XVII

Discrepancias en los periodos de retorno de precipitaciones isométricas: Un análisis comparativo mediante los métodos de Shapiro-Wilk y Test de Levene en distritos altoandinos

Luis Alberto Acosta Sanchez

XVII

Discrepancias en los periodos de retorno de precipitaciones isométricas: Un análisis comparativo mediante los métodos de Shapiro-Wilk y Test de Levene en distritos altoandinos

Luis Alberto Acosta Sanchez

RESUMEN

Este capítulo de revisión narrativa examina las discrepancias técnicas en la estimación de periodos de retorno para precipitaciones isométricas en distritos altoandinos, un desafío crítico para el diseño de infraestructura hidráulica y la gestión de riesgos. A través de un análisis comparativo de la literatura, se explora cómo la aplicación de las pruebas de Shapiro-Wilk (normalidad) y el Test de Levene (homocedasticidad) suele generar resultados contradictorios debido a la naturaleza estocástica y la alta variabilidad climática de las zonas de alta montaña. Se evidencia que la dependencia exclusiva de estos métodos estadísticos, sumada a la sensibilidad de Shapiro-Wilk ante muestras pequeñas, conduce a una selección inadecuada de funciones de distribución, provocando errores significativos en la cuantificación de eventos extremos. Ante este escenario, el estudio argumenta a favor de un enfoque de "pesos de evidencia" que integre pruebas complementarias, como el test de Anderson-Darling y métodos no paramétricos, para mejorar la precisión estadística. Finalmente, se destaca la importancia de incorporar futuras líneas de investigación basadas en modelos de inteligencia artificial y datos satelitales, con el objetivo de reducir la incertidumbre en cuencas de cabecera carentes de estaciones físicas y fortalecer la resiliencia de la planificación frente a la variabilidad climática regional.

Palabras clave: Precipitaciones isométricas, periodos de retorno, Shapiro-Wilk, Test de Levene, hidrología altoandina.

ABSTRACT

This narrative review chapter examines the technical discrepancies in the estimation of return periods for isometric precipitation in high-altitude Andean districts, a critical challenge for hydraulic infrastructure design and risk management. Through a comparative analysis of the literature, it explores how the application of the Shapiro-Wilk test (normality) and Levene's test (homoscedasticity) often generates contradictory results due to the stochastic nature and high climatic variability of high-mountain areas. It demonstrates that the exclusive reliance on these statistical methods, coupled with the sensitivity of the Shapiro-Wilk test to small samples, leads to an inappropriate selection of distribution functions, causing significant errors in the quantification of extreme events. Given this scenario, the study argues for a "weights of evidence" approach that integrates complementary tests, such as the Anderson-Darling test and nonparametric methods, to improve statistical accuracy. Finally, the importance of incorporating future lines of research based on artificial intelligence models and satellite data is highlighted, with the aim of reducing uncertainty in headwater basins lacking physical stations and strengthening the resilience of planning to regional climate variability.

Keywords: Isometric precipitation, return periods, Shapiro-Wilk, Levene test, high Andean hydrology.

INTRODUCCIÓN

La gestión de los recursos hídricos en los distritos altoandinos representa uno de los desafíos más críticos para la ingeniería hidráulica contemporánea. En estos ecosistemas de alta montaña, el cálculo preciso de los periodos de retorno para eventos de precipitación es el pilar fundamental para el diseño de infraestructura segura —como presas, sistemas de drenaje y alcantarillado— y para la implementación de estrategias efectivas de gestión de riesgos ante desastres naturales, particularmente inundaciones y avenidas torrenciales (Villarreal et al., 2021). Sin embargo, la singularidad climática de la región andina, caracterizada por una compleja orografía y una marcada variabilidad estacional, impone retos significativos al modelamiento hidrológico convencional.

La problemática central radica en que las precipitaciones isométricas, aquellas que mantienen patrones de distribución específicos, frecuentemente exhiben comportamientos no lineales y una alta heterogeneidad temporal. La aplicación rígida de métodos estadísticos tradicionales, que a menudo presuponen la normalidad y homocedasticidad de los datos, suele resultar en una subestimación o sobreestimación peligrosa de los eventos extremos. Esta inadecuación metodológica se debe, en gran medida, a que las series históricas de precipitación en zonas altoandinas raramente cumplen con los supuestos teóricos subyacentes en las distribuciones de probabilidad clásicas (Gumbel, Log-Pearson tipo III, o Log-Normal), las cuales son estándar en la práctica hidrológica (Montes & Rodríguez, 2023).

En este escenario, el proceso de selección del modelo probabilístico óptimo suele estar precedido por pruebas de bondad de ajuste, siendo el test de Shapiro-Wilk (para evaluar la normalidad) y la prueba de Levene (para evaluar la homocedasticidad) herramientas de uso frecuente. No obstante, existe una brecha notable en la literatura especializada respecto a cómo las discrepancias detectadas por estas pruebas —donde a menudo un conjunto de datos puede pasar un test de varianza, pero fallar en el de normalidad— afectan directamente la validación y la incertidumbre del modelo seleccionado. Esta contradicción estadística no es trivial; puede llevar al investigador a elegir una función de distribución que no representa fielmente la cola de la distribución de las precipitaciones extremas, comprometiendo la fiabilidad de las obras de ingeniería proyectadas (Heredia et al., 2022).

Ante esta situación, el presente artículo tiene como objetivo analizar comparativamente las implicancias técnicas y las limitaciones del uso de las pruebas de Shapiro-Wilk y Levene en el análisis de frecuencias de precipitaciones en regiones de alta montaña. A través de una revisión narrativa, se pretende esclarecer si la sensibilidad de estos métodos estadísticos ante sesgos inherentes a los datos altoandinos constituye una barrera infranqueable o si, mediante un enfoque de interpretación complementaria, es posible reducir la incertidumbre en la estimación de periodos de retorno. Con este propósito, se exploran las tensiones metodológicas entre la validez estadística pura y la realidad hidrológica de los Andes, proponiendo criterios más robustos para la toma de decisiones en entornos de alta incertidumbre.

METODOLOGÍA

La presente investigación se desarrolló bajo un enfoque cualitativo mediante un diseño de revisión narrativa, estructurado para sintetizar la literatura científica relevante sobre la validación estadística de precipitaciones en entornos de alta montaña. El proceso metodológico se dividió en tres fases críticas: definición de la estrategia de búsqueda, establecimiento de criterios de selección y procesamiento analítico de la información recopilada.

Para la obtención de los insumos bibliográficos, se ejecutó una estrategia de búsqueda sistemática en bases de datos de alto impacto, incluyendo Scopus, Web of Science y SciELO. Se empleó una combinación estratégica de términos descriptores (decs/mesh) tanto en español como en inglés, utilizando operadores booleanos para refinar los resultados. Los términos clave fueron: precipitaciones isométricas, periodos de retorno, Shapiro-Wilk, Test de Levene e hidrología altoandina. Esta selección permitió capturar no solo estudios centrados en la estadística pura, sino también aquellos que contextualizan estas herramientas dentro de la realidad física de los Andes (Gómez & Valdivia, 2020).

En cuanto a la selección de estudios, se delimitaron criterios de inclusión estrictos. Se consideraron únicamente artículos científicos publicados durante los últimos 15 años (2009-2024), garantizando la vigencia de los métodos estadísticos analizados y su aplicabilidad ante los cambios en los patrones climáticos observados recientemente. Se priorizaron investigaciones que aplicaran explícitamente pruebas de bondad de ajuste —específicamente Shapiro-Wilk para evaluar la normalidad— y pruebas de homogeneidad de varianzas, como el Test de Levene, en series meteorológicas de cuencas de cabecera. Se descartaron documentos técnicos breves, notas de opinión y trabajos que no detallaran los procedimientos estadísticos subyacentes, siguiendo las recomendaciones de Martínez y Pardo (2022) sobre la calidad de la evidencia en revisiones hidrológicas.

Finalmente, el procesamiento de la información se realizó mediante una matriz comparativa diseñada para categorizar los estudios seleccionados. Esta clasificación se segmentó bajo tres variables principales: primero, la altitud de las estaciones meteorológicas, con el fin de identificar si existen umbrales altitudinales donde las pruebas estadísticas tienden a perder sensibilidad; segundo, el tamaño de la muestra de las series históricas, dado que la robustez de Shapiro-Wilk es altamente dependiente del número de datos (Razali & Wah, 2011); y tercero, la metodología estadística

empleada como alternativa o complemento para validar los datos de entrada cuando las pruebas primarias presentaban resultados discrepantes.

Este enfoque metodológico permitió no solo sistematizar el conocimiento existente, sino también evaluar la heterogeneidad de los procedimientos aplicados. Al organizar la literatura según estas dimensiones, el estudio logra un análisis exhaustivo de cómo la variabilidad inherente a los datos de precipitación en distritos altoandinos condiciona la eficacia de los estadísticos convencionales, aportando una base sólida para discutir la selección de modelos probabilísticos en entornos de alta incertidumbre climática.

RESULTADOS

El análisis de la literatura revela una tendencia marcada en las estaciones hidrometeorológicas situadas en distritos altoandinos: la recurrente violación del supuesto de normalidad en las series de precipitación máxima diaria. Los hallazgos, basados en la aplicación del test de Shapiro-Wilk, demuestran que las precipitaciones isométricas en zonas de alta montaña rara vez siguen una distribución Gaussiana. Esta desviación no es meramente técnica, sino el reflejo de la compleja dinámica atmosférica y topográfica de los Andes, que introduce asimetrías y colas pesadas en las series de datos (Morales & Torres, 2024). La invalidez de la normalidad compromete directamente la fiabilidad de pruebas paramétricas clásicas que dependen de este supuesto para otorgar validez estadística, obligando a los investigadores a recurrir a transformaciones de datos o a métodos no paramétricos.

Un hallazgo crítico identificado en la revisión es la existencia de discrepancias sistemáticas entre las pruebas de normalidad y las de homocedasticidad. Con frecuencia, el test de Levene —utilizado para evaluar la igualdad de varianzas entre diferentes estratos o periodos— arroja resultados que sugieren homocedasticidad (homogeneidad de varianza), mientras que, simultáneamente, el test de Shapiro-Wilk rechaza categóricamente la normalidad. Esta paradoja genera una incertidumbre significativa en el ajuste de las funciones de distribución (como las familias Gumbel o Pearson). Como señalan Castillo y Zapata (2022), cuando un modelo estadístico se basa en un test de Levene positivo, pero ignora el sesgo distributivo detectado por Shapiro-Wilk, se incurre en una "falsa sensación de ajuste", donde el modelo parece ser estadísticamente estable, pero carece de precisión en la estimación de los cuantiles extremos.

Este fenómeno de discrepancia técnica se traduce en un impacto directo y negativo sobre el cálculo de los periodos de retorno. La síntesis de los estudios analizados evidencia que, ante la imposibilidad de validar los supuestos estadísticos con consistencia, se suele incurrir en la selección inadecuada de la distribución estadística. Por ejemplo, la elección arbitraria de una distribución de Gumbel para datos que presentan curtosis elevada y rechazo de normalidad conduce sistemáticamente a la subestimación de eventos hidrometeorológicos extremos. En el contexto de cuencas de cabecera, esta imprecisión se manifiesta en una cuantificación errónea del riesgo de inundación o, inversamente, en una sobreestimación de los caudales de diseño que eleva innecesariamente los costos de infraestructura (Salazar et al., 2023).

En conclusión, los resultados indican que la divergencia entre Shapiro-Wilk y Levene no es un fenómeno aislado, sino una señal de advertencia sobre la fragilidad del análisis de frecuencias en climas altoandinos. La falta de consenso metodológico al interpretar estos tests estadísticos actúa como un eslabón débil en la cadena de diseño hidráulico, subrayando la necesidad urgente de integrar criterios de selección de distribuciones que sean menos sensibles a los supuestos de normalidad y más resilientes ante la heterocedasticidad oculta en la variabilidad climática de montaña.

DISCUSIÓN

La naturaleza estocástica y la alta variabilidad temporal de las precipitaciones en distritos altoandinos plantean un desafío que trasciende la aplicación mecánica de la estadística tradicional. La revisión realizada sugiere que las pruebas convencionales, como Shapiro-Wilk y Levene, a menudo fallan al intentar capturar la complejidad hidrológica de los Andes debido a que sus supuestos teóricos fueron diseñados bajo entornos de control que no siempre replican las dinámicas de montaña. Las precipitaciones de alta montaña no solo presentan una distribución irregular, sino que a menudo exhiben dependencias de escala y procesos de memoria a largo plazo que la estadística clásica, al asumir independencia y normalidad, es incapaz de integrar correctamente (Hidalgo & Pizarro, 2023).

Una de las principales limitaciones radica en la sensibilidad intrínseca de Shapiro-Wilk frente a muestras pequeñas, una condición frecuente en las estaciones meteorológicas altoandinas debido a la falta de registros históricos continuos. Como han demostrado estudios recientes, el test de Shapiro-Wilk tiende a ser demasiado conservador en

muestras reducidas, rechazando la hipótesis de normalidad ante cualquier perturbación menor en los datos, lo que a menudo induce al analista a descartar distribuciones que podrían ser perfectamente adecuadas (Bautista et al., 2021). Por el contrario, el Test de Levene, aunque es valorado por su relativa robustez frente a desviaciones de la normalidad al evaluar la homocedasticidad, puede proporcionar una falsa sensación de estabilidad. En contextos donde los datos presentan asimetría, Levene puede no detectar inestabilidades latentes que sí afectan la estructura de la varianza en las colas de la distribución, puntos críticos donde se definen los eventos extremos de precipitación.

Ante este panorama de discrepancias, se vuelve imperativo superar la dependencia exclusiva de estas dos pruebas mediante la integración de métodos estadísticos complementarios. La literatura sugiere que la adopción de pruebas de bondad de ajuste con mayor potencia en las colas, como el test de Anderson-Darling, es fundamental para compensar la rigidez observada en Shapiro-Wilk. Anderson-Darling otorga un mayor peso a los extremos, lo cual es vital cuando el objetivo es el diseño de infraestructura hidráulica frente a avenidas extremas (Gutiérrez & Ruiz, 2024).

Asimismo, el uso de pruebas no paramétricas —como el test de Mann-Kendall para analizar tendencias o la prueba de Wilcoxon para comparar medianas— ofrece una alternativa robusta que no requiere de supuestos distributivos estrictos, permitiendo validar las series meteorológicas con una mayor flexibilidad ante la heterogeneidad de los datos altoandinos. La integración de estos métodos, lejos de simplificar el proceso, aporta una capa de profundidad necesaria para validar el modelo antes de proyectar periodos de retorno. En última instancia, la discrepancia detectada no debe entenderse como un error del método, sino como una señal para transitar hacia un enfoque probabilístico multivariado que considere la no linealidad del clima de alta montaña como una propiedad inherente y no como un error de muestreo que deba ser corregido mediante ajustes estadísticos simplistas.

CONCLUSIONES

La investigación realizada permite concluir que las discrepancias observadas entre las pruebas de Shapiro-Wilk y Levene en el análisis de precipitaciones isométricas no representan una deficiencia del método estadístico en sí, sino una manifestación directa de la naturaleza estocástica y la alta variabilidad física propia de los ecosistemas

altoandinos. Estas pruebas, diseñadas bajo supuestos de regularidad, operan al límite de su capacidad cuando se aplican a series meteorológicas de alta montaña, donde los eventos extremos son inherentemente irregulares. Por lo tanto, tales discrepancias deben interpretarse como indicadores de alerta sobre la complejidad de los datos, más que como errores de cálculo.

Ante este panorama, se recomienda encarecidamente que los ingenieros y planificadores abandonen la práctica de depender de una única prueba de bondad de ajuste o de homogeneidad. En su lugar, se propone adoptar un enfoque de "pesos de evidencia" estadística. Este método implica evaluar la consistencia de los datos mediante una batería de tests (paramétricos y no paramétricos) que, al ser interpretados en conjunto, reduzcan el sesgo en la selección del modelo probabilístico, garantizando diseños hidráulicos más seguros y resilientes.

Finalmente, este estudio subraya la necesidad crítica de transitar hacia futuras líneas de investigación que integren modelos de inteligencia artificial y técnicas de machine learning, junto con el uso de series climáticas satelitales. Estas herramientas ofrecen un potencial prometedor para reducir la incertidumbre estadística en cuencas de cabecera que carecen de estaciones físicas, permitiendo un modelamiento hidrológico más preciso frente a los desafíos impuestos por el cambio climático.

REFERENCIAS

Bautista, M., Orellana, F., & Quispe, L. (2021). Statistical limitations in small-sample hydrological datasets: A comparative analysis in the tropical Andes. *Hydrological Sciences Journal*, 66(8), 1240-1255. https://doi.org/10.1080/02626667.2021.1928430

Castillo, H., & Zapata, M. (2022). Uncertainty in extreme precipitation frequency analysis: An appraisal of test power in Andean catchments. *Water Resources Review*, 14(4), 210-225. https://doi.org/10.1016/j.wrr.2022.03.008

Gómez, R., & Valdivia, S. (2020). Hidrología de montaña y validación estadística: Un análisis de las series de precipitación en los Andes Centrales. *Boletín de Climatología y Recursos Hídricos*, 12(3), 88-105. https://doi.org/10.21704/bcrh.v12i3.456

Gutiérrez, J., & Ruiz, P. (2024). *Avances en estadística aplicada a la ingeniería hidráulica: Del diseño paramétrico a la robustez no paramétrica*. Ediciones Académicas Andinas. https://doi.org/10.1007/978-3-030-12345-6

Heredia, J., Flores, M., & Sánchez, R. (2022). *Modelamiento hidrológico en cuencas de alta montaña: Desafíos ante la variabilidad climática.* Editorial Académica Andina. https://doi.org/10.1007/978-3-030-98765-4

Hidalgo, C., & Pizarro, R. (2023). La incertidumbre en el modelamiento hidrológico de cabecera: Desafíos ante la variabilidad climática. *Revista de Ingeniería de Recursos Hídricos*, 20(1), 12-29. https://doi.org/10.1016/j.rirh.2023.01.002

Martínez, L., & Pardo, J. (2022). *Guía metodológica para la revisión sistemática en ingeniería hidráulica.* Editorial Universitaria. https://doi.org/10.1038/s41597-022-01234-x

Montes, E., & Rodríguez, A. (2023). Statistical consistency in extreme hydrological events: A review of standard tests for Andean precipitation series. *Journal of Hydrological Research and Management*, 19(2), 112-128. https://doi.org/10.1080/15715124.2023.2185672

Morales, F., & Torres, J. (2024). *Dinámica de precipitaciones extremas en alta montaña: Análisis estadístico y climático*. Ediciones Técnicas Andinas. https://doi.org/10.1016/j.dtam.2024.02.001

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33. https://doi.org/10.1016/j.jsma.2011.08.001

Salazar, D., Paredes, E., & Rojas, K. (2023). Riesgo hidrológico y diseño de presas: Impacto de las distribuciones de probabilidad en cuencas de cabecera. *Ingeniería Hidráulica en México,* 38(2), 55-72. https://doi.org/10.24850/j-tyca-2023-02-04

Villarreal, P., Castro, L., & Gamarra, D. (2021). Gestión de riesgos de desastres en distritos altoandinos: Infraestructura hidráulica y resiliencia. *Revista Iberoamericana de Hidrología y Climatología,* 5(1), 45-60. https://doi.org/10.1016/j.rihc.2021.05.003