrar en estado de condensación y precipitación[13].

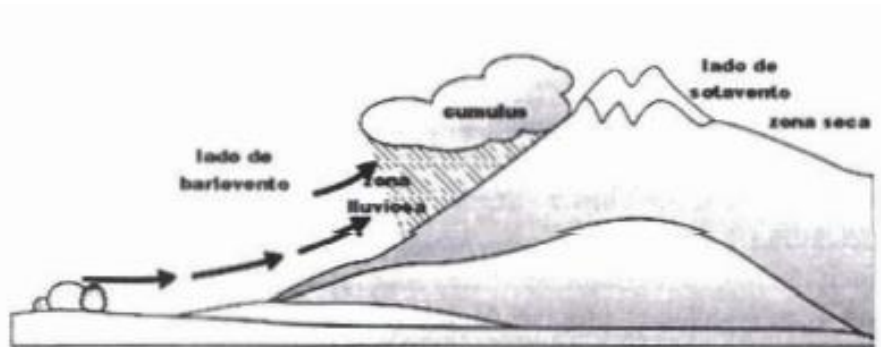
Figura 11. Precipitación de convección.



Fuente. M. V. Béjar, *Hidrología*[13].

Las precipitaciones orográficas, surgen a partir de que la masa de vapor tiene un movimiento horizontal hasta llegar a las barreras naturales que provocan un ascenso mecánico, hasta llegar a la altura que posea las características que permitan su condensación y posterior precipitación[13].

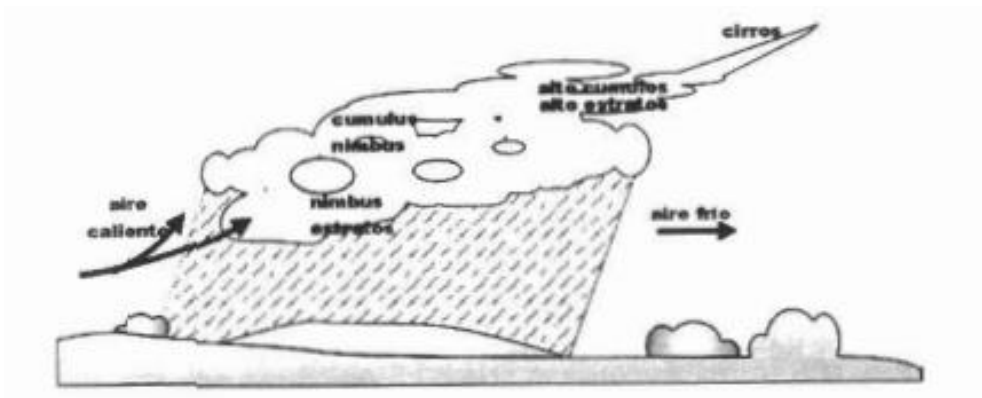
Figura 12. Precipitación orográfica.



Fuente. M. V. Béjar, *Hidrología*[13].

La precipitación ciclónica, es producida a partir del choque de dos masas de aire que, con condiciones de temperatura y humedad distintas, provocando que las masas de vapor más calientes sean impulsadas a zonas más altas donde precipiten[13].

Figura 13. Precipitación ciclónica.



Fuente. M. V. Béjar, *Hidrología*[13].

1.1.3.5.2 Variabilidad temporal y espacial

La precipitación varía en el espacio y el tiempo, por consecuencia, del patrón general de circulación atmosférica y de los factores locales. El valor promedio de la variable climática, luego de un periodo de observación, es conocida como valor normal. En las costas es donde se encontrará la mayor cantidad de humedad, puesto que los océanos tienen el mayor porcentaje de aportación de humedad atmosférica que denotará en la densidad de la precipitación[7].

1.1.3.5.3 Mediciones Hidrológicas

Las mediciones hidrológicas se realizan con el fin de obtener la información necesaria, que permita comprender de mejor manera el comportamiento de los procesos hidrológicos, y a su vez, puedan ser utilizados como datos de entrada en simulaciones hidrológicas que ayuden en la realización de diseños, análisis y toma de decisiones[7].

- **Puntuales**

Son mediciones que se realizan en función del tiempo en lugares fijos del espacio, como en el caso de los pluviómetros que registran la información de precipitación desde un punto específico de la cuenca, para formar una serie de tiempo que puede ser utilizada en análisis estadístico[7].

- **Distribuidas**

Son muestras que se toman a lo largo de una línea o área en el espacio en un momento específico del tiempo. Se toma como ejemplo la medición de la cubierta de nieve, que se hace a través de un aeroplano que mide la radiación reflejada por el campo de nieve[7].

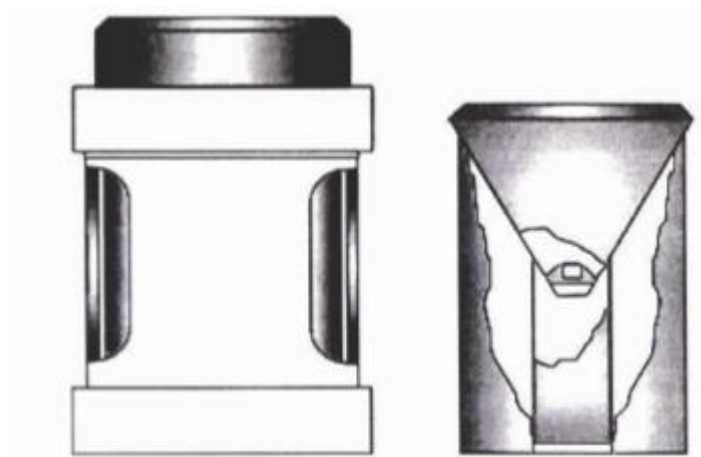
1.1.3.5.4 Registros pluviométricos

- Métodos y equipos de medición.

Los aparatos de medición son instrumentos que están expuestos a la intemperie, para registrar la altura de precipitación a través de un cilindro abierto en la parte superior. Estos aparatos se clasifican en función del registro de precipitaciones en pluviómetros y pluviógrafos.

El pluviómetro es un recipiente de lámina en forma de cilindro, que tiene 20 cm de diámetro y 60 cm de alto aproximadamente. La tapa está compuesta por un embudo receptor, que está conectado con una probeta de sección 10 veces menor que la de la tapa, en donde se mide la altura de la lluvia con una aproximación de decimos de milímetro[13].

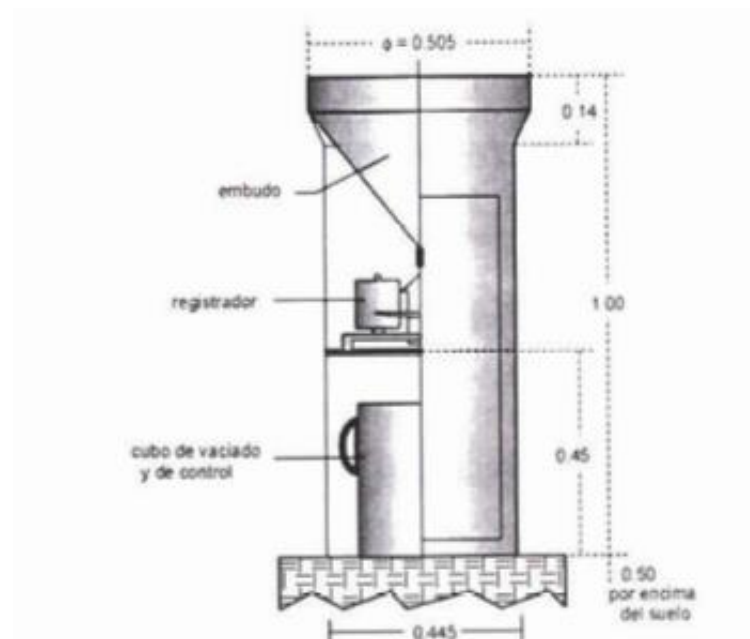
Figura 14.Pluviómetro.



Fuente. M. V. Béjar, *Hidrología*[13].

El pluviógrafo es un instrumento que registra la altura de precipitación en función del tiempo, misma que permite encontrar la intensidad de la precipitación (Dato importante para el diseño de proyectos hidráulicos). Son de forma cilíndrica, y el embudo receptor está conectado a un sistema de flotadores, que dan paso a la oscilación de una aguja sobre un papel registrador, ensamblado en un sistema de reloj[13].

Figura 15.Pluviógrafo.

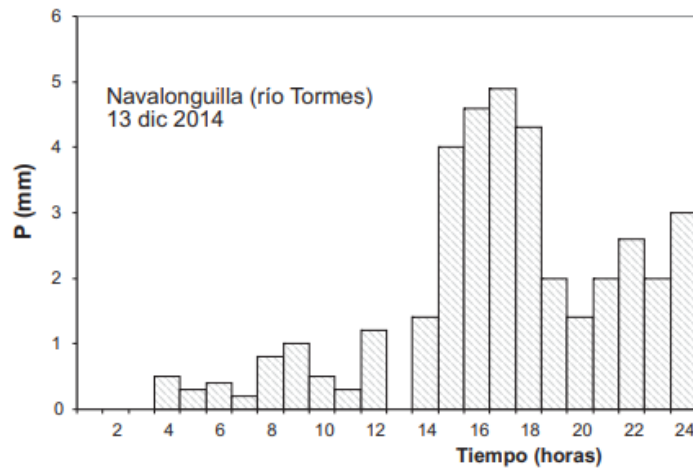


Fuente. M. V. Béjar, *Hidrología*[13].

- Hietograma de precipitación y características.

Son gráficos que pueden representarse mediante barras o línea, pudiendo disponer datos mensuales o anuales, así como diarios, siempre y cuando se tengan los intervalos necesarios para su representación. En el eje de las abscisas se colocará la duración del fenómeno pluvial y en el eje de las ordenadas, el volumen de precipitación medida en milímetros[16].

Figura 16.Ejemplo de hietograma.



Fuente: F. J. Sánchez San Román, Hidrología superficial y subterránea[16].

1.1.3.5.5 Verificación de registros pluviométricos y meteorológicos

1.1.3.5.5.1 Causas de errores en las series pluviométricas

- Medición.

Se debe al registro de datos mal efectuado, puede ser ocasionado por error humano o por defectos de fabrica que tenga el aparato.

- Digitalización.

Este problema puede originarse por una mala calibración de software o defectos en la transmisión debidos a factores como; señal, componentes electrónicos, electricidad, etc.

- Problemas electrónicos

Abarca todos los inconvenientes que tengan relación con el hardware del instrumento, derivados de errores humanos o defectos de fábrica.

- Cambio de posición de la estación

Sucede cuando las estaciones cambian de lugar, usualmente son desplazamientos pequeños, aunque este factor está ligado directamente al relieve del país, si los relieves son muy contrastados podrían generar una variación significativa en los registros de datos.

1.1.3.5.5.2 Problemas de las series

- Datos faltantes

Los datos faltantes pueden originarse por falta de profesionalismo del operador o la ausencia del equipo durante un periodo de tiempo. Los valores ausentes deben ser completados, debido a que en la hidrología se trabajó con series continuas[9].

- Valores sospechosos

Se refiere a la interrogante generada por los valores registrados que, a simple vista, se pueden observar fuera de la tendencia ya denotada entre la información obtenida.

- Comportamientos sospechosos

Son los datos que, al ser proyectados en una gráfica se puede observar que salen de los límites ya definidos en los análisis meteorológicos pasados.

1.1.3.5.5.3 Causas de la heterogeneidad en series climáticas

Son muchas las razones que podrían alterar los datos obtenidos de una estación meteorológica, entre los más importantes podremos mencionar los siguientes:

- Desplazamiento del instrumento: sucede cuando las estaciones cambian de lugar, usualmente son desplazamientos pequeños, aunque este factor está ligado directamente al relieve del país, si los relieves son muy contrastados podrían generar una variación significativa en los registros de datos.
- Por modificación del medio circundante: surge cuando las construcciones o crecimiento de árboles son muy próximas a la estación.

- **Altura del instrumento:** se presenta cuando el aparato está expuesto a corrientes de aire, entre más altura y despejado se encuentre su ubicación, la presencia de este problema es mayor.
- **Defectos de calibración:** podría ocasionarse por error humano o por defectos de fabrica que tenga el aparato.
- **Error de transmisión:** problemas o defectos en la transmisión debidos a factores como; señal, componentes electrónicos, electricidad, etc.

1.1.3.5.4 Homogeneidad de datos pluviométricos

Es el proceso en el que se eliminan las alteraciones de la información obtenida de una estación meteorológica, por la variabilidad que puede ocasionarse por diversos acontecimientos como: error en la observación o transmisión, cambios o traslado del instrumento utilizado. Estos cambios son denominados como heterogeneidades y proyectan un falso comportamiento del clima, por lo que su corrección resulta obligatoria para estudios o investigaciones de variabilidad climática, donde los argumentos finales podrían resultar erróneos[17].

1.1.3.6 Inteligencia artificial (IA)

Se puede describir como la facultad que tiene una máquina para realizar todo proceso que requiere de inteligencia, basándose en la simulación del comportamiento humano a través de la experiencia. Las investigaciones actuales, parten desde una teoría matemática que sigue en desarrollo, tratando de englobar el comportamiento indicativo de la inteligencia. Las divisiones más importantes de la IA son; machine learning, métodos probabilísticos, computación evolutiva, teoría del caos y sistemas difusos. Se cree que las maquinas tienen el potencial necesario para sobrepasar a los humanos, pero actualmente se encuentran limitadas, ya que solo pueden desenvolver tareas rudimentarias[18].

Cuando se habla de inteligencia artificial no se refiere directamente a maquinas físicas (hardware) que realizan tareas del ser humano, dentro de la informática este término está ligado a la máquina virtual que puede resolver los problemas, dicho en otras

palabras, se refiere al sistema del procesamiento de la información que se escribe en un lenguaje de programación.

La IA tiene 5 categorías principales; la IA clásica, el modelo de conexión (basado en redes neuronales artificiales), la programación evolutiva, los autómatas celulares y los sistemas dinámicos. Si bien cada investigador se centra en una categoría, puede existir el caso en el que la IA es híbrida, es el caso en donde se compara la unión de la IA clásica y el modelo de conexión, para explicar el comportamiento del ser humano cuando realiza una tarea y se distrae con un tema que no tiene nada que ver con lo que estaba haciendo[19], [20].

¿Cómo se define la inteligencia artificial? La prueba de Turing es el medio más reconocido para poder definir si una máquina es tan inteligente como el hombre. Una de las interpretaciones de esta prueba, es que si la máquina es tan inteligente y puede engañar al ser humano, haciéndose pasar por uno, aprobaría dicho test[20].

1.1.3.6.1 Ramas de la inteligencia artificial

La inteligencia artificial está compuesta por distintos elementos, dentro de los cuales se pueden destacar tres grandes ramas[21]:

- Lógica difusa
- Redes neuronales artificiales
- Algoritmos genéticos

Debido al carácter de la investigación se centrará netamente en las redes neuronales artificiales.

1.1.3.6.2 Machine Learning (ML) o Aprendizaje Automatizado

Es la capacidad que tiene una máquina para aprender a través de una base de datos, adquirida por consecuencia de una interacción informática y estadística. Las ciencias computacionales se basan en la programación manual de los ordenadores, el ML busca la forma de que los ordenadores se programen a sí mismos, a base de algoritmos computacionales que puedan utilizarse para capturar, almacenar, indexar, recuperar y fusionar los datos de la manera más eficiente[22].

El ML se enfoca en tres tipos de aprendizaje; el supervisado, no supervisado y por reforzamiento[23].

- **Aprendizaje supervisado**

Parte de un conjunto definido de datos y de la conceptualización de como clasificarlos, esto quiere decir que, si se ingresan varias imágenes y se coloca el nombre, el software aprenderá a diferenciar y reconocer[23].

- **Aprendizaje no supervisado**

Se utiliza cuando se tienen grandes cantidades de datos y la clasificación no es importante, por lo que la comprensión del significado tendrá base en la formulación de algoritmos que permitan diferenciar sin poner etiquetas a los datos ingresados[23].

- **Aprendizaje por reforzamiento**

Es un modelo de aprendizaje conductual, donde el algoritmo necesita ser retroalimentado con la supervisión del usuario para que pueda obtener los mejores resultados, cada resultado dado deberá tener una ponderación de acuerdo al acercamiento de la respuesta[23].

1.1.3.6.3 Partes del Machine Learning

- **Data format (Base de datos)**

Conjunto de datos o conjunto finito de vectores reales con múltiples características. Desde un enfoque probabilístico, la información que se extrae de los vectores que pertenecen a una distribución estadística multivariante, se basan en condicionar al conjunto inicial, para lograr que todas las muestras sean independientes y distribuidas de forma idéntica. Así, los valores de salida pueden ser numéricos continuos o categóricos[24].

- **Learnability (Capacidad de aprendizaje)**

Es la capacidad de aprendizaje que tiene un modelo paramétrico, esta actividad se segmenta en dos partes: una estructura estática y un conjunto dinámico de parámetros.

El primero se consigue a través de un algoritmo que no cambia, excepto si se tienen condiciones que proporcionen funciones de remodelado, el segundo es el enfoque de optimización, en donde se ponen parámetros ilimitados que den origen a hipótesis infinitas, estas hipótesis dan como resultado el aprendizaje a través de la experiencia, de acuerdo a un conjunto de parámetros específicos[24].

1.1.3.6.4 Deep learning o Aprendizaje profundo

Es una subdivisión de la inteligencia artificial y que se encuentra a su vez, dentro de Machine Learning, se basa en el aprendizaje de máquinas, a través de grandes modelos de redes neuronales que tienen la capacidad de tomar decisiones precisas, en función de la información de los datos. Se utiliza cuando la información es compleja y existen grandes conjuntos de datos disponibles[25].

El Deep learning es un modelo de aprendizaje que posee técnicas y procedimientos algorítmicos explorados en el ML, con el fin de que las máquinas puedan aprender de la misma forma en que lo hace el ser humano. Estos algoritmos intentan simular el comportamiento biológico del aprendizaje que se da a través de las neuronas[26].

1.1.3.6.5 Algoritmos principales del Deep Learning

Se clasifican de acuerdo a las necesidades o finalidad del estudio y el tipo de aprendizaje:

- **Redes profundas de aprendizaje no supervisado**

Se centra en el análisis de patrones, agrupación de los datos estudiados, sin etiquetar ningún patrón u objetivo[27].

- **Redes profundas para aprendizaje supervisado**

Se clasifican los datos de forma directa, por medio de etiqueta, de manera que toma en cuenta que los patrones son conocidos[28].

- **Redes profundas híbridas**

Es una combinación de las anteriores mencionadas, esperando se tenga la posibilidad de discriminar[28].

Figura 17.Diagrama que muestra como las diferentes partes de un modelo de inteligencia artificial, se relacionan entre si dentro de disciplinas diferentes.

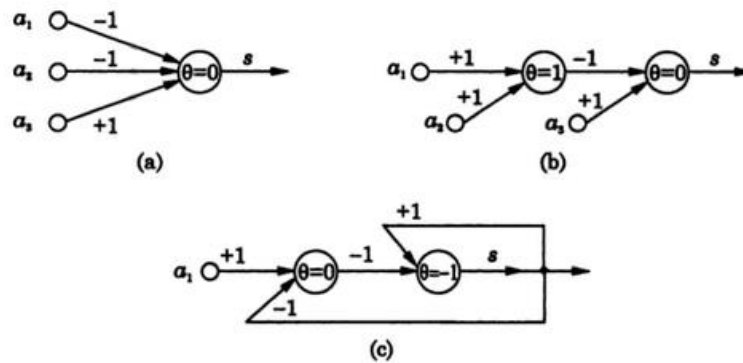
			Salida
		Salida	Mapeo de características
Salida	Mapeo de características	Mapeo de características	Características más complejas
Programa manual	Características manuales	Características	Características simples
Entrada	Entrada	Entrada	Entrada
<i>Sistema basado en reglas</i>	<i>Aprendizaje automático clásico</i>	<i>Aprendizaje de representación</i>	<i>Aprendizaje profundo</i>

Fuente: Y. Bengio, Deep learning[29].

1.1.3.6.6 Red neural artificial (RNA)

Se define como el modelo altamente simplificado de la estructura de una red neuronal biológica. Una RNA consta de unidades de procesamiento interconectadas. El modelo de la unidad de procesamiento consta de una parte de almacenamiento seguida de una parte de salida. La parte almacenamiento recibe N valores de entrada, revisa cada valor y realiza una condición. Esta condición establece si los valores de salida son satisfactorios[30].

Figura 18. Algunos ejemplos del flujo que tienen las redes neuronales artificiales.

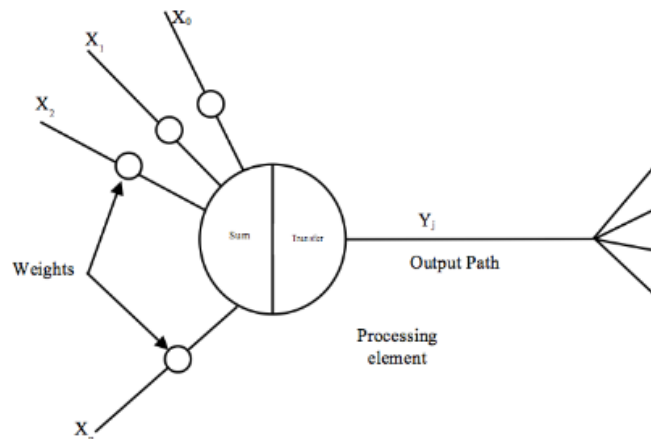


Fuente: B. Yegnanarayana, Artificial neural networks[30].

Las RNA presentan ventajas respecto a otros modelos de procesamiento de la información, ya que pueden sintetizar los algoritmos por medio de un proceso de aprendizaje. Uno de los fuertes de las RNA es que, si algún algoritmo falla seguirá trabajando, lo que les da una considerable ventaja respecto a los algoritmos convencionales. Entre sus desventajas se encuentra que deben entrenar para cada problema en lugar de programarlas, el proceso es largo y dependerá de la complejidad de los valores analizados[21].

La estructura de las neuronas artificiales está basada en las neuronas biológicas, en donde una neurona es equivalente a un elemento procesador, los cuales poseen varias entradas que pueden ser combinadas. Una vez combinadas cambian los valores de acuerdo a la función de transferencia y se envían a la salida del elemento procesador[31].

Figura 19.Diagrama de una neurona artificial.



Fuente: W. Rivas Asanza, Generalidades de las redes neuronales artificiales[31].

1.1.3.6.7 Ventajas de las redes neuronales artificiales

las Redes Neuronales Artificiales (RNAs) tienen muchas ventajas debido a que están basadas en la estructura del sistema nervioso, principalmente el cerebro. A continuación, se mencionan las más importantes[32]:

- **No linealidad.** El procesador neuronal es básicamente no lineal y, por consecuencia la red neuronal también.
- **Aprendizaje.** Las RNAs tienen la habilidad de aprender mediante una etapa que se llama etapa de aprendizaje. Esta consiste en proporcionar a la RNA datos como entrada, a su vez que se le indica cuál es la salida (respuesta) esperada.
- **Auto organización.** Una RNA crea su propia representación de la información en su interior, descargando al usuario de esto.
- **Adaptabilidad.** La red tiene la capacidad de adaptar sus parámetros, aún en tiempo real.
- **Tolerancia a fallos.** Debido a que una RNA almacena la información de forma redundante, ésta puede seguir respondiendo de manera aceptable aun si se daña parcialmente.
- **Uniformidad en el análisis y diseño.** Esto permitirá garantizar características precisas.

- **Tiempo real.** La estructura de una RNA es paralela, por lo cual, si esto es implementado con computadoras o en dispositivos electrónicos especiales, se pueden obtener respuestas en tiempo real.
- **Analogía con las redes biológicas.** Esto permite la utilización mutua del conocimiento de las dos áreas.

1.1.3.7 Lenguajes de programación usados para la inteligencia artificial

Tabla 2. Lenguajes de programación con soporte en inteligencia artificial

Lenguaje	Descripción
Python	Es un lenguaje de código abierto que se utiliza para el desarrollo de cualquier tipo de software, se enfoca en tener un código claro y legible.
R	Es un lenguaje de programación de inteligencia artificial que destaca por su capacidad para procesar big data.
Java	Es un lenguaje pensado para la programación de objetos mediante inteligencia artificial. Además, dispone de varias interfaces de datos pensadas para mejorar la experiencia del usuario.
C++	Es el lenguaje de programación más rápido del mundo, característica que lo hace destacar en la programación de inteligencia artificial.
Prolog	Sirve para diseñar proyectos médicos. Destaca su capacidad para automatizar el backtracking (sistema que permite buscar errores y retroceder en el proceso para solucionarlos).

Fuente: Autor.

1.1.3.8 Lenguaje de programación Phyton

Herramienta que se encuentra en constante desarrollo y posee varias características que la convierten en una herramienta multipropósito. Fue desarrollado con el fin de tener un código abierto con lenguaje de programación de alto nivel, cuya filosofía ha hecho que sea considerado como uno de los más fáciles de interpretar, con sintaxis clara y

concisa. Parte de su desarrollo se ve simplificado en el uso de módulos (son archivos con extensión que definen funciones, clases y variables, incluso tienen códigos ejecutables) que simplifican el proceso de programación[33].

1.1.3.8.1 Librerías de Python

Son complementos funcionales de programación elaborados por terceros, que se utilizan, como soporte en las líneas de programación para ahorrar recursos y optimizar el lenguaje elaborado. A continuación, se detallan las librerías necesarias para la realización de un pronóstico de datos mediante el uso de redes neuronales artificiales:

- Pandas

Pandas es un anexo de Python, sus principios se basan en la manipulación y tratamientos de composiciones de datos. Una de sus funciones es delimitar estructuras desde cero con datos basados en los arreglos de la librería NumPy [34].

- Admite leer y escribir rápidamente ficheros con extensión CSV y bases de datos SQL.
- Ofrece métodos para realizar operaciones matemáticas. Permite manipular series temporales.

- Numpy

Complemento especializado en realizar cálculos numéricos que involucran bases de datos de gran volumen, ancla objetos llamados arrays para coleccionar datos y que su manipulación sea eficiente, la ventaja de utilizar este módulo es que su rapidez a la hora de procesar la información es de 50 veces más eficiente en comparación de otras librerías, lo que convierte a esta función en una herramienta fundamental para vectores y matrices[34].

- Matplotlib

Se especializa en la creación de gráficos en dos dimensiones, entre los más comunes que podemos tener son: diagrama de barras, histograma, diagrama de sectores,

diagrama de caja, diagrama de violín, dispersión de puntos, líneas, áreas, diagramas de contorno y mapas de colores[35].

- Scikit-learn

Una de las librerías por excelencia para el desarrollo de aprendizaje supervisado en Python, la cantidad de algoritmos que se puede resolver es innumerable y puede fusionarse con otras librerías[36].

1.1.4 Hipótesis

1.1.4.1 Hipótesis alternativa

La utilización de redes neuronales artificiales por medio de módulos de inteligencia artificial en el programa Python, permite realizar una homogeneización de las series de datos pluviométricos de la cuenca del río Pastaza.

1.1.4.2 Hipótesis nula

La utilización de redes neuronales artificiales por medio de módulos de inteligencia artificial en el programa Python, no permite realizar una homogeneización de las series de datos pluviométricos de la cuenca del río Pastaza.

1.2 Objetivos

1.2.1 Objetivo General

- Homogeneizar los datos pluviales mensuales de la cuenca del río Pastaza mediante el uso de redes neuronales artificiales.

1.2.2 Objetivos Específicos

- Obtener los datos de registros mensuales de precipitación de la cuenca del Río Pastaza.
- Analizar y aplicar las redes neuronales artificiales para completar los datos de registros pluviales en estaciones de recolección de datos meteorológicos del INAMHI.
- Evaluar los datos obtenidos a través de la homogeneización por medio de redes neuronales artificiales respecto a métodos convencionales.

CAPÍTULO II

METODOLOGÍA

2.1 Materiales

La información necesaria para la elaboración de este trabajo de investigación estará en función de los datos meteorológicos publicados por el INAMHI en el año 2009, debido que se podrá obtener el informe completo de la matriz de datos y la ubicación de las estaciones meteorológicas en un mapa que detalla su ubicación con coordenadas geográficas.

- Hardware

Computador portátil con procesador core i7 de 5ta generación y 8 gb ram.

- software

Además, se hará uso de un editor de código llamado “ATOM”, en donde se utilizará el lenguaje de programación Python en su versión 3.9.13, versión que al momento se encuentra estable y tiene alta compatibilidad con las librerías a usarse.

2.2 Métodos

2.2.1 Nivel o tipo de investigación

El presente trabajo de investigación se desarrollará en base a los siguientes niveles de investigación:

- Investigación bibliográfica

Parte de la investigación se realizará a partir de libros, revistas, artículos científicos y/o páginas web que explican los procesos de precipitación y la importancia de su estudio dentro de la hidrología, para conocer la distribución del agua y como repercute en el ciclo hidrológico.

La búsqueda de información referente al uso de las redes neuronales artificiales es otra de las razones por las que se recurrirá a este tipo de investigación, puesto que los modelos de programación se encuentran detallados en libros y artículos, en donde se describe su funcionamiento y aplicación en el procesamiento y validación de datos estadísticos.

- **Investigación descriptiva**

Es descriptiva porque busca especificar el comportamiento de los datos de precipitación registrados por las estaciones meteorológicas con la finalidad de observar si los antecedentes históricos tienen heterogeneidad.

- **Investigación aplicada**

Es aplicada debido a que se necesitan conocimientos previos de varias áreas como; informática, programación, hidrología y estadística. Con el propósito de desarrollar un software que genere un pronóstico de datos pluviales mediante redes neuronales artificiales.

2.2.2 Población y muestra

- **Población**

Dentro del análisis de los datos pluviométricos se debería únicamente que la precipitación es una variable aleatoria y que la población será todos los datos de precipitación registrados en todas las estaciones pluviométricas disponibles en el anuario meteorológico del año 2009 publicado por el INHAMI.

- **Muestra**

Por el tipo de investigación aplicado para la realización de este trabajo, no se requiere una muestra, puesto que se trabajará con los datos de precipitación registrados en el año 2009 de todas las estaciones pluviométricas que se encuentran instaladas dentro de la cuenca del río Pastaza.

2.2.3 Operacionalización de las variables

2.2.3.1 Variable independiente

- Módulos de inteligencia artificial

Tabla 3. Operacionalización de la variable independiente.

Concepto	Categoría	Indicador	Ítems	Técnicas e instrumentos
Los módulos de inteligencia artificial son librerías que tienen códigos ya implementados para sintetizar el proceso de programación.	Lenguaje de programación.	<ul style="list-style-type: none"> • Código de programación • Método de cálculo 	¿Cómo aplicar el código de programación para el método de cálculo?	Investigación bibliográfica
	Facilitar el proceso de cálculo.	<ul style="list-style-type: none"> • Eficiencia de los resultados • Precisión de los resultados 	¿Cómo el software permite agilizar y optimizar el proceso de cálculo?	Investigación bibliográfica

Fuente: Autor.

2.2.3.2 Variable dependiente

- Homogeneización de series pluviométricas de la cuenca del río Pastaza

Tabla 4. Operacionalización de la variable Dependiente

Concepto	Dimensión	Indicador	Ítems	Técnicas e instrumentos
La homogeneización es el proceso en el cual se ajustan los valores de una matriz de datos, de tal forma que se garantice que provengan de la misma naturaleza.	Homogeneidad de datos.	<ul style="list-style-type: none"> • Garantía de la información. • Métodos estadísticos para rellenar matrices de datos incompletos. 	¿Cuál es el método de cálculo estadístico para homogeneizar matrices de datos?	Investigación bibliográfica
	Heterogeneidad de datos.	<ul style="list-style-type: none"> • Negligencia en la toma de datos. 	¿Cuáles son los problemas que causan la heterogeneidad de las series de datos?	Investigación bibliográfica

Fuente: Autor.

2.2.4 Plan de recolección de la información

A continuación, se detalla la metodología para la recolección de los datos, con el fin de abordar los objetivos estipulados para este proyecto:

- Investigación bibliográfica de homogenización de datos y redes neuronales

Parte de la investigación se realiza a partir de libros, revistas y/o páginas web que explican los procesos de precipitación y la importancia de su estudio dentro de la hidrología, para conocer la distribución del agua y como repercute en el ciclo hidrológico. Además, se indagó los métodos de medición usados en mayor medida para registrar los comportamientos pluviales, y los problemas que se pueden presentar durante la medición, causales de heterogeneidades en las series de datos meteorológicos. Ha sido necesario la búsqueda de información a través de artículos científicos, que explican la importancia de la homogeneización de las series de datos de registros meteorológicos, en donde se recalca que la veracidad de los valores está en función de este análisis.

A través de portales web, se ha revisado la información referente a los datos meteorológicos que ha proporcionado el INAMHI (Instituto Nacional de Meteorología e Hidrología) hasta el año 2013, en donde se ha visualizado la ausencia de información y presencia de variables de heterogeneidad que dan cabida a la inexistencia de un modelo cercano a la realidad que pueda ser usado en los procesos de investigación y desarrollo de proyectos técnicos.

La búsqueda de información referente al uso de las redes neuronales artificiales es otro de las razones por las que se recurre a fuentes bibliográficas, puesto que los modelos de programación se encuentran detallados en libros y artículos, en donde se describe su funcionamiento y aplicación en el procesamiento y validación de datos estadísticos. Además, se denota el porcentaje de factibilidad que se tiene de su inserción en los modelos de análisis de datos hidrológicos.

- Determinación de los datos en análisis (datos pluviométricos)

La recolección de los datos se obtuvo mediante una búsqueda bibliográfica, en donde se encontró información con las características necesarias para la evaluación de los datos, considerando la ausencia o sospecha de la garantía de los valores. Además, se tomó como partida el año 2009, debido a que se encontró información completa como la ubicación de las estaciones en un mapa georreferenciado, lo que permite adicionar a la distancia y ubicación como variable.

Para el desarrollo de la investigación se posee la información de 231 estaciones meteorológicas en una matriz de Excel, de las cuales 18 pertenecen a la cuenca del río Pastaza, el restante servirá para la modelación de los datos que serán destinados al entrenamiento de las redes neuronales artificiales.

A continuación, se adjunta la matriz de datos de las estaciones ubicadas en la cuenca del río Pastaza:

Tabla 5. Datos pluviométricos de la cuenca del río Pastaza.

CÓD.	ENE.	FEB.	MAR.	ABR.	MAY.	JUN.
M008	660.7	369.1	275.8	560.2	350.0	462.0
M390	69.7	86.4	69.7	93.6	46.1	90.3
M378	220.4	161.8	117.2	324.9	250.6	378.9
M258	56.7	43.0	27.1	69.4	33.7	94.8
M029	78.7	71.5	52.4	151.6	110.8	196.6
M380	58.0	50.9	26.3	65.5	54.6	79.6
M128	43.7	40.7	31.4	60.8	20.4	79.2
M377						90.5
M376	93.0	60.0	45.2	80.0	29.8	90.1
M126	25.2					74.3

MA1Y	99.7	39.0	53.1	71.8	39.7	109.4
M127	15.7				134.0	28.4
M369	57.6	41.8	71.7	96.8	10.7	44.0
M004	74.9	41.4	88.6	75.7	21.6	43.3
M375	83.7	27.4	25.5	49.6	58.5	51.1
M371				76.7	39.4	32.8
MA1V	164.0	72.1	83.0	63.4		43.4
M120	202.4	168.4	154.2	64.9	104.6	136.6
CÓD.	JUL.	AGO.	SEP.	OCT.	NOV.	DIC.
M008	331.3	385.2	283.7	483.5	241.0	331.6
M390	50.8				22.6	22.1
M378	387.4	284.4	162.3	137.5	120.6	125.7
M258	53.4	27.4	12.4	44.4	14.8	8.7
M029	198.8	116.5	65.7	64.5	29.3	52.9
M380	87.4	31.7	20.0	30.2	12.4	22.3
M128	37.6	15.3	5.4	40.8	14.4	4.6
M377	66.4	32.6	19.8	84.8	28.8	14.7
M376	69.1	19.6	14.5	37.3	18.6	29.5
M126	59.6	31.7	10.1	88.8	11.5	33.4
MA1Y	89.9	32.5	10.6	55.9	27.2	30.1
M127	1.8	15.7	28.9			

M369	8.9	0.2	0.9	21.7	26.1	42.1
M004	11.5	1.6	10.7	27.8	17.1	68.3
M375	18.2	3.0	11.3	54.5	57.0	7.9
M371	2.7	6.4	7.6	73.4	62.9	50.5
MA1V	0.0	2.1	15.4	49.5	54.2	13.3
M120	36.9	23.4	28.7	81.2	100.5	96.1

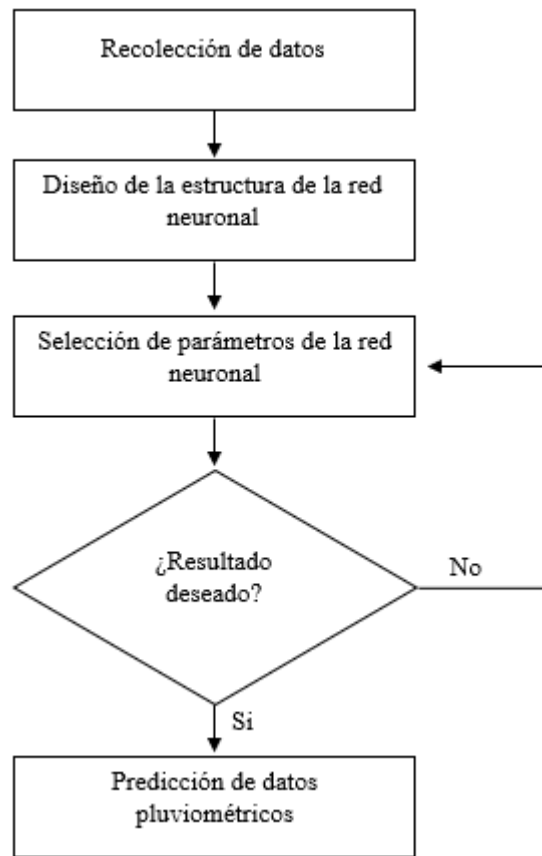
Fuente: Instituto Nacional de Meteorología e hidrología[37].

- Diseño del código del programa a analizar

Se plantea elaborar una red neuronal artificial basada en la estructura perceptrón multicapa, que permita predecir datos pluviales faltantes, haciendo uso de un algoritmo de aprendizaje conveniente. El resultado será obtenido a partir de ingresar los datos en un archivo con formato .CSV, para ser recolectados en el programa, a través de un módulo de inteligencia artificial que tenga la función de ordenar en formato de datos, para posteriormente ser clasificados y examinados por un módulo de regresión. A partir de los datos analizados, se entrenará redes neuronales artificiales que aprenderán a identificar los errores en los registros históricos pluviales, con el fin de lograr la estimación y corroboración de series de los datos faltantes o erróneos. La calidad de los valores resultantes será evaluada en el mismo programa con la función de comprobación que tiene el módulo principal que ordena en formato de datos. Esta función será la principal condicional para validar los resultados que serán impresos en un archivo de salida con formato .CSV, acompañado de una gráfica que permite visualizar la tendencia de las variables finales, ejecutada en el mismo programa.

A continuación, se muestra el esquema que representa la metodología que se seguirá para elaborar el código del programa:

Figura 20. Metodología para la elaboración del código de estimación de datos.



Fuente: Autor.

- Ingreso de los valores de precipitación en el código o en el modelo

En este punto se entrenará la red neuronal artificial con los datos de las estaciones meteorológicas de las cuencas del Ecuador exceptuando la cuenca del río Pastaza. Estos datos son importantes porque están destinados al diseño y construcción de la red neuronal.

- Calibración del programa.

La comparación de los datos de salida con datos reales permitirá identificar si los resultados son eficaces o no, para esto se tomarán estaciones meteorológicas que tengan datos completos, luego se quitarán datos aleatorios para procesar los vectores que contienen la información y de tal manera corroborar si los datos salientes son parecidos o tienen relación con los datos reales. Mediante este proceso se podrá

obtener un valor de eficacia que este en función de la distancia o separación de las estaciones meteorológicas.

- Ingreso de los datos pluviométricos de las estaciones correspondientes a la muestra

Una vez se culmine con el entrenamiento y calibración de la red neuronal artificial, se seleccionará las estaciones meteorológicas ubicadas en la cuenca del río Pastaza que tengan datos faltantes o datos erróneos que sean causales de heterogeneidad, seguido a esto se procederá a seleccionar las estaciones más cercanas, estas servirán de referencia para completar los datos ausentes. Estas estaciones meteorológicas serán colocadas en la matriz de Excel de datos de entrada del código de programación.

2.2.5 Plan de procesamiento y análisis de la información

A continuación, se muestran los pasos que se seguirán para cumplir con el objetivo de este apartado:

- Procesamiento de datos pluviométricos en el código.
- Datos pluviométricos homogenizados.
- Homogeneización de datos pluviométricos con método convencional.
- Análisis de resultados.

CAPÍTULO III RESULTADOS Y DISCUSIÓN

3.1 Recolección y tratamiento de datos

La matriz de datos pluviales es indispensable para el uso de redes neuronales artificiales, puesto que la información servirá para determinar la garantía de los resultados. Los datos obtenidos serán separados en dos grupos; el primer grupo servirá para el entrenamiento y construcción del algoritmo basado en muestras de valores reales, el segundo grupo estará destinado a validar los resultados de los objetivos planteados.

De los datos recolectados de las estaciones meteorológicas, se ha detallado de acuerdo a la cuenca de origen, código, nombre, cantidad de valores faltantes y ubicación.

Debido al tamaño de los datos se muestra únicamente la información levantada de la cuenca del río Pastaza y el resto serán ubicadas en los anexos:

Tabla 6. Estaciones disponibles en el año 2009 de la cuenca del Río Pastaza.

CUENCA DEL RIO PASTAZA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	LONGITUD	LATITUD
MB83	TOTORILLAS	3	-7873.6816	-203.1799
M395	CEBADAS	2	-7865.4252	-192.3329
M396	ALAO	1	-7849.4818	-189.9688
M134	SHIRY XII	SIN REGISTRO	-7876.4050	-186.5138
M407	LICTO	0	-7861.2362	-182.0821

M133	GUASLAN	1	-7867.4136	-173.5808
M041	SANGAY	0	-7795.3618	-170.8765
M393	SAN JUAN	0	-7879.8557	-164.0619
M408	GUANO	0	-7863.2549	-161.9477
M008	PUYO	0	-7869.6679	-149.6558
M390	URBINA	3	-7869.6679	-149.6558
M378	RIO VERDE	0	-7830.2972	-141.4220
M258	QUEROCHACA	0	-7859.4954	-141.2856
M029	BAÑOS	0	-7842.7975	-140.4053
M380	HUAMBALO	0	-7853.8765	-140.0113
M128	PEDRO FERMIN	0	-7862.7353	-136.4711
M377	TISALEO	5	-7867.9258	-135.7131
M376	PILAHUIN	0	-7874.4245	-131.3889
M126	PATATE	4	-7851.0731	-131.2517
MA1Y	CALAMACA	0	-7883.5590	-129.3043
M127	PILLARO	6	-7856.4350	-118.0982
M369	CUSUBAMBA	0	-7871.2371	-107.7123
M004	RUMIPAMBA	0	-7860.4287	-102.8794

MB84	PUJILI	SIN REGISTRO	-7871.9457	-96.7063
M375	SAQUISILI	0	-7867.7113	-84.7221
M371	PASTOCALLE	3	-7864.4612	-73.2838
MA1V	COTOPILALO	1	-7871.2922	-69.1507
M121	EL REFUGIO	SIN REGISTRO	-7858.8951	-66.7264
M120	COTOPAXI	0	-7858.3590	-62.7178

Fuente: Instituto Nacional de Meteorología e hidrología[37].

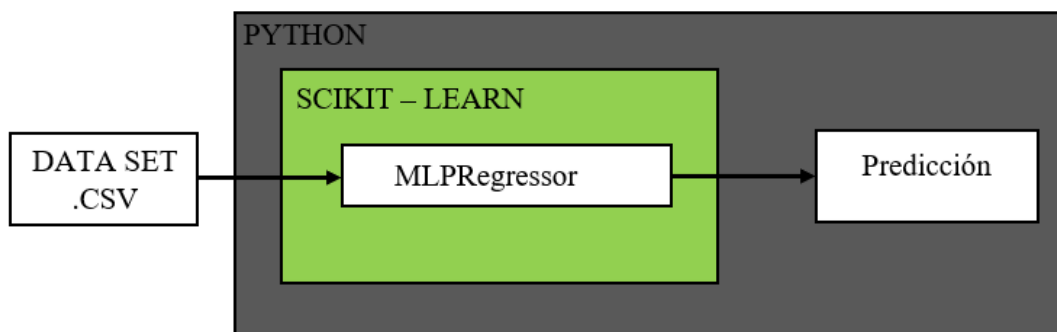
3.2 Diseño del software de estimación basado en redes neuronales artificiales

Para este proyecto se utilizó el lenguaje de programación PYTHON en su versión 3.9.13 a través del editor de código ATOM versión 1.60.0, para una mejor explicación se detalla a continuación el código y las tecnologías utilizadas:

3.2.1 Arquitectura del código de estimación

Se presenta la propuesta de construcción del programa, en donde se tiene la fuente de ingreso de los datos que en este caso será a través de una hoja de cálculo de Excel con extensión CSV, los valores serán importados en la base de datos del programa para ser tratados mediante la extensión MLPRegressor de la librería SCIKIT – LEARN para conseguir la estimación de los datos faltantes.

Figura 22.Diagrama de flujo del programa de estimación de datos.

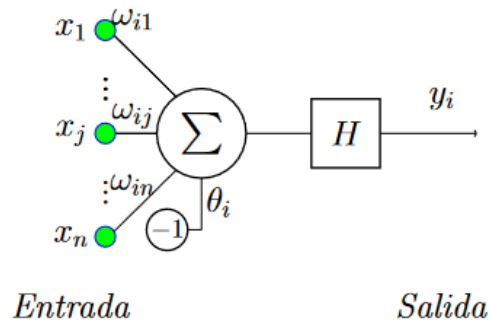


Fuente: Autor.

3.2.2 Arquitectura de la red neuronal artificial

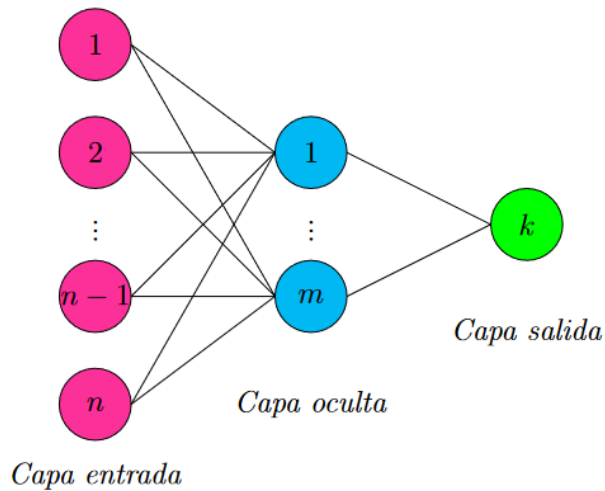
Por el tipo de estudio se ha optado por el modelo de red neuronal artificial perceptrón multicapa por ser una de las arquitecturas más poderosas y que tienen menos gasto computacional en la actualidad. Este tipo de redes neuronales artificiales son la combinación de varios perceptrones simples (forma más simple de una red neuronal artificial), están compuestos por una capa de entrada, tantas capas ocultas como sean necesarias y una capa de salida (o capa de respuesta), su entrenamiento se basa en el método backpropagation (propagación hacia atrás)[38].

Figura 23. Perceptrón simple.



Fuente: Vivas, H. Optimización en el entrenamiento del perceptrón multicapa[39].

Figura 24. Perceptrón multicapa.



Fuente: Vivas, H. Optimización en el entrenamiento del perceptrón multicapa[39].

3.3 Descripción de la codificación

La codificación de este proyecto se basa en el uso de librerías que están disponibles en Python, dentro de las cuales se utilizan las siguientes:

- PANDAS v.1.4.2
- NUMPY v.1.22.4
- MATPLOTLIB v.3.5.2
- SEABORN v.0.11.2
- SCIKIT – LEARN v.1.1.1

La librería SCIKIT LEARN permite importar el modelo de perceptrón multicapa directamente en el código para directamente a la fase entrenamiento con los datos de entrada.

3.2.2.1 Importación de librerías

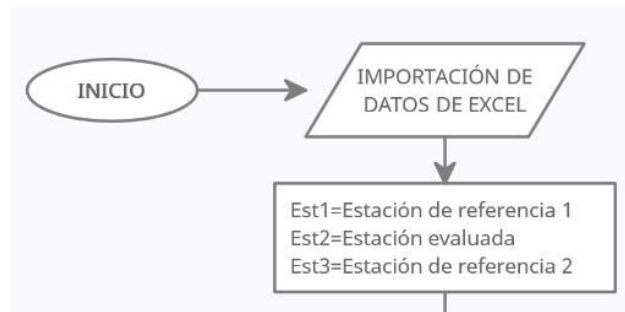
Para iniciar el diseño de la red neuronal artificial importaremos la librería PANDAS para poder procesar y manipular los datos de entrada que serán exportados a través de una matriz Excel, estos datos podrán ser leídos en el código a través de la librería NUMPY que tiene soporte para procesar vectores y matrices de gran dimensión. Para poder graficar los datos procesados en el código será necesario la presencia de la librería SEABORN, la cual es una extensión de la librería MATPLOTLIB que provee una interfaz para el procesamiento gráfico de datos estadísticos.

La librería SKLEARN.PREPROCESSING y SKLEARN.NEURAL_NETWORK son las encargadas de la implementación de la red neuronal artificial, la primera es el apartado que pone los parámetros para el entrenamiento y la segunda escoge el tipo de red neuronal artificial. Para el entrenamiento es necesario establecer un indicador para que el aprendizaje considere si los datos tienen relación y evitar que exista un mal comportamiento de los mismos[36].

3.2.2.2 Importación de datos de entrada

Para el ingreso de los datos se ha elaborado una matriz en Excel que considere dos estaciones de referencia (Est1 y Est3) y la estación con los datos faltantes (Est2). Estos datos serán importados a través de un comando funcional de la librería Pandas, el cual permite tomar información de un archivo con extensión .CSV y ordenarlos de forma matricial.

Figura 25. Algoritmo para la importación de datos de entrada.



Fuente: Autor.

3.2.2.3 Fase de entrenamiento

Primero se debe definir una matriz que elimine las posiciones donde no hay valores, para eso se utilizó la función “.dropna” en la definición de la variable “precipNotNan”, luego se definirá los datos tomados de las estaciones de referencia de la matriz que no contienen los valores nulos para tomar el nombre de “xTrain” y a las etiquetas de los resultados esperados como “yTrain” utilizando la función “.values.flatten()”, a continuación se llama al submódulo “StandardScaler()” para escalar los valores de xTrain en función de la media y la desviación estándar a través del comando “.scaler.transform()”, el resultado será la variable de nombre “xTrainScaled”.

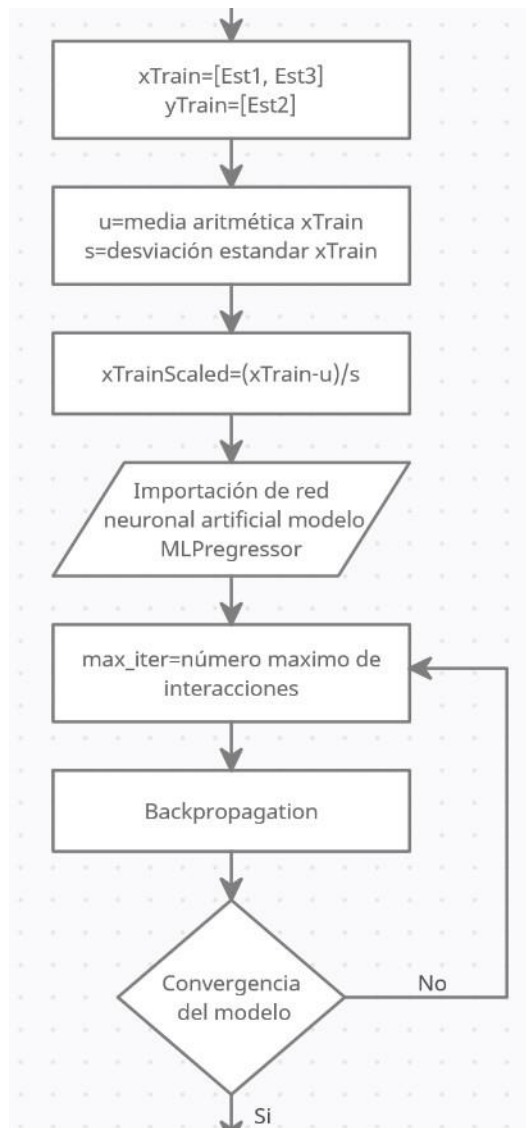
Se debe tomar en cuenta que el submódulo “MLPRegressor” basa su entrenamiento en el número de iteraciones que se necesitan para realizar la convergencia del modelo, razón por la cual cuando se llama al submódulo es necesario establecer ese parámetro a través del comando “max_iter”. Para comenzar el entrenamiento se debe definir una matriz “xTest” que contendrá los valores de las estaciones de referencia no nulos, los mismos que serán filtrados mediante el comando “.dropna”, estos valores serán escalados al igual que “xTrain” y definirán la variable “xTestScaled”. A través de la variable definida por el perceptrón multicapa se usará la regresión generada para elaborar un modelo de estimación en función de los datos de “xTest” a través del comando “.predict()”.

La red neuronal artificial del modelo perceptrón multicapa tiene un algoritmo interno que le permite entrenar a partir del método de propagación hacia atrás (backpropagation), el cual consiste en determinar el peso de las conexiones entre cada

capa de entrada y la capa oculta para la transferencia de la capa oculta hacia la capa de salida. Por cada valor de salida que se obtenga la red neuronal artificial considera un error arbitrario para compararlo con el real obteniendo el mínimo de error cuadrático posible [40].

Para la visualización de los datos resultantes del entrenamiento se hará uso de una gráfica de barras rotas, mismo que permite la visualización del error y el acercamiento del modelo generado por las redes neuronales a los datos de referencia de la estación incompleta, si los datos no tienen una aproximación cercana a los resultados deseados, se deberá cambiar el número de interacciones de la red neuronal artificial.

Figura 26. Algoritmo del entrenamiento y calibración de la RNA.

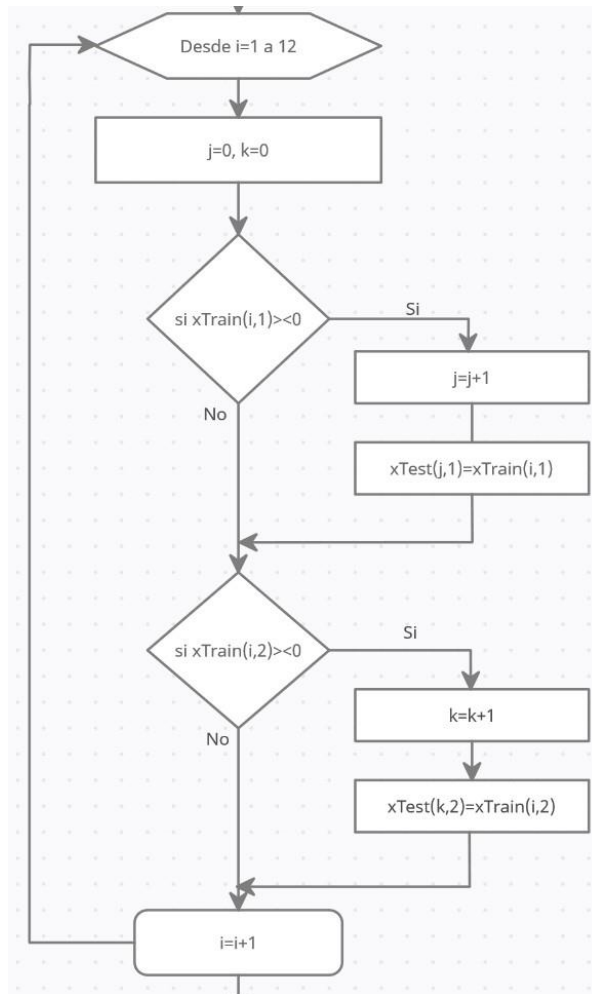


Fuente: Autor.

3.2.2.4 Obtención y exportación de datos faltantes de series pluviales

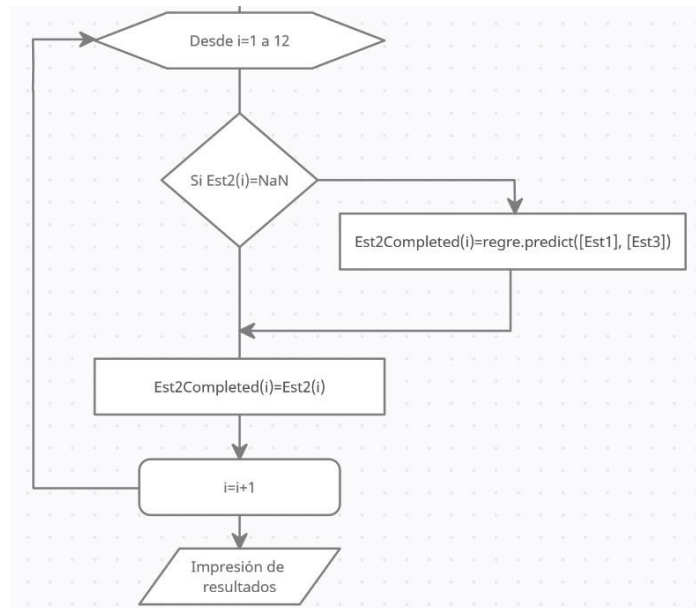
Una vez obtenido el modelo de estimación, se procede a filtrar los datos faltantes en la estación a procesar, para eso se hará uso de un ciclo que defina los valores en las posiciones de un vector denominado “Est2Completed”, se define una condición para que se filtren los valores nulos de la estación “Est2” de tal manera que imprima los valores existentes en el nuevo vector manteniendo su posición, en cambio si no existen valores escala al modelo de regresión generado por la red neuronal artificial y los imprime para completar los valores de la estación. Estos datos serán representados en una gráfica que muestre como se ajustan los valores faltantes de la estación en estudio a la de las estaciones de referencia. Además, se exportarán los datos a una matriz de Excel que contenga un nuevo vector con los datos completos.

Figura 27. Algoritmo para determinar el modelo predictivo.



Fuente: Autor.

Figura 28. Algoritmo para completar datos ausentes.



Fuente: Autor.

3.3 Evaluación y calibración del código

Como se explicó anteriormente, para poder entrenar la red neuronal artificial se hará uso de la información disponible de las estaciones meteorológicas de todas las cuencas del Ecuador con excepción a la cuenca del Río Pastaza que será destinada para la obtención y comparación de los resultados.

Las cuencas utilizadas se han filtrado de acuerdo a los requerimientos del código, entre ellos se ha descartado las cuencas que no poseen un mínimo de 3 estaciones meteorológicas, debido a que es necesaria la presencia de 2 estaciones de referencia y una de estudio, de la misma forma se han descartado las estaciones que tienen datos incompletos por la razón de que en la fase de entrenamiento es necesario conocer los valores resultantes para ajustar el modelo de estimación.

A continuación, se detalla las estaciones meteorológicas que se descartarán por no existir registro de ningún dato:

Tabla 7.Listado de estaciones que no se poseen datos pluviométrica.

CODIGO	NOMBRE DE LA ESTACIÓN
M179	ARENILLAS
M082	CHALPATAN
M154	CAYAPAS
M144	VILCABAMBA
M260	PIFO
M054	OBSERVARIO
M111	MALCHINGUI
M336	PACTO
MB85	QUININDE
MB76	JATÚN - SACHA
M468	MONTALVO
M383	ECHANDIA
M163	BOYACA
M446	SAN ISIDRO
M421	OÑA
MB08	CARTAGENA
M910	MORASPUGRO
M311	CAHUASQUI
M084	BOLIVAR
MB80	GUANDERAS

M102	EL ANGEL
M070	TENA
M563	LORETO
M293	PALMORIENTE
M134	SHIRY XII
MB84	PUJILI
M121	EL REFUGIO
MB90	EL CEBOLLAR
M906	JACARIN
MA70	COSMOAGRO
M173	PLAYAS
M175	MARESPI
MB06	SANTA ELENA

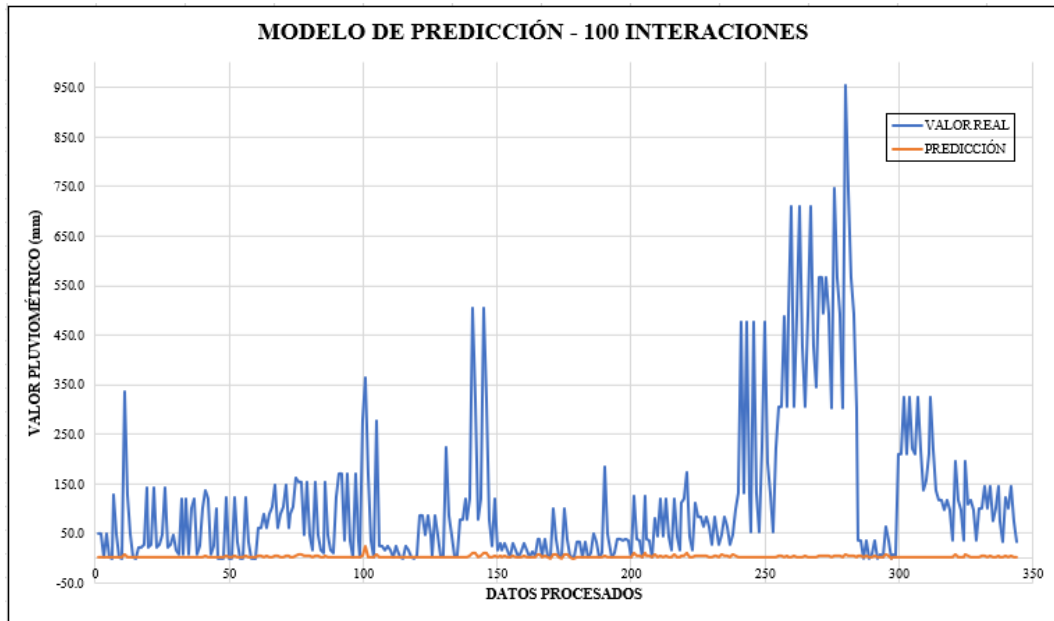
Fuente: Autor.

El número de iteraciones se verá afectado de acuerdo al valor de la salida obtenido y estará definido por la experiencia del programador. El entrenamiento se ha resumido en 4 fases, en donde se describe; el número de iteraciones de la red neuronal artificial, los valores resultantes en comparación a los valores reales, el porcentaje de error y la distancia de las estaciones de referencia en función de la estación de estudio:

- **Fase 1: ejecución del código con 100 iteraciones**

Dentro de la información obtenida de las estaciones meteorológicas de las cuencas del Ecuador, se ha observado que el número mayor de datos faltantes en una estación es 5, por lo tanto para estimar las interacciones que la red neuronal artificial necesita para obtener un modelo que emule el comportamiento de las estaciones de referencia se quitará de forma progresiva 5 valores para ver el comportamiento de los resultados.

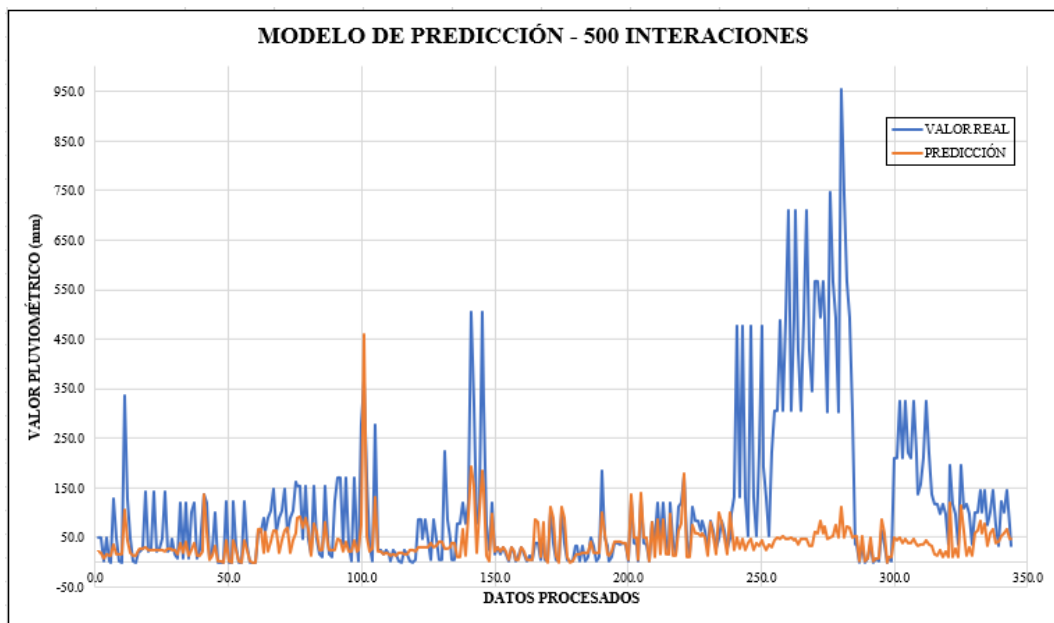
Figura 29.Convergencia con 100 iteraciones del modelo predictivo



Fuente: Autor.

- **Fase 2: ejecución del código con 500 interacciones**

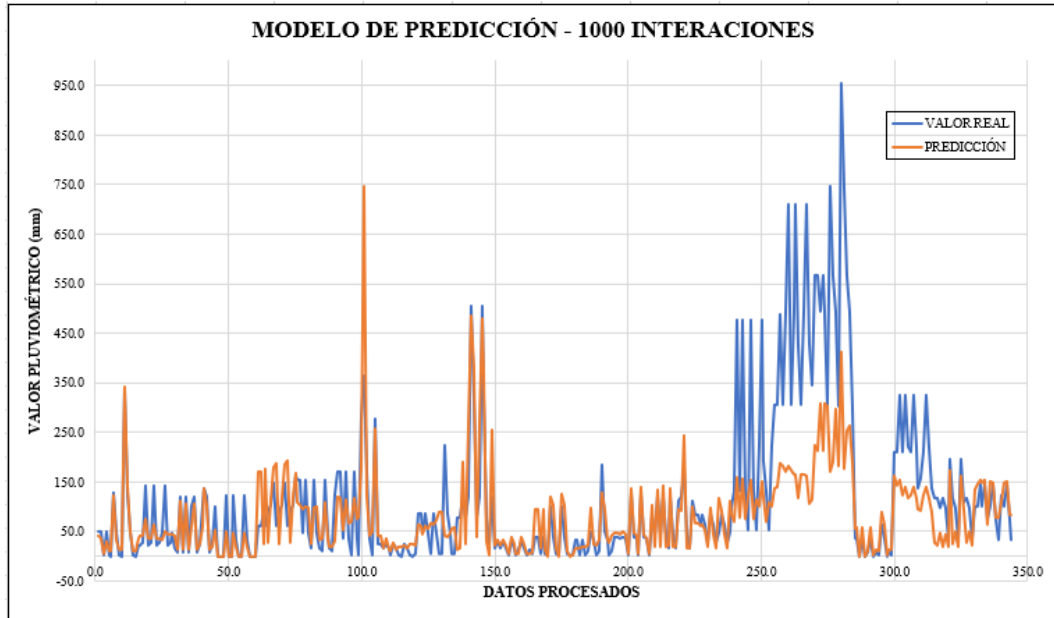
Figura 30.Convergencia con 500 iteraciones del modelo predictivo.



Fuente: Autor.

- Fase 3: ejecución del código con 1000 iteraciones

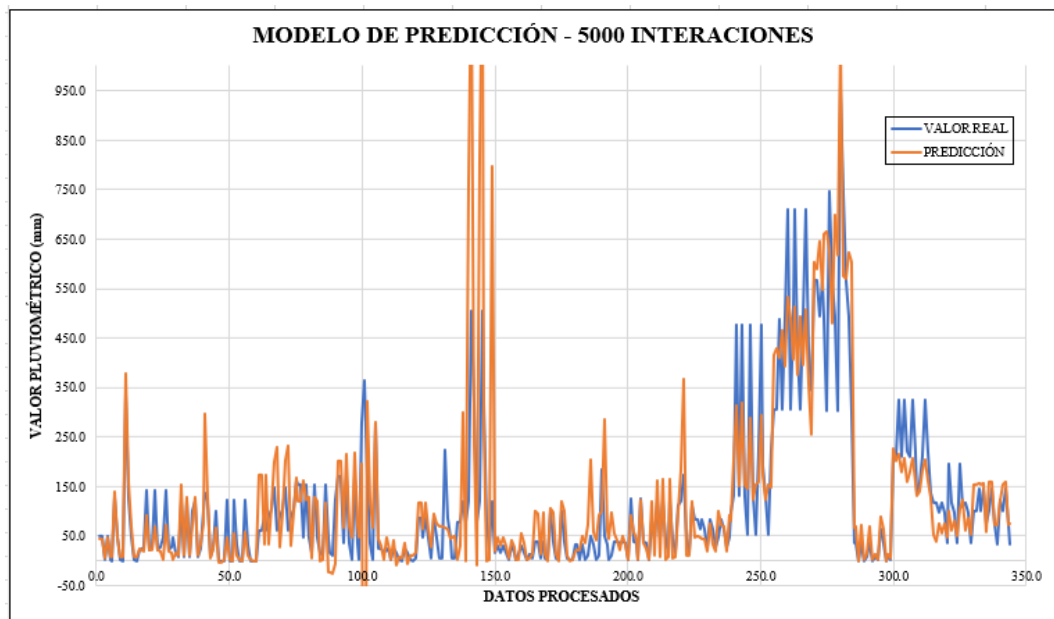
Figura 31. Convergencia con 1000 iteraciones del modelo predictivo.



Fuente: Autor.

- Fase 4: ejecución del código con 5000 iteraciones

Figura 32. Convergencia con 5000 iteraciones del modelo predictivo.



Fuente: Autor.

De los resultados obtenidos se observa que con 5000 iteraciones los datos tienen mayor aproximación a los valores reales, además en los gráficos se prueba que en la cuarta fase el modelo generado por la red neuronal artificial se ajusta al de los datos de referencia, por lo que se utilizará el valor de 5000 interacciones.

3.4 Homogeneización de la Cuenca del Río Pastaza

En la tabla 6 se detalla la información obtenida de las estaciones meteorológicas de la cuenca del Río Pastaza, en donde se observa la cantidad de valores faltantes en cada serie pluviométrica.

Tabla 8. Resumen de estaciones meteorológicas de la cuenca del río Pastaza.

CUENCA DEL PASTAZA				
CODIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	LATITUD	LONGITUD
MB83	TOTORILLAS	3	-78.736816	-2.031799
M395	CEBADAS	2	-78.654252	-1.923329
M396	ALAO	1	-78.494818	-1.899688
M407	LICTO	0	-78.612362	-1.820821
M133	GUASLAN	1	-78.674136	-1.735808
M041	SANGAY	0	-77.953618	-1.708765
M393	SAN JUAN	0	-78.798557	-1.640619
M408	GUANO	0	-78.632549	-1.619477
M008	PUYO	0	-78.696679	-1.496558
M390	URBINA	3	-78.696679	-1.496558
M378	RIO VERDE	0	-78.302972	-1.41422

M258	QUEROCHACA	0	-78.594954	-1.412856
M029	BAÑOS	0	-78.427975	-1.404053
M380	HUAMBALO	0	-78.538765	-1.400113
M128	PEDRO FERMIN	0	-78.627353	-1.364711
M377	TISALEO	5	-78.679258	-1.357131
M376	PILAHUIN	0	-78.744245	-1.313889
M126	PATATE	4	-78.510731	-1.312517
MA1Y	CALAMACA	0	-78.83559	-1.293043
M127	PILLARO	6	-78.56435	-1.180982
M369	CUSUBAMBA	0	-78.712371	-1.077123
M004	RUMIPAMBA	0	-78.604287	-1.028794
M375	SAQUISILI	0	-78.677113	-0.847221
M371	PASTOCALLE	3	-78.644612	-0.732838
MA1V	COTOPILALO	1	-78.712922	-0.691507
M120	COTOPAXI	0	-78.58359	-0.627178

Fuente: Instituto Nacional de Meteorología e hidrología[37].

3.4.1 Estimación de datos de la estación meteorológica Totorillas

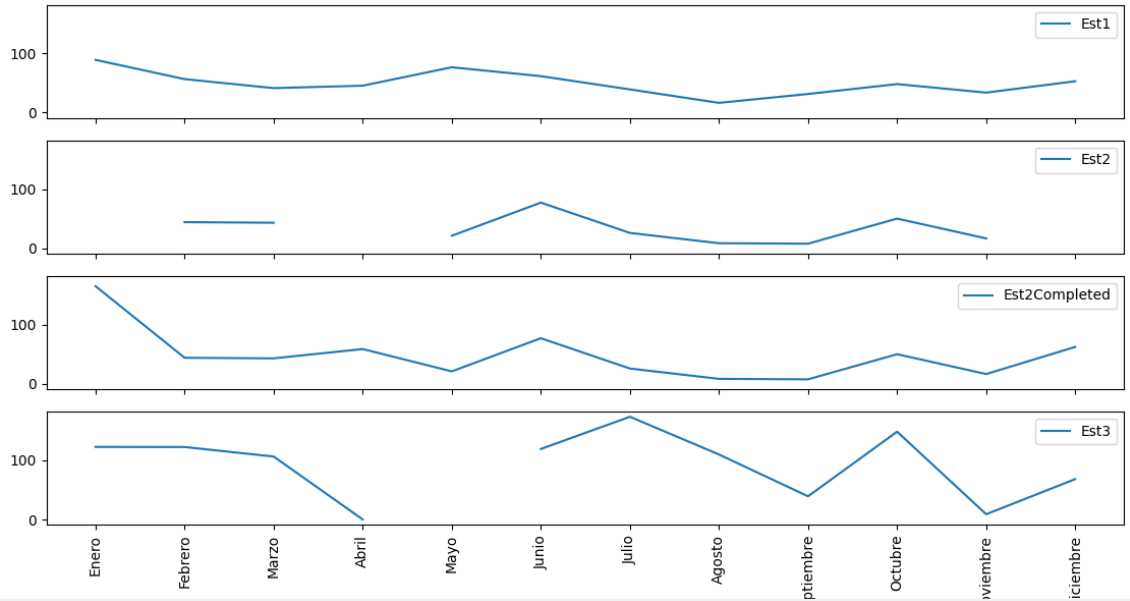
- Estación a evaluar: MB83 – Totorillas
- Estación de referencia 1: M407 – Licto
- Estación de referencia 2: M396 – Alao

Tabla 9. Resultados de la estación Totorillas

Est1	Est2	Est3	Est2Completa
89.1		122.6	164.97
56.6	44.1	122.4	44.1
41.4	43.1	106.3	43.1
45.5		0	58.81
76.6	21.1		21.1
61.6	77.1	119.1	77.1
39.2	25.9	173.5	25.9
16.5	8.5	109.8	8.5
31.5	7.7	39.2	7.7
48.2	50.1	148.3	50.1
33.8	16.6	8.9	16.6
53		68.2	62.41

Fuente: Autor.

Figura 33.Gráfica estimada de la estación Totorillas.



Fuente: Autor.

3.4.2 Estimación de datos de la estación meteorológica Cebadas

- Estación a evaluar: M395 – Cebadas
- Estación de referencia 1: M396 – Alao
- Estación de referencia 2: M407 – Licto

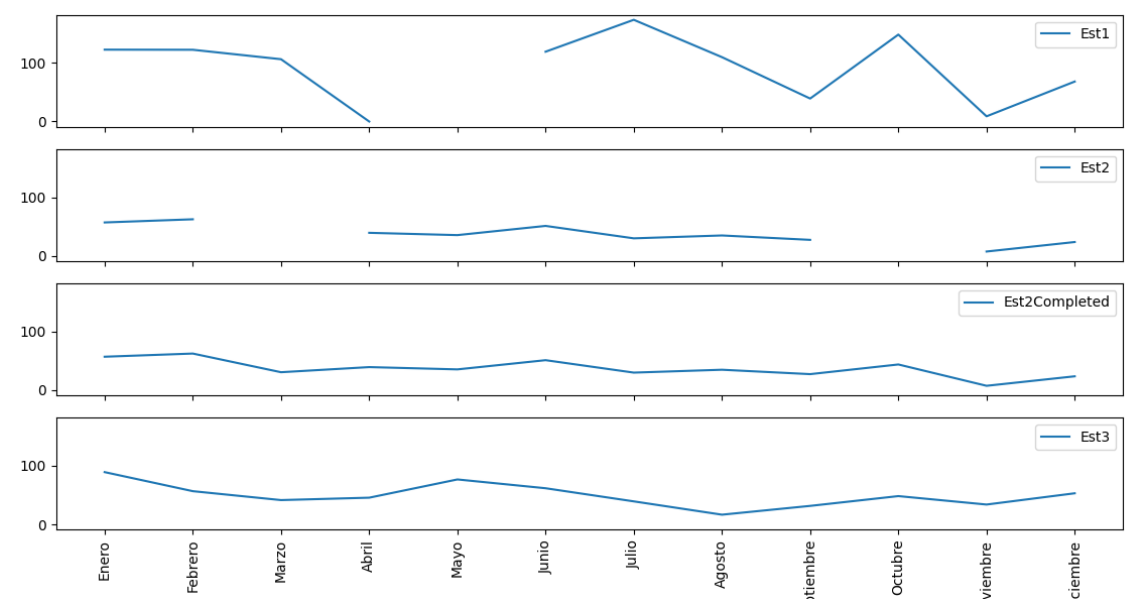
Tabla 10. Resultados de la estación Cebadas.

Est1	Est2	Est3	Est2Completa
122.6	56.9	89.1	56.9
122.4	62.3	56.6	62.3
106.3		41.4	30.48
0	39.2	45.5	39.2
	35.3	76.6	35.3
119.1	51	61.6	51
173.5	29.8	39.2	29.8
109.8	34.7	16.5	34.7
39.2	27.2	31.5	27.2

148.3		48.2	43.69
8.9	7.3	33.8	7.3
68.2	23.5	53	23.5

Fuente: Autor.

Figura 34. Gráfica estimada de la estación Cebadas.



Fuente: Autor.

3.4.3 Estimación de datos de la estación meteorológica Alao

- Estación a evaluar: M396 – Alao
- Estación de referencia 1: M395 – Cebadas
- Estación de referencia 2: M407 – Licto

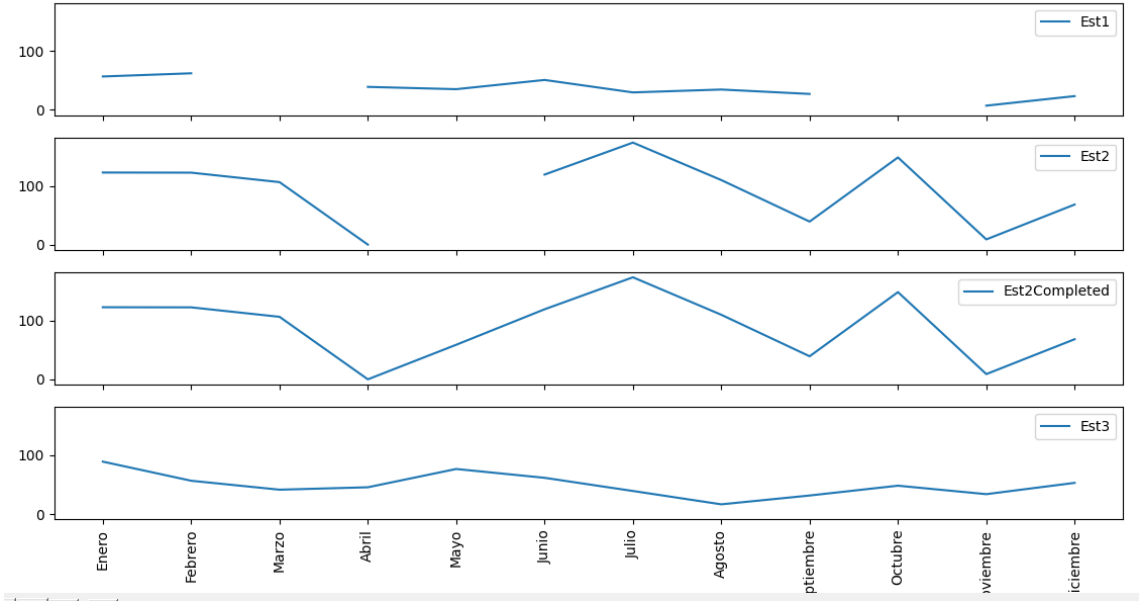
Tabla 11. Resultados de la estación Alao

Est1	Est2	Est3	Est2Completa
56.9	122.6	89.1	122.6
62.3	122.4	56.6	122.4

	106.3	41.4	106.3
39.2	0	45.5	0
35.3		76.6	58.77
51	119.1	61.6	119.1
29.8	173.5	39.2	173.5
34.7	109.8	16.5	109.8
27.2	39.2	31.5	39.2
	148.3	48.2	148.3
7.3	8.9	33.8	8.9
23.5	68.2	53	68.2

Fuente: Autor.

Figura 35. Gráfica estimada de la estación Alao.



Fuente: Autor.

3.4.4 Estimación de datos de la estación meteorológica Guaslán

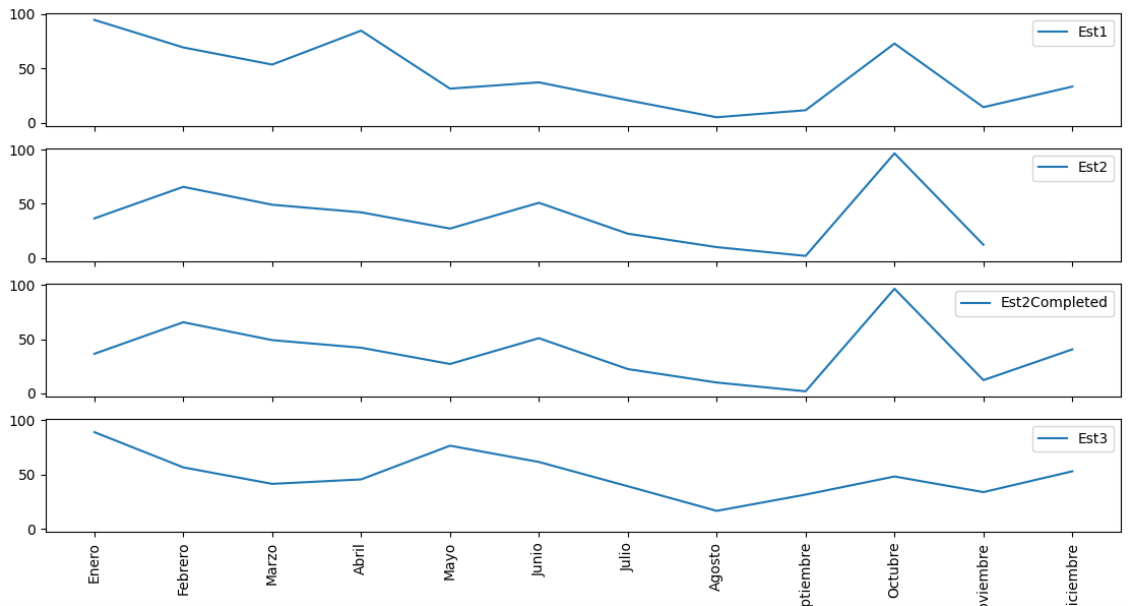
- Estación a evaluar: M133 – Guaslán
- Estación de referencia 1: M393 – San Juan
- Estación de referencia 2: M407 – Licto

Tabla 12. Resultados de la estación Guaslán.

Est1	Est2	Est3	Est2Completa
94.7	36.5	89.1	36.5
69.3	65.6	56.6	65.6
53.5	49.1	41.4	49.1
84.8	42.1	45.5	42.1
31.3	27.1	76.6	27.1
37.1	50.9	61.6	50.9
20.6	22.4	39.2	22.4
4.9	10	16.5	10
11.4	1.9	31.5	1.9
72.9	96.5	48.2	96.5
14.2	12.2	33.8	12.2
33.2		53	40.57

Fuente: Autor.

Figura 36. Gráfica estimada de la estación Guaslán.



Fuente: Autor.

3.4.5 Estimación de datos de la estación meteorológica Urbina

- Estación a evaluar: M390 – Urbina
- Estación de referencia 1: M258 – Querochaca
- Estación de referencia 2: M408 – Guano

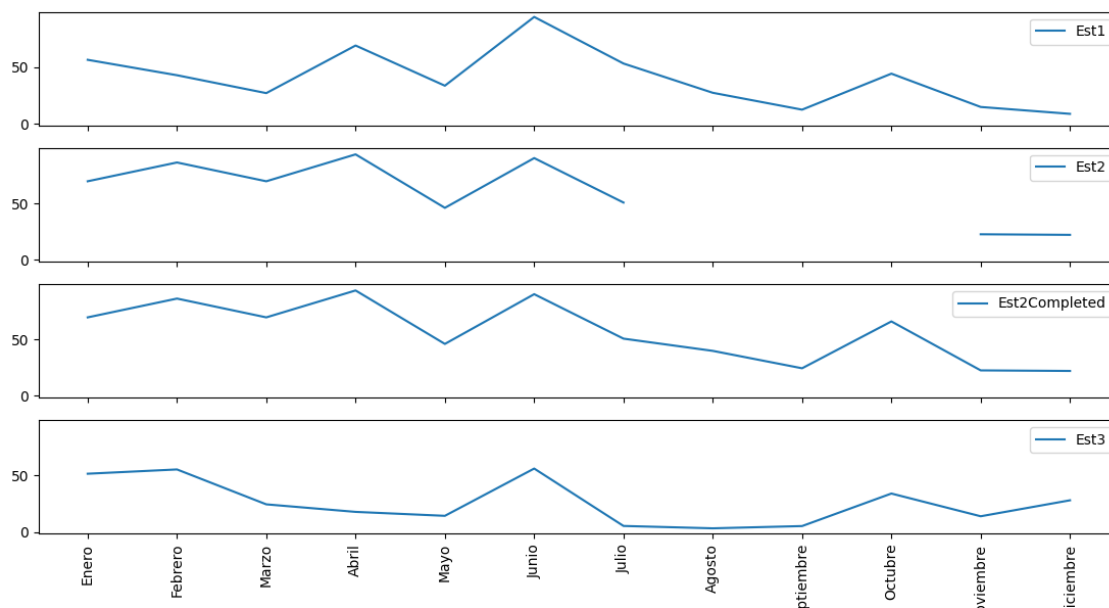
Tabla 13. Resultados de la estación Urbina

Est1	Est2	Est3	Est2Completa
56.7	69.7	51.7	69.7
43	86.4	55.5	86.4
27.1	69.7	24.5	69.7
69.4	93.6	17.8	93.6
33.7	46.1	14.3	46.1
94.8	90.3	56.3	90.3
53.4	50.8	5.3	50.8
27.4		3.2	39.94
12.4		5.2	24.47

44.4		34.1	66.09
14.8	22.6	13.9	22.6
8.7	22.1	28.1	22.1

Fuente: Autor.

Figura 37. Gráfica estimada de la estación Urbina



Fuente: Autor.

3.4.6 Estimación de datos de la estación meteorológica Tisaleo

- Estación a evaluar: M377 – Tisaleo
- Estación de referencia 1: M376 – Pilahuín
- Estación de referencia 2: M128 – Pedro Fermín

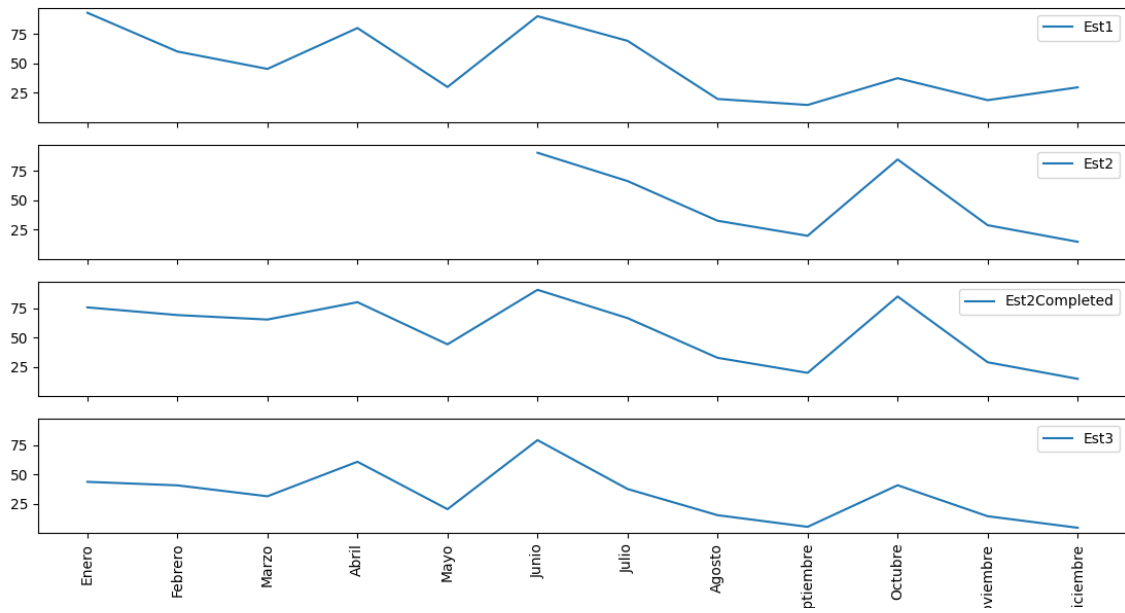
Tabla 14. Resultados de la estación Tisaleo.

Est1	Est2	Est3	Est2Completa
93		43.7	75.53
60		40.7	69.01

45.2		31.4	65.19
80		60.8	79.96
29.8		20.4	44.08
90.1	90.5	79.2	90.5
69.1	66.4	37.6	66.4
19.6	32.6	15.3	32.6
14.5	19.8	5.4	19.8
37.3	84.8	40.8	84.8
18.6	28.8	14.4	28.8
29.5	14.7	4.6	14.7

Fuente: Autor.

Figura 38. Gráfica estimada de la estación Tisaleo.



Fuente: Autor.

3.4.7 Estimación de datos de la estación meteorológica Pillaro

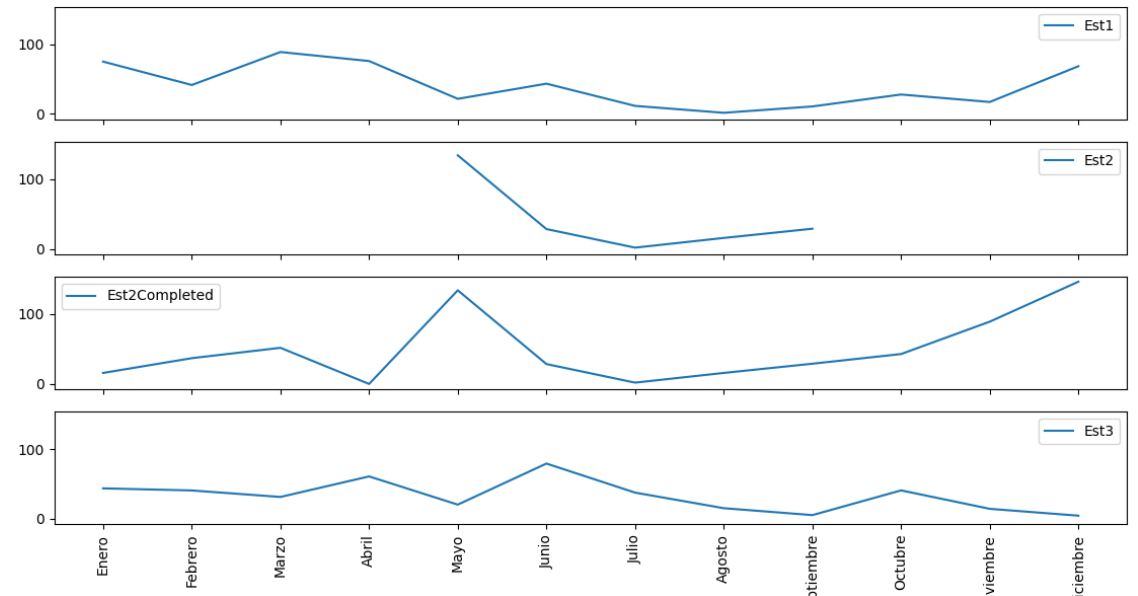
- Estación a evaluar: M127 – Píllaro
- Estación de referencia 1: M004 – Rumipamba
- Estación de referencia 2: M128 – Pedro Fermín

Tabla 15. Resultados de la estación Píllaro.

Est1	Est2	Est3	Est2Completa
74.9	15.7	43.7	15.7
41.4		40.7	36.82
88.6		31.4	51.66
75.7		60.8	0
21.6	134	20.4	134
43.3	28.4	79.2	28.4
11.5	1.8	37.6	1.8
1.6	15.7	15.3	15.7
10.7	28.9	5.4	28.9
27.8		40.8	42.74
17.1		14.4	89.11
68.3		4.6	146.30

Fuente: Autor.

Tabla 16. Gráfica estimada de la estación Píllaro.



Fuente: Autor.

3.4.8 Estimación de datos de la estación meteorológica Pastocalle

- Estación a evaluar: M371 – Pastocalle
- Estación de referencia 1: MA1V – Cotopilalo
- Estación de referencia 2: M375 – Saquisilí

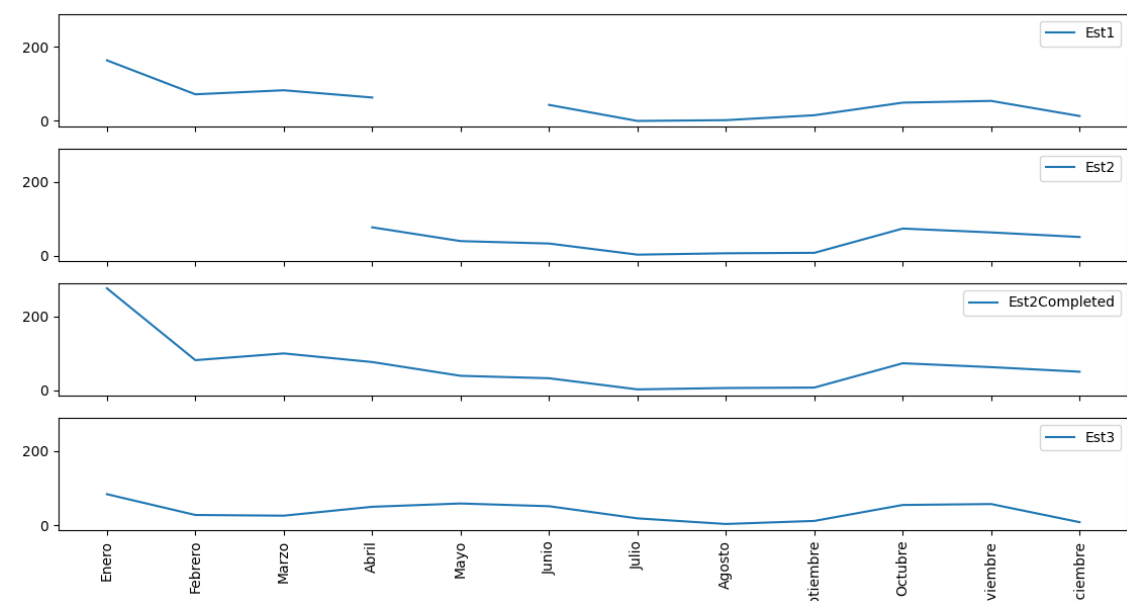
Tabla 17. Resultados de la estación Pastocalle.

Est1	Est2	Est3	Est2Completa
164		83.7	276.30
72.1		27.4	81.84
83		25.5	100.01
63.4	76.7	49.6	76.7
	39.4	58.5	39.4
43.4	32.8	51.1	32.8
0	2.7	18.2	2.7
2.1	6.4	3	6.4

15.4	7.6	11.3	7.6
49.5	73.4	54.5	73.4
54.2	62.9	57	62.9
13.3	50.5	7.9	50.5

Fuente: Autor.

Figura 39. Gráfica estimada de la estación Pastocalle.



Fuente: Autor.

3.4.9 Estimación de datos de la estación meteorológica Patate

- Estación a evaluar: M126 – Patate
- Estación de referencia 1: M128 – Pedro Fermín
- Estación de referencia 2: M029 – Baños

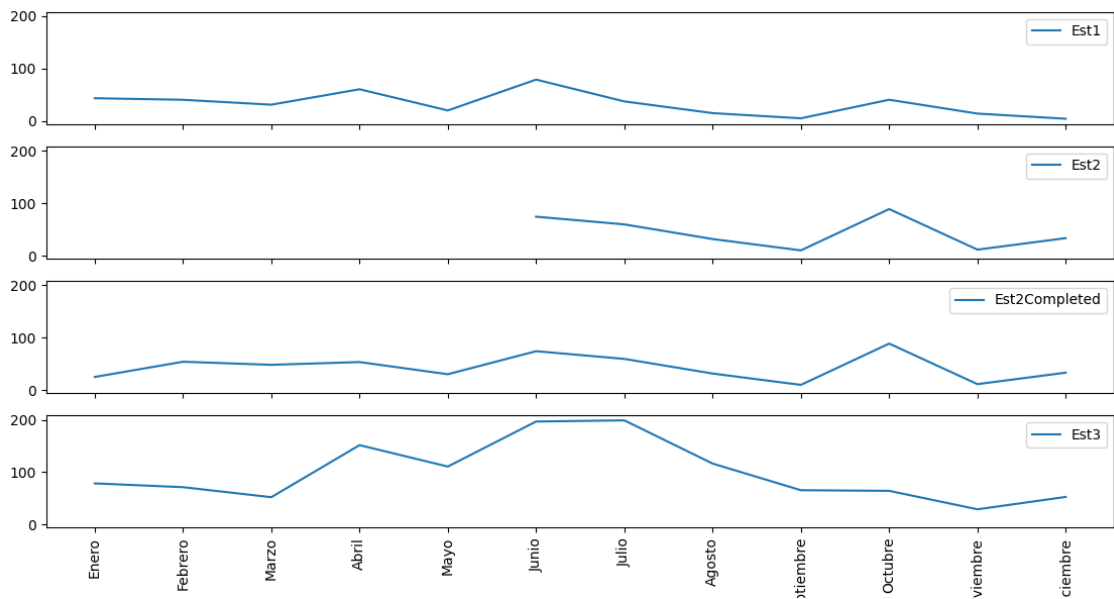
Tabla 18. Resultados de la estación Patate.

Est1	Est2	Est3	Est2Completa
43.7	25.2	78.7	25.2
40.7		71.5	54.21

31.4		52.4	48.35
60.8		151.6	53.58
20.4		110.8	30.39
79.2	74.3	196.6	74.3
37.6	59.6	198.8	59.6
15.3	31.7	116.5	31.7
5.4	10.1	65.7	10.1
40.8	88.8	64.5	88.8
14.4	11.5	29.3	11.5
4.6	33.4	52.9	33.4

Fuente: Autor.

Figura 40. Gráfica estimada de la estación Patate.



Fuente: Autor.

3.4.10 Estimación de datos de la estación meteorológica Cotopilalo

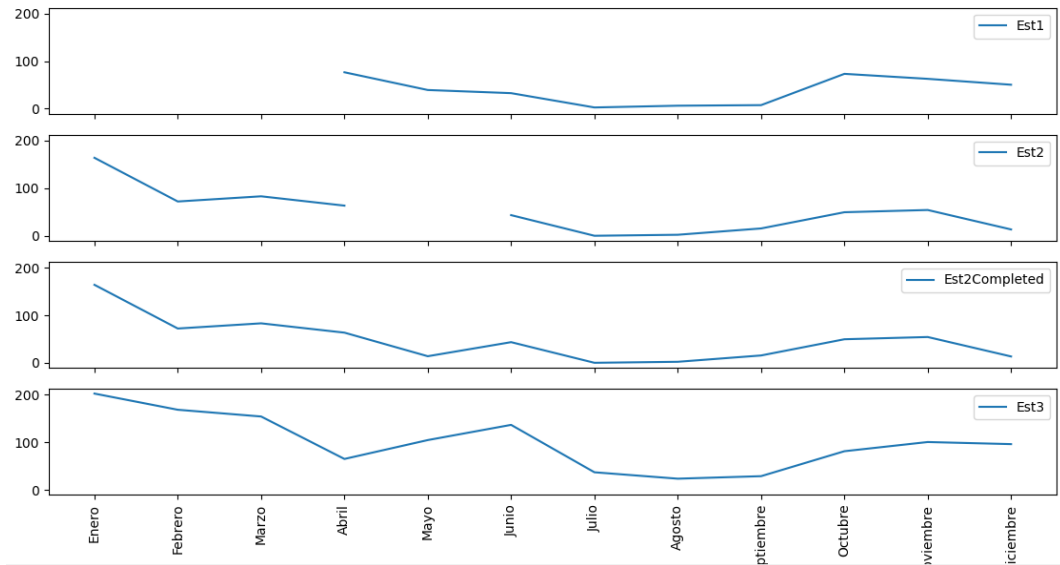
- Estación a evaluar: MA1V – Cotopilalo
- Estación de referencia 1: M371 – Pastocalle
- Estación de referencia 2: M120 – Cotopaxi

Tabla 19. Resultados obtenidos de la estación Cotopilalo.

Est1	Est2	Est3	Est2Completa
	164	202.4	164
	72.1	168.4	72.1
	83	154.2	83
76.7	63.4	64.9	63.4
39.4		104.6	13.80
32.8	43.4	136.6	43.4
2.7	0	36.9	0
6.4	2.1	23.4	2.1
7.6	15.4	28.7	15.4
73.4	49.5	81.2	49.5
62.9	54.2	100.5	54.2
50.5	13.3	96.1	13.3

Fuente: Autor.

Figura 41.Gráfica estimada de la estación Cotopilalo.



Fuente: Autor.

3.5 Homogeneización de la Cuenca del Río Pastaza por método manual

Para el desarrollo de este apartado se ha elegido el método de las razones de distancia, debido a que necesita de dos estaciones de referencia para estimar los valores faltantes. Además, en su cálculo incluye como variable la distancia que existe entre las estaciones de referencia y la estación a evaluar[41].

$$PX = PA + a * (PB - PA)/(a + b)$$

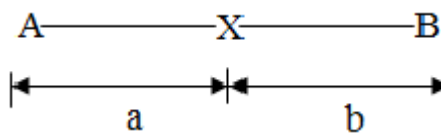
Donde,

PA Y PB, Estaciones de referencia con valores completos.

PX, Estación incompleta.

a y b, distancia de la estación incompleta a la estación de referencia.

Figura 42.Disposición espacial para la completación por razones de distancia.



Fuente: M. L. Yupa, estimación de datos faltantes de precipitación en la subcuenca del río Patate[41].

3.5.1 Estimación de datos de la estación meteorológica Totorillas

- Estación a evaluar: MB83 – Totorillas
- Estación de referencia 1: M407 – Licto
- Distancia entre estación de referencia 1 y estación incompleta: 27.14 km
- Estación de referencia 2: M396 – Alao
- Distancia entre estación de referencia 2 y estación incompleta: 30.65 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 89.1 + 27.14 * \frac{(122.6 - 89.1)}{(27.14 + 30.65)}$$

$$PX = 104.83$$

$$PX = 45.5 + 27.14 * \frac{(0 - 45.5)}{(27.14 + 30.65)}$$

$$PX = 24.13$$

$$PX = 53 + 27.14 * \frac{(68.2 - 53)}{(27.14 + 30.65)}$$

$$PX = 60.14$$

Tabla 20. Resultados con método manual de la estación Totorillas.

Est1	Est2	Est3	Est2Completa
89.1		122.6	104.83
56.6	44.1	122.4	44.1
41.4	43.1	106.3	43.1
45.5		0	24.13
76.6	21.1		21.1

61.6	77.1	119.1	77.1
39.2	25.9	173.5	25.9
16.5	8.5	109.8	8.5
31.5	7.7	39.2	7.7
48.2	50.1	148.3	50.1
33.8	16.6	8.9	16.6
53		68.2	60.14

Fuente: Autor.

3.5.2 Estimación de datos de la estación meteorológica Cebadas

- Estación a evaluar: M395 – Cebadas
- Estación de referencia 1: M396 – Alao
- Distancia entre estación de referencia 1 y estación incompleta: 17.94 km
- Estación de referencia 2: M407 – Licto
- Distancia entre estación de referencia 2 y estación incompleta: 12.26 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 106.3 + 17.94 * \frac{(41.4 - 106.3)}{(17.94 + 12.26)}$$

$$PX = 67.75$$

$$PX = 148.3 + 17.94 * \frac{(48.2 - 148.3)}{(17.94 + 12.26)}$$

$$PX = 88.84$$

Tabla 21. Resultados con método manual de la estación Cebadas.

Est1	Est2	Est3	Est2Completa
122.6	56.9	89.1	56.9
122.4	62.3	56.6	62.3
106.3		41.4	67.75
0	39.2	45.5	39.2
	35.3	76.6	35.3
119.1	51	61.6	51
173.5	29.8	39.2	29.8
109.8	34.7	16.5	34.7
39.2	27.2	31.5	27.2
148.3		48.2	88.84
8.9	7.3	33.8	7.3
68.2	23.5	53	23.5

Fuente: Autor.

3.5.3 Estimación de datos de la estación meteorológica Alao

- Estación a evaluar: M396 – Alao
- Estación de referencia 1: M395 – Cebadas
- Distancia entre estación de referencia 1 y estación incompleta: 17.94 km
- Estación de referencia 2: M407 – Licto
- Distancia entre estación de referencia 2 y estación incompleta: 15.73 km

$$\begin{aligned}
 \mathbf{PX} &= \mathbf{PA} + \mathbf{a} * \frac{(\mathbf{PB} - \mathbf{PA})}{(\mathbf{a} + \mathbf{b})} \\
 \mathbf{PX} &= 35.3 + 17.94 * \frac{(76.6 - 35.3)}{(17.94 + 15.73)} \\
 \mathbf{PX} &= 57.31
 \end{aligned}$$

Tabla 22. Resultados con método manual de la estación Alao.

Est1	Est2	Est3	Est2Completa
56.9	122.6	89.1	122.6
62.3	122.4	56.6	122.4
	106.3	41.4	106.3
39.2	0	45.5	0
35.3		76.6	57.31
51	119.1	61.6	119.1
29.8	173.5	39.2	173.5
34.7	109.8	16.5	109.8
27.2	39.2	31.5	39.2
	148.3	48.2	148.3
7.3	8.9	33.8	8.9
23.5	68.2	53	68.2

Fuente: Autor.

3.5.4 Estimación de datos de la estación meteorológica Guaslán

- Estación a evaluar: M133 – Guaslán
- Distancia entre estación de referencia 1 y estación incompleta: 17.40 km
- Estación de referencia 1: M393 – San Juan
- Distancia entre estación de referencia 2 y estación incompleta: 11.65 km
- Estación de referencia 2: M407 – Licto

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 33.2 + 17.40 * \frac{(53 - 33.2)}{(17.40 + 11.65)}$$

$$PX = 45.06$$

Tabla 23. Resultados con método manual de la estación Guaslán.

Est1	Est2	Est3	Est2Completa
94.7	36.5	89.1	36.5
69.3	65.6	56.6	65.6
53.5	49.1	41.4	49.1
84.8	42.1	45.5	42.1
31.3	27.1	76.6	27.1
37.1	50.9	61.6	50.9
20.6	22.4	39.2	22.4
4.9	10	16.5	10
11.4	1.9	31.5	1.9
72.9	96.5	48.2	96.5
14.2	12.2	33.8	12.2
33.2		53	45.06

Fuente: Autor.

3.5.5 Estimación de datos de la estación meteorológica Urbina

- Estación a evaluar: M390 – Urbina
- Estación de referencia 1: M258 – Querochaca
- Distancia entre estación de referencia 1 y estación incompleta: 14.63 km
- Estación de referencia 2: M408 – Guano
- Distancia entre estación de referencia 2 y estación incompleta: 15.36 km

$$\begin{aligned}
 \mathbf{PX} &= \mathbf{PA} + \mathbf{a} * \frac{(\mathbf{PB} - \mathbf{PA})}{(\mathbf{a} + \mathbf{b})} \\
 \mathbf{PX} &= 27.4 + 14.63 * \frac{(3.2 - 27.4)}{(14.63 + 15.36)} \\
 \mathbf{PX} &= 15.59
 \end{aligned}$$

$$PX = 12.4 + 14.63 * \frac{(5.2 - 12.4)}{(14.63 + 15.36)}$$

$$PX = 8.89$$

$$PX = 44.4 + 14.63 * \frac{(34.1 - 44.4)}{(14.63 + 15.36)}$$

$$PX = 39.38$$

Tabla 24. Resultados con método manual de la estación Urbina.

Est1	Est2	Est3	Est2Completa
56.7	69.7	51.7	69.7
43	86.4	55.5	86.4
27.1	69.7	24.5	69.7
69.4	93.6	17.8	93.6
33.7	46.1	14.3	46.1
94.8	90.3	56.3	90.3
53.4	50.8	5.3	50.8
27.4		3.2	15.94
12.4		5.2	8.89
44.4		34.1	39.38
14.8	22.6	13.9	22.6
8.7	22.1	28.1	22.1

Fuente: Autor.

3.5.6 Estimación de datos de la estación meteorológica Tisaleo

- Estación a evaluar: M377 – Tisaleo
- Estación de referencia 1: M376 – Pilahuín
- Distancia entre estación de referencia 1 y estación incompleta: 8.67 km
- Estación de referencia 2: M128 – Pedro Fermín
- Distancia entre estación de referencia 2 y estación incompleta: 5.84 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 93 + 8.67 * \frac{(43.7 - 93)}{(8.67 + 5.84)}$$

$$PX = 63.54$$

$$PX = 60 + 8.67 * \frac{(40.7 - 60)}{(8.67 + 5.84)}$$

$$PX = 48.47$$

$$PX = 45.2 + 8.67 * \frac{(31.4 - 45.2)}{(8.67 + 5.84)}$$

$$PX = 36.95$$

$$PX = 80 + 8.67 * \frac{(60.8 - 80)}{(8.67 + 5.84)}$$

$$PX = 68.53$$

$$PX = 29.8 + 8.67 * \frac{(20.4 - 29.8)}{(8.67 + 5.84)}$$

$$PX = 24.18$$

Tabla 25. Resultados con método manual de la estación Tisaleo.

Est1	Est2	Est3	Est2Completa
93		43.7	63.54
60		40.7	48.47
45.2		31.4	36.95
80		60.8	68.53
29.8		20.4	24.18
90.1	90.5	79.2	90.5

69.1	66.4	37.6	66.4
19.6	32.6	15.3	32.6
14.5	19.8	5.4	19.8
37.3	84.8	40.8	84.8
18.6	28.8	14.4	28.8
29.5	14.7	4.6	14.7

Fuente: Autor.

3.5.7 Estimación de datos de la estación meteorológica Píllaro

- Estación a evaluar: M127 – Píllaro
- Estación de referencia 1: M004 – Rumipamba
- Distancia entre estación de referencia 1 y estación incompleta: 17.41 km
- Estación de referencia 2: M128 – Pedro Fermín
- Distancia entre estación de referencia 2 y estación incompleta: 21.50 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 41.4 + 17.41 * \frac{(40.7 - 41.4)}{(17.41 + 21.50)}$$

$$PX = 41.09$$

$$PX = 88.60 + 17.41 * \frac{(31.40 - 88.60)}{(17.41 + 21.50)}$$

$$PX = 63.01$$

$$PX = 75.70 + 17.41 * \frac{(60.80 - 75.70)}{(17.41 + 21.50)}$$

$$PX = 69.03$$

$$PX = 27.80 + 17.41 * \frac{(40.80 - 27.80)}{(17.41 + 21.50)}$$

$$PX = 33.62$$

$$PX = 17.10 + 17.41 * \frac{(14.40 - 17.10)}{(17.41 + 21.50)}$$

$$PX = 15.89$$

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 68.30 + 17.41 * \frac{(4.60 - 68.30)}{(17.41 + 21.50)}$$

$$PX = 39.80$$

Tabla 26. Resultados con método manual de la estación Píllaro.

Est1	Est2	Est3	Est2Completa
74.9	15.7	43.7	15.7
41.4		40.7	41.09
88.6		31.4	63.01
75.7		60.8	69.03
21.6	134	20.4	134
43.3	28.4	79.2	28.4
11.5	1.8	37.6	1.8
1.6	15.7	15.3	15.7
10.7	28.9	5.4	28.9
27.8		40.8	33.62
17.1		14.4	15.89
68.3		4.6	39.80

Fuente: Autor.

3.5.8 Estimación de datos de la estación meteorológica Pastocalle

- Estación a evaluar: M371 – Pastocalle
- Estación de referencia 1: MA1V – Cotopilalo
- Distancia entre estación de referencia 1 y estación incompleta: 8.88 km
- Estación de referencia 2: M375 – Saquisilí
- Distancia entre estación de referencia 2 y estación incompleta: 13.16 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 164 + 8.88 * \frac{(83.70 - 164)}{(8.88 + 13.16)}$$

$$PX = 131.66$$

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 72.1 + 8.88 * \frac{(27.4 - 72.1)}{(8.88 + 13.16)}$$

$$PX = 54.10$$

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 83 + 8.88 * \frac{(25.5 - 83)}{(8.88 + 13.16)}$$

$$PX = 59.84$$

Tabla 27. Resultados con método manual de la estación Pastocalle.

Est1	Est2	Est3	Est2Completa
164		83.7	131.66
72.1		27.4	54.10
83		25.5	59.84

63.4	76.7	49.6	76.7
	39.4	58.5	39.4
43.4	32.8	51.1	32.8
0	2.7	18.2	2.7
2.1	6.4	3	6.4
15.4	7.6	11.3	7.6
49.5	73.4	54.5	73.4
54.2	62.9	57	62.9
13.3	50.5	7.9	50.5

Fuente: Autor.

3.5.9 Estimación de datos de la estación meteorológica Patate

- Estación a evaluar: M126 – Patate
- Estación de referencia 1: M128 – Pedro Fermín
- Distancia entre estación de referencia 1 y estación incompleta: 14.21 km
- Estación de referencia 2: M029 – Baños
- Distancia entre estación de referencia 2 y estación incompleta: 13.69 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 40.7 + 14.21 * \frac{(71.5 - 40.7)}{(14.21 + 13.69)}$$

$$PX = 56.39$$

$$PX = 31.4 + 14.21 * \frac{(52.4 - 31.4)}{(14.21 + 13.69)}$$

$$PX = 42.10$$

$$PX = 60.8 + 14.21 * \frac{(52.4 - 60.8)}{(14.21 + 13.69)}$$

$$PX = 107.04$$

$$PX = 20.4 + 14.21 * \frac{(110.8 - 20.4)}{(14.21 + 13.69)}$$

$$PX = 66.44$$

Tabla 28. Resultados con método manual de la estación Patate.

Est1	Est2	Est3	Est2Completa
43.7	25.2	78.7	25.2
40.7		71.5	56.39
31.4		52.4	42.10
60.8		151.6	107.04
20.4		110.8	66.44
79.2	74.3	196.6	74.3
37.6	59.6	198.8	59.6
15.3	31.7	116.5	31.7
5.4	10.1	65.7	10.1
40.8	88.8	64.5	88.8
14.4	11.5	29.3	11.5
4.6	33.4	52.9	33.4

Fuente: Autor.

3.5.10 Estimación de datos de la estación meteorológica Cotopilalo

- Estación a evaluar: MA1V – Cotopilalo
- Estación de referencia 1: M371 – Pastocalle
- Distancia entre estación de referencia 1 y estación incompleta: 14.21 km
- Estación de referencia 2: M120 – Cotopaxi
- Distancia entre estación de referencia 2 y estación incompleta: 13.69 km

$$PX = PA + a * \frac{(PB - PA)}{(a + b)}$$

$$PX = 39.4 + 14.21 * \frac{(104.6 - 39.4)}{(14.21 + 13.69)}$$

$$PX = 72.61$$

Tabla 29. Resultados con método manual de la estación Cotopilalo.

Est1	Est2	Est3	Est2Completa
	164	202.4	164
	72.1	168.4	72.1
	83	154.2	83
76.7	63.4	64.9	63.4
39.4		104.6	72.61
32.8	43.4	136.6	43.4
2.7	0	36.9	0
6.4	2.1	23.4	2.1
7.6	15.4	28.7	15.4
73.4	49.5	81.2	49.5
62.9	54.2	100.5	54.2
50.5	13.3	96.1	13.3

Fuente: Autor.

3.6 Análisis de resultados

En la tabla 30 se muestra la diferencia entre los resultados obtenidos del método manual y el método de redes neuronales artificiales.

Tabla 30. Resumen de resultados de homogeneización.

ESTACIÓN		MÉTODO		Diferencia
Código	Nombre	Manual	Software desarrollado	
MB83	Totorillas	104.83	164.97	60.14
		24.13	58.81	34.68
		60.14	62.41	2.27
M395	Cebadas	67.75	30.48	-37.27
		88.84	43.69	-45.15
M396	Alao	57.31	58.77	1.46
M133	Guaslán	45.06	40.57	-4.49
M390	Urbina	15.94	39.94	24
		8.89	24.47	15.58
		39.38	66.09	26.71
M377	Tisaleo	65.54	75.53	9.99
		48.47	69.01	20.54
		36.95	65.19	28.24
		68.53	79.96	11.43
		24.18	44.08	19.9
M127	Píllaro	41.09	36.82	4.27

		63.01	51.66	-11.35
		69.03	0	-69.03
		33.62	42.74	9.12
		15.89	89.11	73.22
		39.8	146.3	106.5
M371	Pastocalle	131.66	276.3	144.64
		54.1	81.84	27.74
		59.84	100.01	40.17
M126	Patate	56.39	54.21	-2.18
		42.1	48.35	6.25
		107.04	53.58	-53.46
		66.44	30.39	-36.05
MA1V	Cotopilalo	72.61	13.8	-58.81

Fuente: Autor.

Si bien los resultados de ambos métodos tienen diferencia, no significa un margen de error, puesto que la homogeneización es garantizar la naturaleza de los datos, no se podría valorar la información mediante simple observación y es necesario recurrir a métodos estadísticos no paramétricos.

La precipitación es una variable aleatoria debido a que su dominio es el espacio muestral y su rango es el conjunto de números reales, lo que hace que no se pueda examinar con métodos paramétricos puesto que la información no proviene de una población distribuida normalmente, razón por la cual es necesario usar pruebas que consideren su naturaleza.

3.6.1 Coeficiente de determinación (R^2)

Para la validación de los datos obtenidos se necesita calcular el coeficiente de determinación, puesto que mide la capacidad predictiva del modelo, dicho en otras palabras, se determina la aleatoriedad de los datos y debido a que la precipitación es una variable aleatoria, este parámetro nos permitirá observar la calidad de los resultados [42].

Para determinar el R^2 se hará uso de una hoja de cálculo de Excel, en donde se obtendrá una gráfica de dispersión a partir de los resultados de la estimación tanto del método manual como el método basado en redes neuronales artificiales, seguido a esto se dibujará la línea de tendencia que refleje el coeficiente de determinación.

En la tabla 31 se muestra los coeficientes de determinación encontrados para cada estación tanto del método manual como del método basado en redes neuronales artificiales.

Tabla 31. Coeficiente de determinación R^2 .

NOMBRE DE ESTACIÓN	R^2	
	SOFTWARE DESARROLLADO	MANUAL
Totorillas	53.90%	41.90%
Cebadas	50.20%	20.18%
Alao	5.34%	5.25%
Guaslán	7.52%	7.77%
Urbina	56.00%	60.32%
Tisaleo	38.95%	15.08%
Píllaro	29.75%	10.92%

Pastocalle	77.34%	64.20%
Patate	10.15%	22.50%
Cotopilalo	75.63%	73.99%

Fuente: Autor.

En función de los resultados obtenidos se puede observar que existe diferencia entre los valores de ambos métodos y mediante el coeficiente de determinación que mide la fiabilidad del ajuste de un modelo predictivo, se puede describir estadísticamente que el método basado en redes neuronales artificiales tiene mayor acercamiento en función de la ecuación generada para la línea de tendencia.

3.6.2 Prueba de rachas

Uno de los métodos no paramétricos que permite analizar el comportamiento de los resultados es la prueba de rachas, mismo que determina si los datos muestrales en una secuencia están en un orden aleatorio. Para validar la aleatoriedad de los datos se debe plantear una hipótesis que será aceptada en función de los valores críticos tabulados del método[43].

A continuación, se realizará la prueba de rachas en una muestra de la homogeneización obtenida en la estación Totorillas por el método basado en redes neuronales artificiales, el resto de pruebas se resumirán en una tabla que contenga los resultados de ambos métodos.

Tabla 32. Datos mensuales de la estación totorillas.

Mes	Datos
Enero	164.97
Febrero	44.1
Marzo	43.1
Abril	58.81
Mayo	21.1
Junio	77.1

Julio	25.9
Agosto	8.5
Septiembre	7.7
Octubre	50.1
Noviembre	16.6
Diciembre	62.41

Fuente: Autor.

- Cálculo de media aritmética (\bar{x})

$$\bar{x} = (164.97 + 44.10 + 43.10 + 58.81 + 21.10 + 77.10 + 25.90 + 8.50 + 7.70 + 50.10 + 16.60 + 62.41) / 12$$

$$\bar{x} = 48.37$$

- Determinación de N1 y N2

Tabla 33. Cálculo N1 y N2.

Datos	Datos $\geq \bar{x}$	# Racha
164.97	Si	1
44.1	No	2
43.1	No	
58.81	Si	3
21.1	No	4
77.1	Si	5
25.9	No	6
8.5	No	
7.7	No	
50.1	Si	7
16.6	No	8
62.41	Si	9

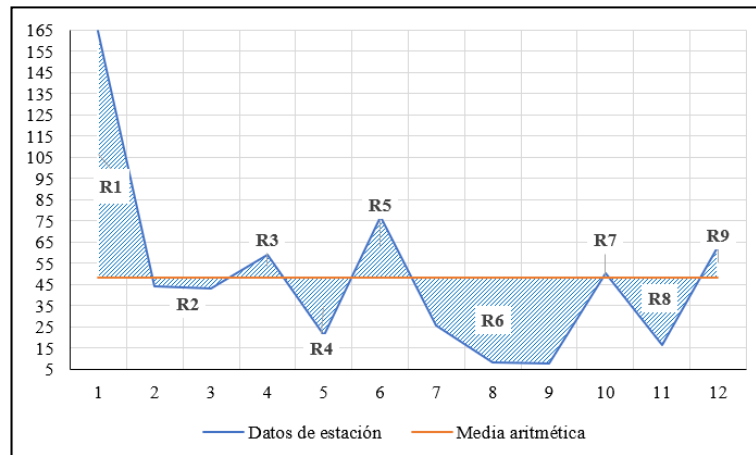
Fuente: Autor.

Tomando los datos de la tabla 31, se tiene que n1 será igual al número de valores que sean mayores a la media estadística y n2 será igual a la cantidad de valores que sean menores a la media estadística, por lo tanto:

$$n1=5 \text{ y } n2=7$$

El número de rachas estará definido por los grupos que se formen partiendo de la media aritmética, en el gráfico 42 se explica de mejor manera como se agrupan los valores para enumerar las rachas, en donde se obtiene que $G=9$.

Figura 43. Determinación del número de rachas.



Fuente: Autor.

Si $n1$ y $n2$ son menores a 20 y el nivel de confiabilidad que se requiere es del 95%, el estadístico de prueba es el número de rachas G . Los valores críticos se encuentran en la tabla 32.

Tabla 34. Valores críticos para el número de rachas.

		Valores críticos para el número de rachas G																		
		Valor de n_2																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Valor de n_1	2	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
	3	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3
	6	6	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	4	1	1	1	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4
	6	6	8	9	9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	5	1	1	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5
	6	6	8	9	10	10	11	11	11	12	12	12	12	12	12	12	12	12	12	12
	6	1	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5	6
	6	6	8	9	10	11	12	12	13	13	13	13	13	14	14	14	14	14	14	14
	7	1	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6	6	6	6
	6	6	8	10	11	12	13	13	14	14	14	14	14	15	15	15	16	16	16	16
	8	1	2	3	3	3	4	4	5	5	5	6	6	6	6	6	6	7	7	7
	6	6	8	10	11	12	13	14	14	15	15	16	16	16	16	16	17	17	17	17
	9	1	2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	7	8	8
	6	6	8	10	12	13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
	10	1	2	3	3	4	5	5	6	6	6	7	7	7	7	8	8	8	8	8
	6	6	8	10	12	13	14	15	16	16	17	17	18	18	18	19	19	19	19	20
	11	1	2	3	4	4	5	5	6	6	7	7	7	8	8	8	8	9	9	9
	6	6	8	10	12	13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	8	9	9	9	10	
6	6	8	10	12	13	14	16	16	17	18	19	19	20	20	21	21	21	22	22	
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10	
6	6	8	10	12	14	15	16	17	18	19	19	20	20	21	21	22	22	23	23	
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	11	11	11	
6	6	8	10	12	14	15	16	17	18	19	20	20	21	22	22	23	23	23	24	
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12	
6	6	8	10	12	14	15	16	18	18	19	20	21	22	22	23	23	24	24	25	
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12	
6	6	8	10	12	14	16	17	18	19	20	21	21	22	23	23	24	25	25	25	
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13	
6	6	8	10	12	14	16	17	18	19	20	21	22	23	23	24	25	25	26	26	
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13	
6	6	8	10	12	14	16	17	18	19	20	21	22	23	24	25	25	26	26	27	
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13	
6	6	8	10	12	14	16	17	18	20	21	22	23	23	24	25	26	26	27	27	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	14	14	
6	6	8	10	12	14	16	17	18	20	21	22	23	24	25	25	26	27	27	28	

NOTAS:

1. Los valores en esta tabla son los valores críticos G_c , suponiendo una prueba de dos colas con un nivel de significancia de $\alpha = 0.05$.
2. La hipótesis nula de aleatoriedad se rechaza si el número total de rachas G es menor que o igual al valor más bajo, o si es mayor que o igual al valor más alto.

De "Tables for Testing Randomness of Groupings in a Sequence of Alternatives", *The Annals of Mathematical Statistics*, vol. 14, núm. 1. Reproducido con permiso del Institute of Mathematical Statistics.

Fuente: Triola M. Estadística[44].

Los valores críticos obtenidos son 3 y 11. El criterio de decisión será el siguiente:

Se rechazará la aleatoriedad si el número de rachas G es menor o igual al valor crítico pequeño y si es mayor o igual al valor crítico más grande encontrado en la tabla. Entoces se tiene;

$$3 \leq G \leq 11$$

Puesto que $G=9$ no es menor o igual que el valor crítico 3, ni tampoco es mayor o igual que el valor crítico de 11, no se rechaza la aleatoriedad.

Tabla 35. Resumen de resultados de Prueba de Rachas.

NOMBRE DE ESTACIÓN	SOFTWARE DESARROLLADO	MANUAL
Totorillas	Aleatorio	Aleatorio
Cebadas	Aleatorio	Aleatorio
Alao	Aleatorio	Aleatorio
Guaslán	Aleatorio	Aleatorio
Urbina	Aleatorio	Aleatorio
Tisaleo	Aleatorio	Aleatorio
Píllaro	Aleatorio	Aleatorio
Pastocalle	Aleatorio	Aleatorio
Patate	Aleatorio	Aleatorio
Cotopilalo	Aleatorio	Aleatorio

Fuente: Autor.

En función de los resultados de la prueba de rachas con un índice de confiabilidad del 95%, se puede observar que ambos métodos cumplen con la aleatoriedad.

3.6.3 Correlación de rangos de Spearman

Si bien los datos obtenidos distan entre ambos métodos, no significa un margen de error, debido a que su análisis no puede ser paramétrico. Una forma correcta de analizar resultados sería mediante la obtención de la asociación de las variables o la correlación entre ambos métodos, por lo que la hipótesis nula y alternativa son las siguientes:

$H_0: p_s = 0$ (No existe correlación entre las dos variables)

$H_1: p_s \neq 0$ (Existe una correlación entre las dos variables)

Donde, p_s denota el coeficiente de correlación de rangos de la población completa [44].

A continuación, se determinará la correlación que tienen los valores calculados por el método basado en redes neuronales artificiales y el método convencional, se detallará el proceso de los valores de la primera estación analizada y luego se resumirá los resultados de las series restantes.

Se necesita hallar la diferencia al cuadrado de los rangos de las dos variables, para esto deberemos ubicar la posición que ocupan los valores de manera ascendente, en el caso de que existieran dos valores iguales su rango sería igual a la media aritmética de su posición. En la tabla 35 se presentan los rangos de las variables obtenidas a partir de la homogeneización por ambos métodos de la estación Totorillas.

Tabla 35. Cálculo de la diferencia al cuadrado de rangos de la estación Totorillas.

Método				d=RangoX-RangoY	d ²
Manual X	RangoX	Software Y	RangoY		
164.97	12	104.83	12	0	0
44.1	7	44.1	8	-1	1
43.1	6	43.1	7	-1	1
58.81	9	24.13	5	4	16
21.1	4	21.1	4	0	0
77.1	11	77.1	11	0	0
25.9	5	25.9	6	-1	1
8.5	2	8.5	2	0	0
7.7	1	7.7	1	0	0
50.1	8	50.1	9	-1	1
16.6	3	16.6	3	0	0
62.41	10	60.14	10	0	0
				∑ d²	20

Fuente: Autor.

El valor de n será igual a la cantidad de datos que contiene cada variable, en este caso:

$$n=12$$

Teniendo los datos necesarios, se determinará el coeficiente de correlación por rangos para datos muestrales apareados (r_s):

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 * 20}{12(12^2 - 1)}$$

$$r_s = 0.930$$

Puesto que $n < 30$, los valores críticos se tomarán en la tabla 36.

Tabla 36. Valores críticos del coeficiente de correlación de rangos de Spearman (r_s).

TABLA A-9 Valores críticos del coeficiente de correlación de rangos de Spearman r_s				
n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$
5	.900	—	—	—
6	.829	.886	.943	—
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.700	.783	.833
10	.564	.648	.745	.794
11	.536	.618	.709	.755
12	.503	.587	.678	.727
13	.484	.560	.648	.703
14	.464	.538	.626	.679
15	.446	.521	.604	.654
16	.429	.503	.582	.635
17	.414	.485	.566	.615
18	.401	.472	.550	.600
19	.391	.460	.535	.584
20	.380	.447	.520	.570
21	.370	.435	.508	.556
22	.361	.425	.496	.544
23	.353	.415	.486	.532
24	.344	.406	.476	.521
25	.337	.398	.466	.511
26	.331	.390	.457	.501
27	.324	.382	.448	.491
28	.317	.375	.440	.483
29	.312	.368	.433	.475
30	.306	.362	.425	.467

NOTAS:

1. Para $n > 30$, utilice $r_s = \pm z/\sqrt{n-1}$ donde z corresponde al nivel de significancia. Por ejemplo, si $\alpha = 0.05$, then $z = 1.96$.
2. Si el valor absoluto del estadístico de prueba r_s excede al valor crítico positivo, entonces rechace $H_0: \rho_s = 0$ y concluya que existe una correlación.

Basado en datos de "Biostatistical Analysis, 4th edition", © 1999, de Jerrold Zar, Prentice Hall, Inc., Upper Saddle River, Nueva Jersey, y "Distribution of Sums of Squares of Rank Differences to Small Numbers with Individuals", *The Annals of Mathematical Statistics*, vol. 9, núm. 2, con permiso del Institute of Mathematical Statistics.

Fuente: Triola M. Estadística [44].

Los valores críticos son ± 0.587 (con un nivel de significancia de 0.05 y $n=12$). Puesto que el estadístico de prueba $r_s=0.930$, el estadístico muestral excede al valor crítico por lo que se puede decir que existe una correlación significativa.

Tabla 37. Resultados de correlación de rangos de la estación Totorillas.

NOMBRE DE ESTACIÓN	r_s	NIVEL DE INSIGNIFICANCIA
Totorillas	0.930	0.05
Cebadas	0.790	0.05
Alao	1.000	0.05
Guaslán	0.993	0.05
Urbina	0.923	0.05
Tisaleo	0.958	0.05
Píllaro	0.434	0.20
Pastocalle	0.895	0.05
Patate	0.811	0.05
Cotopilalo	0.853	0.05

Fuente: Autor.

Los resultados de la correlación por rango demuestran que, de las 10 estaciones evaluadas 9 de ellas tiene una asociación alta, mientras que la restante tiene una correlación moderada. De tal manera se puede obtener una media aritmética de los niveles de significancia igual a 0.065.

3.7 Verificación de la hipótesis

En concordancia con la hipótesis alternativa anteriormente planteada en esta investigación y después de haber realizado el proceso de análisis de los resultados obtenidos por el método manual y con el software a base de módulos de inteligencia artificial centrados en redes neuronales artificiales, se obtiene valores aleatorios que respetan la naturaleza de las series pluviométricas y se determinó un 93.5% como mediaaritmética del índice de fiabilidad a partir de la evaluación de la asociación de valores entre el método manual y el método con redes neuronales artificiales, por lo que se da veracidad a la hipótesis planteada. Comprobando así, que la utilización de redes neuronales artificiales por medio de módulos de inteligencia artificial en el programa Phyton, permite realizar la homogeneización de las series pluviométricas de la cuenca del río Pastaza.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- El coeficiente de determinación obtenido en el método basado en redes neuronales artificiales es 8.27% mayor al método convencional, lo que establece que tiene mayor acercamiento hacia el modelo predictivo.
- El uso de 5000 iteraciones realiza la convergencia óptima entre el modelo de referencia y el modelo estimado creado por la red neuronal artificial, lo que permite tener valores de salida que se escalen de mejor manera y se acerquen a la realidad.
- El método basado en módulos de inteligencia artificial crea un modelo que estima datos aleatorios con un índice de fiabilidad del 95% y la correlación existente de los resultados de ambos métodos tiene un promedio del índice de asociación de variables del 87%.
- La homogeneización con el uso de redes neuronales artificiales mediante módulos de inteligencia artificial permite optimizar el tiempo en la estimación de datos pluviométricos faltantes de una estación meteorológica.
- La distancia de las estaciones de referencia respecto a la estación evaluada para la estimación de datos faltantes de una estación meteorológica mediante el uso de redes neuronales artificiales, tendrá influencia en la calidad de los valores, puesto que entre más cercanas se encuentren sus características climáticas serán similares.

4.2 Recomendaciones

- Para desarrollar una red neuronal artificial que estime los datos faltantes de una estación meteorológica es recomendable investigar las arquitecturas probadas para garantizar el cumplimiento de los propósitos planteados y optimizar tiempo mediante la evasión de las pruebas del funcionamiento.
- Para estimar los valores faltantes de una estación meteorológica, se recomienda tomar las estaciones de referencia más cercanas a la estación evaluada, si bien se demostró que la distancia puede ser una variable despreciable, se debe tomar en cuenta que el comportamiento de la precipitación es aleatorio y la distancia no define el comportamiento que tendrá en lugares más lejanos.
- Se recomienda Python para desarrollar software con base en redes neuronales artificiales, puesto que se posee librerías enfocadas en inteligencia artificial que optimizan la creación del código de programación.
- Es recomendable hacer evaluaciones de resultados con métodos convencionales en función de estaciones meteorológicas que contengan registros completos para poder tener un aproximado del acercamiento a valores reales.

Referencias bibliográficas

- [1] E. Najarchi, Mohsen; Mahavi, Ali; Hazaveh, “Determination of missing precipitation data by various methodologies,” vol. 7, no. 6, pp. 929–935, 2020.
- [2] R. Houari, A. Bounceur, A. K. Tari, and M. T. Kecha, “Handling missing data problems with sampling methods,” 2014. doi: 10.1109/INDS.2014.25.
- [3] J. Kajornrit, K. W. Wong, and C. C. Fung, “Estimation of Missing Precipitation Records Using Modular Artificial Neural Networks,” in *Neural Information Processing*, 2012, pp. 52–59.
- [4] W. E. Béjar Chacón, K. Yonatan ValerianoValdez, J. C. Ilachoque Umasi, and J. Sulla Torres, “Predicción de caudales medios diarios en la cuenca del Amazonas aplicando redes neuronales artificiales y el modelo neurodifuso ANFIS,” *Research in Computing Science*, vol. 113, no. 1, 2016, doi: 10.13053/rcs-113-1-2.
- [5] M. Yépez Heredia, “Los recursos naturales y el manejo de Cuencas Hidrográficas,” Pontificia Universidad Católica del Ecuador, Ecuador, 2015.
- [6] L. C. Tamayo-bacacela and A. L. Balarezo-aguilar, “Relleno de series anuales de datos meteorológicos mediante métodos estadísticos en la zona costera e interandina del Ecuador , y cálculo de la precipitación media,” pp. 81–90, 2016.
- [7] V. te Chow, D. R. Maidment, L. W. Mays, J. G. Saldarriaga, and Others, “Hidrología aplicada,” *Aguas superficiales*, vol. 146, 1994.
- [8] D. F. C. Aranda, *Procesos del ciclo hidrológico*. UASLP, 1984.
- [9] G. Monsalve Sáenz, “Hidrología en la Ingeniería,” *Alfaomega, Colombia*, 1999.
- [10] J. Faustino and F. Jiménez Otárola, “Manejo de cuencas hidrográficas,” 2000.
- [11] F. Gaspari, A. Rodríguez Vagaría, G. Senisterra, M. I. Delgado, and S. Besteiro, “Elementos metodológicos para el manejo de cuencas hidrográficas,” 2013.
- [12] R. F. P. López and E. R. Patrón, “Cuencas hidrográficas. Fundamentos y perspectivas para su manejo y gestión,” 2013.
- [13] M. V. Béjar, *Hidrología*. Instituto Tecnológico de Costa Rica, 2004.
- [14] A. F. B. Puyol and M. A. J. Villa, *Principios y fundamentos de hidrología superficial*. Universidad Autònoma Metropolitana, Unidad Iztapalpa, 2013.
- [15] W. Chereque Morán, “Hidrología: para estudiantes de ingeniería civil,” 1989.
- [16] F. J. Sánchez San Román, *Hidrología superficial y subterránea*. S.l: F. Javier Sánchez San Román, 2017.
- [17] J. A. Guijarro, “CLIMATOL: Software libre para la depuración y homogeneización de datos climatológicos,” 2004.

- [18] P. C. Jackson, *Introduction to artificial intelligence*. Courier Dover Publications, 2019.
- [19] M. A. Boden, *Inteligencia artificial*. Turner, 2017.
- [20] M. Alfonseca, “¿Basta la prueba de Turing para definir la “inteligencia artificial?” 2014.
- [21] P. P. Cruz, *Inteligencia artificial con aplicaciones a la ingeniería*. Alfaomega, 2011.
- [22] T. M. Mitchell, *The discipline of machine learning*, vol. 9. Carnegie Mellon University, School of Computer Science, Machine Learning~..., 2006.
- [23] J. Hurwitz and D. Kirsch, “Machine learning for dummies,” *IBM Limited Edition*, vol. 75, 2018.
- [24] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.
- [25] J. D. Kelleher, *Deep learning*. MIT press, 2019.
- [26] B. García Navarro and others, “Implementación de técnicas de deep learning,” 2015.
- [27] L. Deng and D. Yu, “Deep learning: methods and applications,” *Foundations and trends in signal processing*, vol. 7, no. 3--4, 2014.
- [28] E. Ordoñez Ramos, “Deep Learning para la visión artificial e identificación del personal administrativo y docente de la Universidad Nacional Micaela Bastidas de Apurímac 2018,” 2020.
- [29] Y. Bengio, I. Goodfellow, and A. Courville, *Deep learning*, vol. 1. MIT press Massachusetts, USA:, 2017.
- [30] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
- [31] W. Rivas Asanza, B. Mazón Olivo, and E. Mejía Peñafiel, “Generalidades de las redes neuronales artificiales.” Machala: Universidad Técnica de Machala, 2018.
- [32] D. J. Matich, “Redes Neuronales: Conceptos Básicos y Aplicaciones.” *Historia Santiago*, 2001.
- [33] J. Knowlton, *Python: Crear-modificar-reutilizar*, Anaya Multimedia., vol. 333. 2009.
- [34] W. McKinney, “pandas: a Foundational Python Library for Data Analysis and Statistics,” *Python for High Performance and Scientific Computing*, 2011.
- [35] G. R. Kanagachidambaresan and G. Manohar Vinoothna, “Visualizations,” in *EAI/Springer Innovations in Communication and Computing*, 2021. doi: 10.1007/978-3-030-57077-4_3.

- [36] F. Pedregosa and G. Varoquaux, “Scikit-learn: Machine Learning in Python ,” *JMLR*, vol. 12, pp. 2825–2830, 2011.
- [37] INAMHI, “Red de estaciones automáticas,” *Condiciones actuales del tiempo para las últimas 24 horas*, 2020.
- [38] F. Salazar Escobar and M. LLumitasig Galarza, “Simulación de pronósticos de ventas en la empresa IMPACTEX mediante redes neuronales,” Universidad Técnica de Ambato, Ecuador, 2021.
- [39] H. Vivas, “Optimización en el entrenamiento del perceptrón multicapa,” Universidad del Cauca, Popayán, 2014.
- [40] G. Ovando, M. Bocco, and S. Sayago, “REDES NEURONALES PARA MODELAR PREDICCIÓN DE HELADAS,” *Agricultura Técnica*, vol. 65, no. 1, 2005, doi: 10.4067/s0365-28072005000100007.
- [41] M. L. Yupa, K. Fonseca Largo, A. Quichimbo Miguítama, R. Lara Landázuri, and J. Tiche Toasa, “ESTIMACIÓN DE DATOS FALTANTES DE PRECIPITACIÓN EN LA SUBCUENCA DEL RÍO PATATE,” *Revista Bases de la Ciencia. e-ISSN 2588-0764*, vol. 2, no. 3, 2017, doi: 10.33936/rev_bas_de_la_ciencia.v2i3.1079.
- [42] R. Dios Palomares, “El modelo lineal sin termino independiente y el coeficiente de determinacion. Un estudio Monte carlo,” *QUESTIO*, vol. 22, no. 1, 1998.
- [43] J. L. Devore, *Probabilidad y estadística para ingeniería*. 2004.
- [44] F. M. Triola, “Probabilidad y estadística - Mario F. Triola - Google Libros,” in *Probabilidad y estadística*, vol. 1, 2004.

Anexos

Información levantada de las cuencas del Ecuador disponibles en el año 2009

Tabla 36. Estaciones disponibles en el año 2009 de la cuenca del Río Arenillas.

CUENCA DEL RÍO ARENILLAS				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M179	ARENILLAS	SIN REGISTRO	-80.085494	-3.566703
M482	CHACRAS	0	-80.231674	-3.567951

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 37. Estaciones disponibles en el año 2009 de la cuenca del Río Balao.

CUENCA DEL RÍO BALAO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M179	ARENILLAS	SIN REGISTRO	-80.085494	-3.566703
M482	CHACRAS	0	-80.231674	-3.567951

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 38. Estaciones disponibles en el año 2009 de la cuenca del Río Cañar.

CUENCA DEL RIO CAÑAR				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M031	CAÑAR	0	-78.955650	-2.571764
M411	INGAPIRCA	0	-78.891746	-2.559728
M477	PUERTO INCA	1	-79.569331	-2.549798
M412	SUSCALPAMBA	2	-79.082846	-2.480522

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 39. Estaciones disponibles en el año 2009 de la cuenca del Río Carchi.

CUENCA DEL RIO CARCHI				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M082	CHALPATAN	SIN REGISTRO	-77.813607	0.739267
M308	TUFIÑO	0	-77.855640	0.806664

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 40. Estaciones disponibles en el año 2009 de la cuenca del Río Cayapas.

CUENCA DEL RIO CAYAPAS				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y

M444	TEAONE	0	-78.695305	0.799419
M154	CAYAPAS	SIN REGISTRO	-78.978502	0.859019

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 41. Estaciones disponibles en el año 2009 de la cuenca del Río Catamayo.

CUENCA DEL RIO CATAMAYO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M438	JIMBURA	0	-79.492639	-4.660619
M150	AMALUZA	0	-79.457801	-4.617472
M433	EL LUCERO	0	-79.499251	-4.430636
M151	ZAPOTILLO	0	-80.271016	-4.411913
M147	YANGANA	0	-79.198750	-4.400173
M439	SABIANGO	0	-79.840948	-4.391949
M146	CARIAMANGA	1	-79.582581	-4.363550
M434	SOZORANGA	1	-79.819680	-4.354876
M544	CALAISACA	0	-79.722168	-4.346953
M145	QUINARA	0	-79.268063	-4.345046
M437	SAUCILLO	1	-80.232199	-4.310300
M144	VILCABAMBA	SIN REGISTRO	-79.241335	-4.294186

M149	GONZANAMA	0	-79.457878	-4.260416
M143	MALACATOS	0	-79.295766	-4.246719
MB88	NAMBOCOLA	0	-79.457643	-4.166945
MB87	POZUL	2	-80.089328	-4.143777
M148	CELICA	0	-79.983161	-4.133297
M515	CATACOCCHA	0	-79.672230	-4.085316
M435	ALAMOR	0	-80.060244	-4.046938

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 42. Estaciones disponibles en el año 2009 de la cuenca del Río Chone.

CUENCA DEL RIO CHONE				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M452	ZAPOTE	0	-80.085450	-0.948399
M462	JUNIN	0	-80.240441	-0.944116
MA34	RIO GRANDE	4	-79.974730	-0.739500
M162	CHONE	1	-80.139718	-0.713013
MB91	PUCE - BAHÍA	0	-80.429980	-0.663025
MA33	RIO SANTO	2	-80.030698	-0.576343

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 43. Estaciones disponibles en el año 2009 de la cuenca del Río Esmeraldas.

CUENCA DEL RIO ESMERALDAS				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M363	SIGCHOS	2	-78.906097	-0.708430
M364	LORETO	1	-78.435113	-0.569605
M353	RUMIPAMBA	0	-78.424067	-0.434088
M362	LAS PAMPAS	0	-78.981539	-0.431811
M113	UYUMBICHO	0	-78.534551	-0.396043
M003	IZOBAMBA	0	-78.560711	-0.372920
M354	SAN JUAN	0	-78.643808	-0.290859
M348	SANTA ANITA	0	-79.268134	-0.236195
M002	LA TOLA	0	-78.375456	-0.234241
M260	PIFO	SIN REGISTRO	-78.348860	-0.230416
M054	OBSERVARIO	SIN REGISTRO	-78.509391	-0.215971
M116	CHIRIBOGA	0	-78.794510	-0.216446
M335	LA CHORRERA	0	-78.545537	-0.207706
MA86	NAYON - GRAN	0	-78.436146	-0.182822
M024	IÑAQUITO	0	-78.492997	-0.172275

M357	CANAL 10 TV	2	-78.532467	-0.170529
M346	YARUQUI INA	0	-78.322192	-0.164521
M343	EL QUINCHE	1	-78.310635	-0.106765
M345	CALDERÓN	1	-78.430002	-0.103325
M361	NONO	0	-78.583787	-0.077468
M344	CANGAHUA	2	-78.583787	-0.077468
M358	CALACALI	1	-78.522669	-0.002450
M025	LA CONCORDIA	0	-79.398507	0.023158
MA2T	TOMALON	0	-79.398507	0.023158
M359	CAYAMBE	0	-78.143806	0.050482
M111	MALCHINGUI	SIN REGISTRO	-78.339685	0.051376
M009	LA VICTORIA	0	-78.206224	0.056350
M339	NANEGALITO	0	-78.688437	0.062629
M336	PACTO	SIN REGISTRO	-78.777037	0.143295
M023	OLMEDO	0	-78.777037	0.143295
M337	SAN JOSE	0	-78.051009	0.144497
M327	CHONTAL BAJO	2	-78.846669	0.229821

M325	GARCIA MORENO	0	-78.638334	0.231365
M326	SELVA ALEGRE	0	-78.584131	0.247778
MB85	QUININDE	SIN REGISTRO	-78.441926	0.317220
M909	GUALSAQUI	0	-78.441926	0.317220
MB76	JATÚN - SACHA	SIN REGISTRO	-79.745015	0.352772
M318	APULA - INTAG	1	-78.521142	0.357566
M441	SAGUE	1	-79.656017	0.887601

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 44. Estaciones disponibles en el año 2009 de la cuenca del Río Guayas.

CUENCA DEL RIO GUAYAS				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M397	COMPUD	3	-78.954149	-2.360048
M399	ACHUPALLAS	0	-78.783146	-2.299748
M136	CHUNCHI	0	-78.940633	-2.293921
M405	GUASUNTOS	0	-78.826525	-2.24804
M403	ALASI	0	-78.862097	-2.217791
M218	INGENIO SAN	0	-79.466038	-2.217903

M037	MILAGRO	0	-79.624912	-2.131258
M391	PALLATANGA	1	-78.982725	-2.015374
M130	CHILLANES	0	-79.0815	-1.99244
M388	RIO SAN ANTONIO	0	-79.253056	-1.966659
M402	CHIMBO DJ P	1	-79.021797	-1.956228
MB81	NOBOL	1	-79.021797	-1.956228
M409	PANGOR	1	-78.897586	-1.844677
M051	BABAHOYO	0	-79.557061	-1.812059
M468	MONTALVO	SIN REGISTRO	-79.320614	-1.790333
M404	CAÑI - LIMBE	3	-79.00728	-1.78717
MA47	CHIMBO-3	0	-79.067284	-1.712381
M476	DAULE	1	-80.025134	-1.710049
M129	CALUMA	2	-79.313255	-1.63442
MA37	LAGUACOTO	0	-79.01543	-1.629176
M171	CAMPOSANO	2	-80.436162	-1.606042
M458	COLIMES	0	-80.541998	-1.591683
M466	VINCES	0	-79.776644	-1.562034
M475	COLIMES	1	-80.037729	-1.556155
M172	PUEBLO VIEJO	0	-79.565768	-1.53142

M465	VENTANAS	0	-794.48866	-1.456789
M383	ECHANDIA	SIN REGISTRO	-79.310558	-1.438988
M385	SALINAS	2	-79.035699	-1.416597
M166	OLMEDO - MANÁ	0	-80.238643	-1.406524
M471	ZAPOTAL	3	-79.338837	-1.36508
M470	MOCACHE	0	-79.5176	-1.189393
M368	MORASPUNGO	0	-79.242996	-1.187646
M367	PINLLOPATA	0	-79.042591	-1.153513
M123	EL CORAZÓN	0	-79.103776	-1.127092
M370	RAMON CAMPA	0	-79.093472	-1.144764
M006	PICHILINGUE	0	-79.485043	-1.110055
M122	PILALO	0	-79.012117	-0.953168
M124	SAN JUAN LA	0	-79.265399	-0.926206
M374	SAN ANTONIO	0	-79.267804	0.877243
M283	INMORIEC	3	-79.374999	-0.790341
MB89	LIBERTAD	0	-79.140996	-0.538421
M026	PUERTO ILA	1	-79.35942	-0.483591
M160	EL CARMÉN	5	-79.47802	-0.286719

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 45. Estaciones disponibles en el año 2009 de la cuenca del Río Jama.

CUENCA DEL RIO JAMA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M163	BOYACA	SIN REGISTRO	-80.209487	-0.578274
M446	SAN ISIDRO	SIN REGISTRO	-80.191253	-0.376839
M167	JAMA	3	-80.306504	-0.212396

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 46. Estaciones disponibles en el año 2009 de la cuenca del Río Jipijapa.

CUENCA DEL RIO JIPIJAPA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M459	SAN PABLO	0	-80.628588	-1.592044
M169	JULCUY	0	-80.669534	-1.493677
M451	EL ANEGADO	0	-80.574968	-1.492333
M455	JOI - JIPIJAPA	0	-80.667774	-1.382487
M457	PUERTO CAYO	0	-80.774589	-1.360556
M449	SANCAN	0	-80.623744	-1.269623
M448	LA LAGUNA	1	-80.661736	-1.163699

M450	CAMARONES	1	-80.816511	-1.139767
M453	CHORILLOS	1	-80.726725	-1.064454

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 47. Estaciones disponibles en el año 2009 de la cuenca del Río Jubones.

CUENCA DEL RIO JUBONES				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M142	SARAGURO	0	-79.256419	-3.638627
M421	OÑA	SIN REGISTRO	-79.175946	-3.490447
MB08	CARTAGENA	SIN REGISTRO	-79.185882	-3.435053
M420	NABON	0	-79.086214	-3.358818
M040	PASAJE	0	-79.79733	-3.34539
M481	USHCURRUMI	0	-79.609744	-3.345793
M422	HDA. S. LUCIA	0	-79.28002	-3.296515
M032	SANTA ISABLE	0	-79.299536	-3.271554
MB86	PUCARA	0	-79.488673	-3.24653
M419	GIRON	2	-79.169462	-3.177991

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 48. Estaciones disponibles en el año 2009 de la cuenca del Río Mira.

CUENCA DEL RIO MIRA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M321	TOPO - IMBABURA	0	-78.169316	0.212251
M105	OTAVALO	0	-78.266387	0.235803
M001	INGUINCHO	0	-78.408705	0.255606
M910	MORASPUGRO	SIN REGISTROS	-78.349992	0.269414
M324	SAN FRANCISCO	1	-77.913047	0.295899
M310	MARIANO ACOSTA	0	-77.984281	0.299789
M317	COTACACHI	0	-78.272521	0.306722
M328	HDA. LA MARIA	3	-78.273506	0.350384
M021	ATUNTAQUI	1	-78.231325	0.354588
M315	PIMAMPIRO	1	-77.925938	0.385373
M314	AMBUQUI	1	-78.009612	0.432809
M086	SAN VICENTE	2	-78.045367	0.494766
M312	PABLO ARENA	1	-78.197994	0.502201

M311	CAHUASQUI	SIN REGISTRO	-78.21726	0.517909
M301	FF CC CARCHI	0	-78.136592	0.527825
M104	MIRA	0	-78.036355	0.533589
M084	BOLIVAR	SIN REGISTRO	-77.875902	0.535057
M106	LITA	0	-78.033232	0.548883
MB80	GUANDERAS	SIN REGISTRO	-77.715736	0.594778
M103	SAN GABRIEL	0	-77.819273	0.604341
M102	EL ANGEL	SIN REGISTRO	-77.94683	0.627361
M305	JULIO ANDRA	0	-77.722832	0.657319

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 49. Estaciones disponibles en el año 2009 de la cuenca del Río Muisne.

CUENCA DEL RIO MUISNE				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M153	MUISNE	1	-80.054844	0.614963

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 50. Estaciones disponibles en el año 2009 de la cuenca del Río Napo.

CUENCA DEL RIO NAPO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M485	ZATZAYACU	0	-77.859074	-1.203329
M070	TENA	SIN REGISTRO	-77.815344	-0.995356
M710	CHONTA PUNTA	0	-77.344464	-0.943784
M007	NUEVO ROCAF.	0	-75.380243	-0.929838
MB07	HUATICOCHA	2	-77.483422	-0.770785
M563	LORETO	0	-77.305345	-0.70331
MA54	SIERRAZUL	5	-77.928707	-0.68259
M203	EL REVENTAD	0	-77.970063	-0.427627
M436	CUYUJA	0	-78.053688	-0.423219
M490	SARDINAS	3	-77.801769	-0.377378
M188	PAPLLACTA	1	-78.149797	-0.37154
M293	PALMORIENTE	SIN REGISTRO	-77.05785	-0.322376
MB77	LUMBAQUI	1	-77.327364	0.03544
M697	PUERTO LIBR	0	-77.496368	0.332946

M698	LA BONITA	4	-77.47098	0.415163
M101	EL CAMELO	0	-77.597293	0.662872

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 51. Estaciones disponibles en el año 2009 de la cuenca del Río Portoviejo.

CUENCA DEL RIO PORTOVIEJO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M447	24 DE MAYO	0	-80.453425	-1.289792
MB82	LA TEODOMIRA	1	-80.425015	-1.173984
M454	RIO CHICO	0	-80.326025	-1.06399
M005	PORTOVIEJO	0	-80.50103	-1.049867
M464	RIO CHAMOTE	0	-80.265986	-1.04632
M165	ROCAFUERTE	0	-80.483854	-0.931712

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 52. Estaciones disponibles en el año 2009 de la cuenca del Río Puyango.

CUENCA DEL RIO PUYANGO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y

M180	ZARUMA	2	-79.638774	-3.725689
M773	PIÑAS	0	-79.732006	-3.702344

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 53. Estaciones disponibles en el año 2009 de la cuenca del Río Rosa.

CUENCA DEL RIO SANTA ROSA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M292	GRANJA	0	-79.932244	-3.344797

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 54. Estaciones disponibles en el año 2009 de la cuenca del Río Santiago.

CUENCA DEL RIO SANTIAGO				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M033	LA ARGELIA	0	-79.224622	-4.065504
M503	SAN FRANCISCO	0	-79.094291	-3.993325
M502	EL PANGUI	1	-78.692584	-3.963345
M190	YANZATZA	1	-78.76883	-3.865705
M432	SAN LUCAS	0	-79.285109	-3.759336
M506	PAQUISHA	1	-78.60665	-3.656346

M189	GUALAQUIZA	0	-78.590915	-3.424155
M418	CUMBE	0	-79.032691	-3.10613
M424	SIGSIG I	0	-78.802934	-3.072032
M139	GUALACEO	0	-78.79287	-2.90386
MB90	EL CEBOLLAR	SIN REGISTRO	-79.036185	-2.905009
M427	SAYAUSI	0	-79.096741	-2.887883
M426	RICAURTE	0	-78.967374	-2.873111
M429	SURUCUCHO	1	-79.152629	-2.847685
M906	JACARIN	SIN REGISTRO	-78.951765	-2.843006
M431	SEVILLA DE ORO	0	-78.651291	-2.819745
M138	PAUTE	0	-78.775469	-2.799317
M417	PISCICOLA	0	-79.193463	-2.79544
M141	EL LABRADO	1	-79.092749	-2.754009
M045	LAS PALMAS	0	-78.644598	-2.737898
M137	BIBLIAN	4	-78.909172	-2.730313
M501	MENDEZ	0	-78.313524	-2.723928
MA41	CHANLUD	1	-79.050853	-2.697557
M414	CHANIN	0	-78.762736	-2.691468

M497	LOGROÑO	0	-78.210617	-2.646819
M410	RIO MAZAR	0	-78.664809	-2.593977
M217	PEÑAS COLORADOS	0	-78.579862	-2.591994

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 55. Estaciones disponibles en el año 2009 de la cuenca del Río Taura.

CUENCA DEL RIO TAURA				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
MA2U	INGENIO	0	-79.397397	-2.393621
MA70	COSMOAGRO	SIN REGISTRO	-79.217104	-2.253026

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Tabla 56. Estaciones disponibles en el año 2009 de la cuenca del Río Zapotal.

CUENCA DEL RIO ZAPOTAL				
CÓDIGO	NOMBRE DE LA ESTACIÓN	CANTIDAD DE DATOS FALTANTES	X	Y
M173	PLAYAS	SIN REGISTRO	-80.436673	-2.647467
M175	MARESPI	SIN REGISTRO	-80.401623	-2.418013

MB06	SANTA ELENA	SIN REGISTRO	-80.949458	-2.250066
MA2V	GUAYAQUIL	0	-79.912679	-2.217371

Fuente: Instituto Nacional de Meteorología e hidrología[37].

Código

#IMPORTACIÓN DE LIBRERÍAS

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import r2_score
```

#IMPORTACIÓN DE LOS DATOS DE LAS ESTACIONES

```
precipStations=pd.read_csv('C:/Users/Alexander Rogel/Desktop/programa de
precipitación/txt/Est1_Est2_Est3.csv',index_col=0,parse_dates=True)
```

#GRÁFICA DE LOS DATOS DE LAS ESTACIONES IMPORTADAS

```
fig,ax = plt.subplots(1,1,figsize=(12,8))
corr = precipStations.corr()
sns.heatmap(corr,xticklabels=corr.columns.values,
yticklabels=corr.columns.values,ax=ax)
plt.show()
```

```
precipNotNan=precipStations.dropna()
print(precipStations.describe())
print(precipNotNan.describe())
```

#ENTRENAMIENTO Y CALIBRACIÓN DE LA RED NEURONAL


```

xTrain=precipNotNan[['Est1','Est3']]
yTrain=precipNotNan[['Est2']].values.flatten()

scaler = StandardScaler().fit(xTrain)
xTrainScaled=scaler.transform(xTrain)
print(xTrain[:10])
print(xTrainScaled[:10])
print(xTrainScaled.mean(axis=0))
print(xTrainScaled.std(axis=0))

regr=MLPRegressor(random_state=1,max_iter=5000).fit(xTrainScaled,yTrain)

xTest=precipStations[['Est1','Est3']].dropna()
xTestScaled=scaler.transform(xTest)

#GENERACIÓN DEL MODELO PREDICTIVO
yPredict=regr.predict(xTestScaled)

#GRÁFICA DEL MODELO DE ESTIMACIÓN
#fig,ax=plt.subplots(figsize=(12,8))
#ax.plot(precipStations.index,precipStations['Est2'],label='Est2')
#ax.plot(xTest.index,yPredict,label='Est2Predict')
#plt.legend()
#plt.xticks(rotation='vertical')
#plt.show()

#GENERACIÓN DE LOS VALORES FALTANTES
precipStations['Est2Completed']=0

for index, row in precipStations.iterrows():
    if np.isnan(row['Est2']) and ~np.isnan(row['Est1']) and ~np.isnan(row['Est3']):
        rowScaled=scaler.transform([[row['Est1'],row['Est3']]])

```

```

    precipStations.loc[index,['Est2Completed']] = regr.predict(rowScaled)
elif ~np.isnan(row['Est2']):
    precipStations.loc[index,['Est2Completed']] = row['Est2']
else:
    precipStations.loc[index,['Est2Completed']] = np.nan

```

#GRÁFICA COMPARATIVA DE LOS DATOS AUSENTES Y DATOS OBTENIDOS

```

print(precipStations.head())
print(precipStations.tail())

```

```

fig, axs = plt.subplots(4,1,figsize=(12,8),sharex=True,sharey=True)
axs[0].plot(precipStations.index, precipStations['Est1'],label='Est1')
axs[0].legend()
axs[1].plot(precipStations.index, precipStations['Est2'],label='Est2')
axs[1].legend()
axs[2].plot(precipStations.index,
precipStations['Est2Completed'],label='Est2Completed')
axs[2].legend()
axs[3].plot(precipStations.index, precipStations['Est3'],label='Est3')
axs[3].legend()
plt.legend()
plt.xticks(rotation='vertical')
plt.show()

```

#EXPORTACIÓN DE LOS RESULTADOS

```

precipStations.to_csv('C:/Users/Alexander Rogel/Desktop/programa de
precipitación/txt/completedStations1.csv')

```