

EFICIENCIA DE LOS MÉTODOS ESTADÍSTICOS DE IMPUTACIÓN DE LA ESTACIÓN METEOROLÓGICA DE TUMPEKA

Campos,J.¹, Bellido, J.², Espada,C.³, Huaranca, J.⁴, Ibarra,A.⁵

- 1 Docente investigador en el área de Hidráulica de la Facultad de Ingeniería Civil (USFX), Destacamento 317 N° 573, Campus Universitario Ex REFISUR, Sucre, Bolivia. E-mail: campos.edgar@usfx.bo.
- 2 Docente investigador en el área de Ingeniería de Sistemas de la Facultad de Ingeniería Civil (USFX), Destacamento 317 N° 573, Campus Universitario Ex REFISUR, Sucre, Bolivia. E-mail: bellido.boris@usfx.bo.
- 3 Estudiante investigadora en el área de Hidráulica de la Facultad de Ingeniería Civil (USFX), Destacamento 317 N° 573, Campus Universitario Ex REFISUR, Sucre, Bolivia. E-mail: carlyveronica93@gmail.com.
- 4 Estudiante investigadora en el área de Hidráulica de la Facultad de Ingeniería Civil (USFX), Destacamento 317 N° 573, Campus Universitario Ex REFISUR, Sucre, Bolivia. E-mail: huaranca@gmail.com .
- 5 Estudiante investigador en el área de Hidráulica de la Facultad de Ingeniería Civil (USFX), Destacamento 317 N° 573, Campus Universitario Ex REFISUR, Sucre, Bolivia. E-mail: sensey.dumb12@gmail.com

RESUMEN

La investigación pretende hacer conocer la eficiencia, al aplicar el o los métodos de relleno de datos faltantes de precipitación pluvial, tomando en cuenta las características del lugar, ya que esta información es crucial para la gestión adecuada de los recursos hídricos. Los datos de precipitación se utilizan para estimar la disponibilidad de agua en una región determinada, lo que es importante para la planificación y gestión de proyectos de infraestructura hidráulica, como represas, canales, sistemas de riego y abastecimiento de agua.

En Sucre-Bolivia, la captación de estos datos pluviales se dan a través de pluviógrafos, de estaciones pluviométricas y recientemente por telemetría. Los cuales están sujetos a errores humanos y sistemáticos.

Este artículo analiza la utilidad y eficiencia de los métodos Razón Normal, U.S. National Weather Service y Regresión Lineal, según las características de la zona estudiada, para el relleno de datos faltantes de precipitación de la estación de Tumpeka ubicada en la cuenca hidrográfica de Ravelo, la cual es muy importante para el municipio de Sucre ya que es su principal fuente de abastecimiento de agua potable.

De esta manera se pretende maximizar el aprovechamiento de los datos meteorológicos y que estos sean utilizados como una alternativa eficaz para las estaciones de Bolivia en las cuales no existen registros totales de precipitación pluvial o si existen, estas cuentan con escasa información.

Palabras clave: Tumpeka, eficiencia, imputación, meteorología.

SUMMARY

The research aims to make known the efficiency, by applying the method(s) of filling in missing rainfall data, taking into account the characteristics of the place, since this information is crucial for the proper management of water resources. Precipitation data is used to estimate water availability in a given region, which is important for planning and managing hydraulic infrastructure projects such as dams, canals, irrigation and water supply systems.

In Sucre-Bolivia, these rainfall data are collected through rain gauges, rainfall stations and recently by telemetry. Which are subject to human and systematic errors.

This article analyzes the usefulness and efficiency of the Normal Ratio, U.S. methods. National Weather Service and Linear Regression, according to the characteristics of the studied area, to fill in missing precipitation data from the Tumpeka station located in the Ravelo hydrographic basin, which is very important for the municipality of Sucre since it is its main source of drinking water supply.

In this way, the aim is to maximize the use of meteorological data and for these to be used as an effective alternative for the Bolivian stations in which there are no total records of rainfall or, if they do exist, they have little information.

Keywords: Tumpeka, efficiency, imputation, meteorology

INTRODUCCIÓN

“Las estaciones meteorológicas desempeñan un papel fundamental en la recopilación y seguimiento de datos atmosféricos cruciales para comprender los patrones climáticos y proporcionar pronósticos precisos del tiempo. Sin embargo, la calidad de los datos recopilados puede verse comprometida debido a diversos factores, incluidas las fallas técnicas, el mantenimiento irregular y las condiciones ambientales extremas.” (Climatología y Meteorología - I.E.S. T-004 Normal «Gral. Toribio de Luzuriaga», s. f.) Según (Serrano Vincenti et al., 2012) la presencia de datos faltantes en los registros meteorológicos puede tener un impacto significativo en la precisión de los análisis climáticos y en la fiabilidad de los pronósticos.

El relleno de datos es estimar la precipitación de los registros faltantes con el menor error posible respecto de su ocurrencia real con las estaciones más cercanas y consistentes. “Si al graficar los valores acumulados de lluvia en ambos ejes, la

recta obtenida presenta una sola pendiente, los registros históricos de la serie analizada son consistentes.” (Chávez-De La Lima et al., 2022) Dentro de las técnicas para relleno de datos pluviométricos existe una amplia variedad; sin embargo, la elección del método debería depender de las características geomorfológicas del área de estudio.

U.S.National Weather Service.

Este procedimiento ha sido verificado teóricamente como empíricamente y considera que el dato faltante de una estación X por ejemplo, puede ser estimada en base a los datos observados en las estaciones circundantes. (Quispe Mamani, 2018)

El método puede ser aplicado para estimar valores diarios, mensuales o anuales faltantes. El método consiste en ponderar los valores observados en una cantidad W, igual al recíproco del cuadrado de la distancia entre cada estación vecina y la estación(NAMA-ADMIN, s. f.)

Como se observa en la ecuación:

$$P_x = \frac{\sum P_i * w_i}{\sum w_i} \quad (1)$$

Pi = Precipitación observada para la fecha de la faltante, en las estaciones auxiliares circundantes.

wi = 1/Di Siendo Di la distancia entre cada estación circundante y la estación incompleta, en km.

Como la serie histórica de precipitación de la Estación Meteorológica de Tumpeka, presentó datos faltantes que difieren en más del 10% de esta Estación, se utilizó el método del U.S. Weather(1) para las estaciones meteorológicas con radio de distancia menor a 20 kilómetros, como se muestra en la tabla 7.

Tabla 7

Distancia entre estaciones y la estación de Tumpeka.

Tumpeka	Potolo	10.73 km.
	Cajamarca	5,89 km.
	Ravelo	18,445 km.

Nota: verificación de la distancia menor a 20 km. Razón normal:

Este método se basa en la relación constante entre los valores de precipitación de dos estaciones, una de las cuales tiene datos completos y la otra no.(Carrera-Villacrés et al., 2016)

Los valores mensuales, anuales o medios tiende a ser constante. Para aplicar este método, se necesitan tres o más estaciones cercanas y confiables, que estén uniformemente espaciadas con respecto a la estación en estudio. (Carrera-Villacrés et al., 2016)

$$P_x = \frac{N_x}{n} * \sum_{i=1}^n \frac{P_i}{N_i} \quad (2)$$

Donde:

Px = registro de precipitación a ser completado

n = número de estaciones auxiliares

Pi = precipitaciones de las estaciones vecinas
Regresión Lineal

En el modelo de regresión lineal se requiere de una estación “donante” y preferiblemente cercana a la estación con datos faltantes (Mendenhall,1990).

La fórmula es:

$$y = \beta_0 + \beta_1 x \quad (3)$$

Donde:

“y” es el dato de precipitación a rellenar.

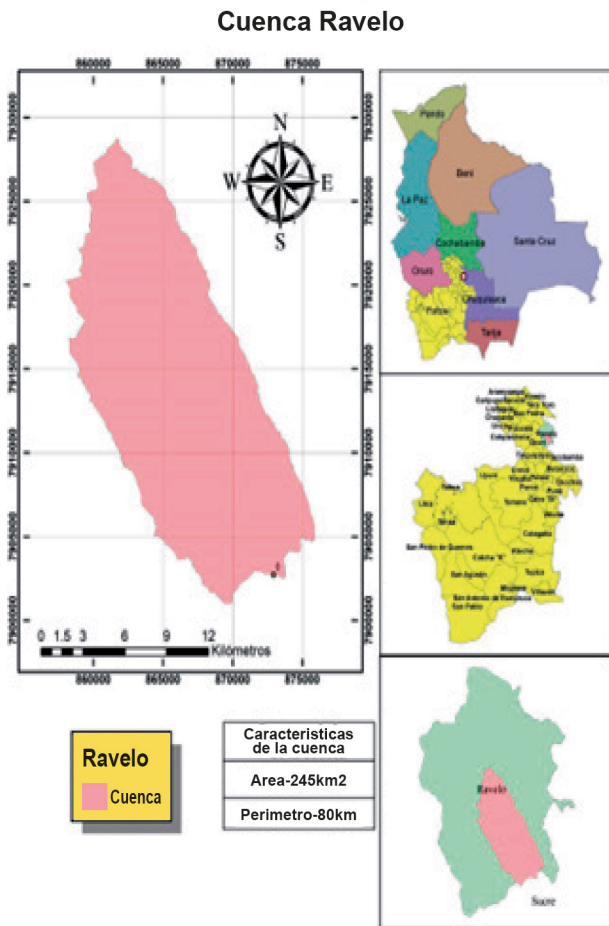
Los coeficientes “B0” y “B1” se obtienen de la regresión.

La variable independiente “x” son las precipitaciones de la estación “donante”.

Este método es mayormente usado cuando no se disponen de muchos datos de precipitación.

La cuenca Ravelo, forma parte de la cuenca Cachimayu, según datos obtenidos de del ministerio y medio ambiente de agua (MMAyA) perteneciente a los departamentos de Chuquisaca y Potosí y recorre los municipios de Sucre, Ocurí, Ravelo; se ubica entre las coordenadas geográficas 18°48’ 17” a 18°58’99” de latitud Sur y 65°51’3” a 65°180’30 de longitud Oeste, ocupa un área aproximada de 245 Km2 y un perímetro de 80 Km.

Figura 1
Características de la cuenca de estudio



Nota: esta imagen muestra las características de la cuenca de estudio y su ubicación geográfica

La cuenca Ravelo representa un total del 15% de la Macrocuenca de Cachimayu, siendo la fuente principal de abastecimiento de recursos hídricos para Sucre.

Como se encuentra ubicada en una zona montañosa, dificulta la obtención de datos y su accesibilidad. Y a pesar de que la cuenca de Ravelo, cuenta con estaciones meteorológicas, los datos recopilados tienden a ser incompletos.

A causa de ese problema, se generan incertidumbres en la toma de decisiones para la Obra de Toma de Tumpeka.

El objetivo general es identificar la eficacia de diferentes métodos estadísticos de imputación en registros mensuales de la estación meteorológica de Tumpeka.

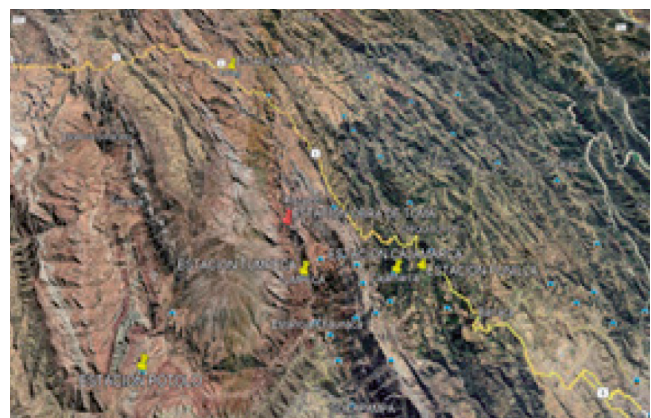
MATERIALES Y MÉTODOS

Se utilizó un programa comparativo detectando las precipitaciones mensuales completas de la estación Tumpeka, año 2022. (tabla 1). Se desechó los valores de dicha gestión, para utilizar los métodos estadísticos y que de esta manera se pueda realizar la comparación entre valores obtenidos por los métodos estadísticos y los valores verdaderos.

Obtención de datos:

Para la investigación se solicitó la información necesaria de la empresa ELAPAS (Empresa Local de Agua Potable y Alcantarillado de SUCRE) gracias a solicitudes y los convenios existentes entre la universidad y la empresa, se obtuvo información de las bases de datos de las estaciones de Ravelo, Tumpeka, Potolo y Cajamarca como se muestra en la figura 2, marcadas de color amarillo.

Figura 2
Estaciones meteorológicas



Nota: esta imagen muestra las estaciones de estudio.

Se utilizó el software PC-400 debido a que los datos se encontraban en formato datalogger, este versátil software admitió la visualización en un formato más accesible para ejecutar el trabajo de más 14000 datos de precipitación

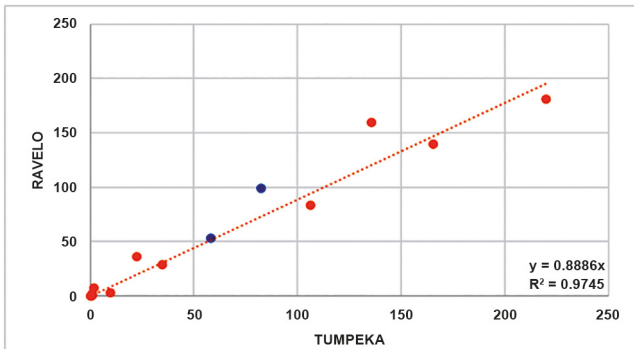
medidas cada 5 minutos de las gestiones 2019 a las 2023.

Una vez exportados los datos a Excel y con ayuda de las herramientas Python y la biblioteca Pandas, se armaron las series de precipitación mensual total, se verificó la falta de registros de los últimos años en las distintas estaciones. Como se puede ver en las tablas 1, 2, 3 y 4 .

Análisis de consistencia:

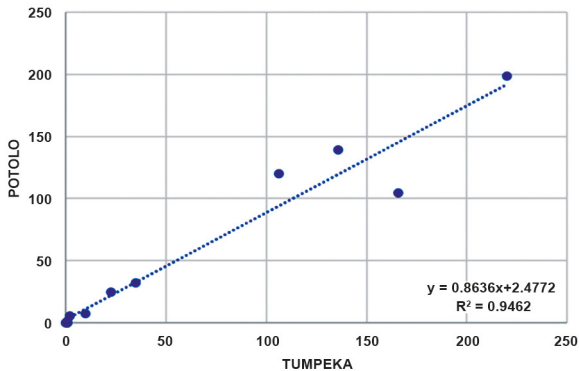
Se aplicó el análisis de consistencia para la estación de Tumpeka, con las estaciones de Ravelo, como muestra ver la figura 3; con las estación de Potolo, como se muestra en la figura 4 y con la estación de Cajamarca, como se muestra en la figura 5; obteniendo en todos los casos datos consistentes.

Figura 3
Consistencia de la cuenca de Ravelo y Tumpeka



Nota: esta imagen muestra la consistencia existente entre ambas estaciones

Figura 4
Consistencia de la cuenca de Potolo y Tumpeka



Nota: esta imagen muestra la consistencia existente entre ambas estaciones

Figura 5
Consistencia de la cuenca de Cajamarca y Tumpeka

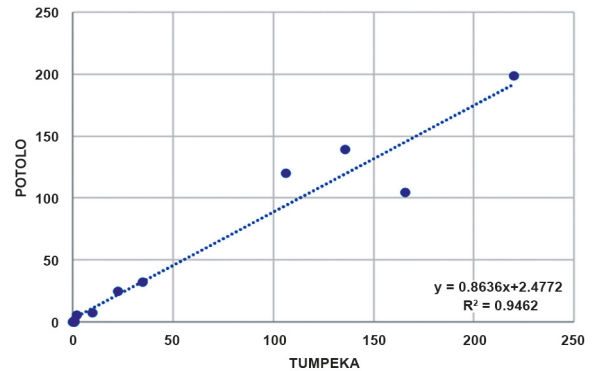


Tabla 3
Valores de precipitaciones de la estación de Potolo.

Estación:Potolo	Precipitación Mensual Total				
Años	2019	2020	2021	2022	2023
Enero	124.8	177.1	111.2	198.7	89.3
Febrero	218.3	161.3	135.9	139.2	145
Marzo	72.7	102.8	91.6	104.5	146.1
Abril	110	7.5	23.2	24.9	7.3
Mayo	20	0.2	0	0.8	1.2
Junio	5.2	0	0	0	0
Julio	10.9	0	0	0	1.2
Agosto	0	0	14	0	
Septiembre	25.8	44.8	49.4	5.5	
Octubre	76	101.2	22.5	32.2	
Noviembre	68.8	27	58.3	7.3	
Diciembre	78.3	88.8	426.3	120.1	
PMT	67.5666667	59.225	77.7	52.76667	

Nota: esta tabla muestra la cantidad de datos existentes en la estación.

Tabla 4
Valores de precipitaciones de la estación de Tumpeka.

Estación:Ravelo	Precipitación Mensual Total				
Años	2019	2020	2021	2022	2023
Enero		276.7	172.2	181.2	57
Febrero		290.1	188.5	159.5	171
Marzo		103.4	166.3	139.9	159.7
Abril	68.7	25.7	29.7	35.9	24.9
Mayo	16.3	3.7	0.1	2.3	
Junio	6.5	0.9	1	0.7	
Julio	15.8	0	0.3	0.1	
Agosto	0.4	11.3	15.7	0.1	
Septiembre	30.3	91.6	33.4	7.2	
Octubre	50.8	153.4	28.6	28.9	
Noviembre	125.2	7.6	45.3	2.9	
Diciembre	119.7	160.8	511.9	83.3	
PMT		93.7667	99.4167	53.5	

Nota: esta tabla muestra la cantidad de datos existentes en la estación.

RESULTADOS

Los valores obtenidos, después de utilizar los diferentes métodos para el análisis, mostraron la existencia de errores de hasta el 650% en el método U.S.National Weather Service. El método adopta errores elevados en los meses de Abril Mayo y Junio (épocas secas). No obstante los meses de septiembre noviembre diciembre enero (épocas lluviosas), se presentan con un bajo error relativo, tanto que el error para esos meses difiere entre el 4% y 15%(Tabla 6).

El método de la Razón Normal muestra un error de hasta el 275%, los valores más elevados están en los meses de Enero, Marzo, Abril, Mayo, Junio y Septiembre (épocas secas y lluviosas). No obstante para los meses de Febrero, Octubre, Noviembre y Diciembre (épocas lluviosas), se presenta un bajo error relativo.

En el método de la regresión lineal se obtienen porcentajes altos de error en los meses de Enero y Abril, también se muestran errores bajos en los meses de Octubre, Noviembre y Diciembre.

Cabe destacar que, aunque en los meses de Mayo, Junio, Julio y Agosto se presentan porcentajes de error bastante altos en los 3 métodos, en realidad representan una pequeña diferencia de unos 2 o 3 milímetros respecto al valor original

Tabla 5
Valores comparativos aplicando el método U.S.National Weather Service, para la gestión 2022 en la estación de Tumpeka.

2022	TUMPEKA VALOR OBTENIDO	TUMPEKA VALOR VERDADERO	Error	Error relativo %
ENE.	211	220	9.00	4%
FEB.	230	135.7	94.30	69%
MAR.	144	165.6	21.60	13%
ABR.	53	22.5	30.50	136%
MAY.	6	0.8	5.20	650%
JUN.	3	0.8	2.20	275%
JUL.	0	0	0	0%
AGO.	0	0.6	0.60	100%
SEP.	2	1.9	0.10	5%
OCT.	47	34.9	12.10	35%
NOV.	9	9.8	0.80	8%
DIC.	123	106.2	16.80	16%

Nota: esta tabla muestra el relleno de datos y los resultados del error relativo porcentual, aplicando el método U.S.National Weather Service.

Tabla 6
Valores comparativos aplicando el método de la Razón Normal, para la gestión 2022 en la estación de Tumpeka.

2022	TUMPEKA VALOR OBTENIDO	TUMPEKA VALOR VERDADERO	Error	Error relativo %
ENE.	146	220	74.00	34%
FEB.	149	135.7	13.30	10%
MAR.	108	165.6	57.60	35%
ABR.	43	22.5	20.50	91%
MAY.	3	0.8	2.20	275%
JUN.	3	0.8	2.20	275%
JUL.	0	0	0	0%
AGO.	0	0.6	0.60	100%
SEP.	6	1.9	4.10	216%
OCT.	37	34.9	2.10	6%
NOV.	10	9.8	0.20	2%
DIC.	100	106.2	6.20	6%

Nota: esta tabla muestra el relleno de datos y los resultados del error relativo porcentual, aplicando el método de la Razón Normal.

Tabla 7
Valores comparativos aplicando el método Regresión lineal, para la gestión 2022 en la estación de Tumpeka.

2022	TUMPEKA VALOR OBTENIDO	TUMPEKA VALOR VERDADERO	Error	Error relativo %
ENE.	221	220	1.31	1%
FEB.	158	135.7	22.18	16%
MAR.	178	165.6	12.61	8%
ABR.	73	22.5	50.66	225%
MAY.	0	0.8	0.80	100%
JUN.	2	0.8	1.20	150%
JUL.	0	0	0	0%
AGO.	0	0.6	0.60	100%
SEP.	0	1.9	1.90	100%
OCT.	26	34.9	8.76	25%
NOV.	18	9.8	7.98	81%
DIC.	146	106.2	40.03	38%

Nota: esta tabla muestra el relleno de datos y los resultados del error relativo porcentual, aplicando el método de Regresión lineal.

CONCLUSIONES:

Al realizar el análisis se pudo observar, que los métodos varían su eficacia dependiendo de la época (lluviosa y seca) y también de la cantidad de datos disponibles en las estaciones más cercanas .

Esta discrepancia de valores obtenidos por los diferentes métodos en la cuenca estudiada, identifica la no existencia de un único método

recomendable para el relleno de datos de la zona específicamente estudiada.

REFERENCIAS BIBLIOGRÁFICAS

Degaetano, A. T.; Eggleston, K. L. and Knapp, W. W. 1995. A method to estimate missing maximum and minimum temperature observations. *J. Appl. Meteorol.* 34(2):371-380.

Barrios, A., Trincado, G., & Garreaud, R. (2018). Alternative approaches for estimating missing climate data: Application to monthly precipitation records in South-Central Chile. *Forest Ecosystems*, 5(1), 28. <https://doi.org/10.1186/s40663-018-0147-x>

Chávez-De La Lima, J. D., Olguín-López, J. L., Guevara-Gutiérrez, R. D., Mancilla-Villa, Ó. R., Palomera-García, C., Cruz-Sandoval, G., Chávez-De La Lima, J. D., Olguín-López, J. L., Guevara-Gutiérrez, R. D., Mancilla-Villa, Ó. R., Palomera-García, C., & Cruz-Sandoval, G. (2022). Consistencia, homogeneidad y distribución de la precipitación pluvial y temperatura, Región Sierra de Amula, Jalisco, México. *Revista Geográfica de América Central*, 68, 353-378. <https://doi.org/10.15359/rgac.68-1.13>

Climatología y Meteorología – I.E.S. T-004 Normal «Gral. Toribio de Luzuriaga». (s. f.). Recuperado 19 de agosto de 2023, de <https://ens9004-inf.d.mendoza.edu.ar/sitio/climatologia-y-meteorologia/>
Guía de prácticas climatológicas. (2018).

NAMA-ADMIN. (s. f.). Libro: Hidrología para estudiantes de ingeniería civil - Wendor Chereque Morán. Hidro SM. Recuperado 20 de agosto de 2023, de <https://www.hidrosm.com/2021/02/libro-hidrologia-para-estudiantes-de.html>
Quispe Mamani, J. C. (2018). Curvas IDF para la estación meteorológica de Viacha, departamento de La Paz. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, 5(1), 99-105.

Serrano Vincenti, S., Zuleta, D., Moscoso, V., Jácome, P., Palacios, E., & Villacís, M. (2012). Análisis estadístico de datos meteorológicos mensuales y diarios para la determinación de variabilidad climática y cambio climático en el Distrito Metropolitano de Quito. *La Granja*, 16(2), 23. <https://doi.org/10.17163/lgr.n16.2012.03>.