UNIVERSIDAD NACIONAL AGRARIA LA MOLINA - ESCUELA DE POSGRADO

DOCTORADO EN INGENIERÍA Y CIENCIAS AMBIENTALES



DISEÑO Y ANÁLISIS DE EXPERIMENTOS EN INGENIERÍA Y CIENCIAS AMBIENTALES

**Actividad:** Trabajo final encargado de teoría

Docente:

Ph.D. Christian René Encina Zelada

Presenta:

José Augusto Zevallos Ruiz

Lima – Perú

10 de diciembre del 2024

# INTRODUCCIÓN

La región de Piura, en la costa norte de Perú, enfrenta desafíos significativos asociados a eventos de precipitaciones extremas vinculados al fenómeno El Niño. Estudios previos han demostrado una relación consistente entre las precipitaciones intensas y las anomalías positivas en la temperatura de la superficie del mar (SST), que favorecen la formación de sistemas convectivos al modificar las condiciones en la capa límite planetaria (Tapley & Waylen, 1990; Waylen & Caviedes, 1986). Durante estos eventos, las precipitaciones acumuladas en Piura pueden ser hasta treinta veces superiores a lo normal, resultado de incursiones anómalas hacia el polo de la Corriente del Niño que generan inestabilidad atmosférica extendida y condiciones propicias para el ascenso del aire, especialmente sobre pendientes montañosas cercanas (Takahashi, 2004). Estos fenómenos meteorológicos incrementan significativamente el riesgo de inundaciones, lo que subraya la necesidad de herramientas predictivas robustas para la gestión del riesgo.

En el contexto del cambio climático, el aumento en la frecuencia e intensidad de eventos extremos refuerza la urgencia de desarrollar modelos predictivos eficaces que apoyen la planificación y la mitigación de riesgos. Tradicionalmente, se han empleado modelos hidráulicos basados en las ecuaciones de aguas poco profundas, los cuales han sido aplicados para simular inundaciones históricas en Piura, como las asociadas al evento El Niño de 2017, y proyectar escenarios futuros de períodos de retorno hasta 500 años (Álvarez et al., 2021; Muñoz et al., 2022). Aunque útiles, estas metodologías tienden a enfocarse en evaluaciones posteriores a los eventos, dejando aspectos como el llenado de datos faltantes sin explorar completamente.

Por otro lado, los modelos hidrológicos han mostrado avances en la predicción a corto plazo gracias a la integración de datos satelitales, mejorando la precisión en eventos de inundaciones en regiones montañosas como los Andes tropicales (Llauca et al., 2023). Asimismo, sistemas globales como el GloFAS han demostrado ser efectivos para pronosticar inundaciones en cuencas con grandes áreas de captación, alcanzando una precisión del 65% en ríos peruanos (Bischiniotis et al., 2019). Sin embargo, estas herramientas suelen depender de conjuntos de datos completos, lo que representa un desafío en áreas donde las estaciones de monitoreo presentan discontinuidades significativas en sus registros.

En este contexto, las técnicas de regresión múltiple ofrecen una alternativa prometedora para abordar el problema del llenado de datos faltantes, particularmente en zonas con baja densidad de estaciones meteorológicas. La regresión múltiple permite relacionar las precipitaciones en estaciones cercanas mediante la integración de variables predictoras, como la elevación, latitud y longitud, logrando estimaciones precisas a nivel local (Naoum & Tsanis, 2004). Por ejemplo, estudios realizados en la isla de Creta han demostrado cómo la incorporación de estas variables mejora significativamente las predicciones de precipitación en terrenos complejos y con limitadas estaciones de monitoreo.

Además, la aplicación de modelos de regresión múltiple ha sido exitosa en diferentes regiones, como en la isla de Creta, donde se lograron estimaciones robustas de la precipitación media anual al incluir parámetros orográficos y geográficos (Naoum & Tsanis, 2004). Estas metodologías no solo capturan la variabilidad espacial de las precipitaciones, sino que también son aplicables en diferentes escalas, desde cuencas hasta áreas más extensas, lo que resalta su versatilidad.

El presente estudio tiene como objetivo principal utilizar modelos de regresión múltiple para llenar vacíos en datos de precipitación diaria en la cuenca del río Piura. Para ello, se utilizarán datos de cinco estaciones meteorológicas: Chusis, Chalaco, Huamarca, Huancabamba y Miraflores, aprovechando la información de las estaciones vecinas para realizar las estimaciones. Este enfoque busca mejorar la disponibilidad de datos en la región, lo que es esencial para la modelación hidrológica y la planificación de recursos hídricos.

La metodología planteada permitirá no solo abordar la problemática de datos faltantes, sino también generar insumos que puedan integrarse a modelos hidrológicos e hidráulicos para la gestión del riesgo de inundaciones en Piura. Así, se busca contribuir al fortalecimiento de herramientas de predicción, esenciales para mitigar los impactos de los eventos extremos que afectan recurrentemente a esta región.

# MARCO TEÓRICO

**Regresión lineal múltiple**

La regresión múltiple es una técnica estadística utilizada para modelar la relación entre una variable dependiente y varias variables independientes . Su objetivo es predecir o explicar su variabilidad basándose en las variables predictoras. La ecuación general de un modelo de regresión múltiple es:

Donde:

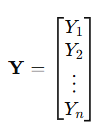
* : Variable dependiente.
* ​: Variables independientes o predictoras.
* ​: Intercepto del modelo.
* Coeficientes de regresión que representan el cambio esperado en por unidad de cambio en cada ​, manteniendo las demás constantes.
* : Término de error o residual, que captura la variabilidad no explicada por el modelo.

**Forma Matricial del Modelo**

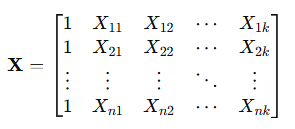
El modelo de regresión múltiple puede expresarse en términos matriciales como:

Donde:

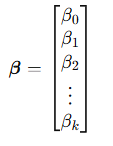
* es el vector de observaciones de la variable independiente .



* es la matriz de diseño , que incluye una columna de unos para el termino de intercepto y las observaciones de las variables independientes



* es el vector de coeficientes :



* es el vector de errores :

**Estimación de los Coeficientes**

Los coeficientes del modelo se estiman minimizando la suma de los cuadrados de los errores , siguiendo el método de los mínimos cuadrados ordinarios (OLS). La estimación matricial de se obtiene como:

Donde:

* es la transpuesta de la matriz de diseño.
* es la matriz inversa del producto.
* es el producto de la transpuesta de con .

El vector contiene los coeficientes estimados del modelo

**Prueba Kolmogórov-Smirnov**

La prueba de Kolmogórov-Smirnov es una prueba no paramétrica que evalúa si una muestra sigue una distribución específica. Es utilizada para verificar la normalidad de los datos en un análisis ANOVA. La estadística de prueba DDD para esta prueba es:

Donde:

* es la función de distribución empírica de la muestra,
* es la función de distribución acumulada teórica.

Si el valor de D es suficientemente grande, se rechaza la hipótesis nula de que los datos siguen la distribución especificada.

**Prueba de Durbin-Watson**

La prueba de Durbin-Watson es una prueba estadística utilizada para detectar la autocorrelación en los residuos de un modelo de regresión. Su estadística de prueba es:

Donde:

* son los residuos,
* es el número de observaciones

Un valor cercano a 2 indica que no hay autocorrelación en los residuos, mientras que valores alejados de 2 sugieren la presencia de autocorrelación.

**Error cuadrático medio**

Es un estimador que mide el promedio de los errores al cuadrado. El error es la diferencia entre el valor estimado por un modelo y el valor real medido de la variable. Esta diferencia se debe a que el modelo generalmente no captura toda la información necesaria para reproducir la realidad o a la existencia de errores de medición aleatoria (ver Ecuación).

# METODOLOGÍA

**4.1. Descripción de las bases de la información analizada**

**4.2. Metodología empleada**

Figura 1 Flujograma metodológico empleado en el presente trabajo

**4.3. Datos**

# RESULTADOS Y DISCUSIONES

**Prueba Kolmogórov-Smirnov**

Para el test de normalidad se tiene = distribución normal muestral, = distribución teórica Weibull.

De la siguiente expresión se obtuvo que:

El valor de :

Como el valor de no supera al , se concluye que los residuos si cumplen el criterio de normalidad.

**Prueba de Durbin-Watson**

La prueba de Durbin-Watson es una prueba estadística utilizada para detectar la autocorrelación en los residuos de un modelo de regresión. En este caso, el valor de Durbin-Watson es 1.0659, con un p-valor de 0.0001056. Dado que el p-valor es muy bajo (menor que el umbral común de significancia de 0.05), se rechaza la hipótesis nula de que no hay autocorrelación. El valor de 1.0659 está cerca de 1, lo que sugiere una autocorrelación positiva de primer orden en los residuos, es decir, existe una relación entre los errores residuales consecutivos en el modelo. Esto indica que los residuos no son independientes, lo cual puede ser un problema para la validez de los resultados del análisis ANOVA, ya que una de las suposiciones clave es la independencia de los errores.

# CONCLUSIONES Y RECOMENDACIONES

# REFERENCIAS BIBLIOGRÁFICAS

Luo, J., Ledgard, S. F., De Klein, C. A. M., Lindsey, S. B., & Kear, M. (2008). Effects of dairy farming intensification on nitrous oxide emissions. *Plant and Soil*, *309*(1–2), 227–237. https://doi.org/10.1007/s11104-007-9444-9

Rui, Y., Wang, Y., Chen, C., Zhou, X., Wang, S., Xu, Z., Duan, J., Kang, X., Lu, S., & Luo, C. (2012). Warming and grazing increase mineralization of organic P in an alpine meadow ecosystem of Qinghai-Tibet Plateau, China. *Plant and Soil*, *357*(1), 73–87. https://doi.org/10.1007/s11104-012-1132-8

Zhan, W., Yang, Z., Liu, J., Chen, H., Yang, G., Zhu, E., Hu, J., Jiang, L., Liu, L., Zhu, D., He, Y., Zhao, C., Xue, D., & Peng, C. (2021). Effect of grazing intensities on soil n2 o emissions from an alpine meadow of zoige plateau in China. *Atmosphere*, *12*(5). https://doi.org/10.3390/atmos12050541

# ANEXOS

Cálculo de los valores de para tratamiento y bloque.

# Datos necesarios Tratamiento

alpha **<-** 0.05 # Nivel de significancia

a **<-** 4 # Número de tratamientos

df\_error **<-** 30 # Grados de libertad asociados al error (ν)

# Obtención del valor crítico de q

q\_critico **<-** qtukey**(**p **=** 1 **-** alpha, nmeans **=** a, df **=** df\_error**)**

# Mostrar el valor crítico de q

q\_critico

# Datos necesarios Bloque

alpha **<-** 0.05 # Nivel de significancia

a **<-** 3 # Número de bloques

df\_error **<-** 30 # Grados de libertad asociados al error (ν)

# Obtención del valor crítico de q

q\_critico **<-** qtukey**(**p **=** 1 **-** alpha, nmeans **=** a, df **=** df\_error**)**

# Mostrar el valor crítico de q

q\_critico