**Título de la publicación:** "Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis"  
**Año de publicación:** 2003  
**Autores:** Jorge Marquínez, Javier Lastra, Pilar García

El estudio desarrolla modelos para estimar la distribución espacial de la precipitación media mensual en una región montañosa del norte de España, utilizando técnicas de regresión lineal múltiple y sistemas de información geográfica (GIS). Se analizaron datos de precipitación de 117 estaciones meteorológicas durante el periodo 1966-1990, dividiendo los datos en conjuntos para el desarrollo del modelo (84 estaciones) y para la validación (33 estaciones).

Se seleccionaron cinco variables topográficas como predictores: altitud, distancia al litoral, distancia al oeste, pendiente y promedio de elevación en áreas homogéneas. Estas variables fueron derivadas de un modelo digital de elevación (DEM) con resolución de 200 m, y se evaluó su influencia en la distribución de la precipitación en diferentes periodos (seco, húmedo y anual).

La regresión permitió explicar entre el 58% y el 67% de la variabilidad espacial de la precipitación media anual, dependiendo del modelo y de las variables seleccionadas. Se observó un error estándar del 10% aproximadamente y un error absoluto medio que varió entre 8.1 y 26.1 mm, representando entre el 13% y el 19% de la precipitación observada.

El modelo desarrollado, denominado "Modelo 4", mostró el mejor ajuste. Incorporó variables derivadas de subcuencas, como la elevación media de la subcuenca (DEMb) y la pendiente media (slopeb), además de variables relacionadas con las posiciones de los frentes dominantes, como la distancia al oeste (dwest) y la distancia al litoral (dsea).

Para validar el modelo, se utilizaron métodos estadísticos como el coeficiente de determinación ajustado (adj-R²), que compensó la tendencia del R² a sobreestimar el ajuste, y análisis de residuales. Los residuales mostraron distribución normal según el test de Kolmogorov-Smirnov y no presentaron autocorrelación significativa según el test de Durbin-Watson.

Las relaciones entre las variables independientes y la precipitación sugieren patrones no lineales, con modelos cuadráticos y cúbicos mejorando significativamente el ajuste. Por ejemplo, la precipitación aumentó con la altitud hasta los 400 m, se estabilizó, y luego volvió a aumentar por encima de los 1000 m.

Los errores de validación mostraron un sesgo medio de ±2.4 mm para el periodo anual y ±1.2 mm para los periodos seco y húmedo. Los errores fueron similares o inferiores a los reportados en estudios previos que aplicaron enfoques de regresión o geoestadísticos en regiones montañosas.

El análisis de regresión residual no mostró tendencias sistemáticas en los residuales respecto a las predicciones, indicando la robustez del modelo. El modelo permitió estimar la precipitación en áreas sin estaciones cercanas, dependiendo de las características topográficas locales y no únicamente de las estaciones más próximas.

Los resultados destacan la importancia de considerar variables derivadas de subcuencas para mejorar la precisión en la predicción de precipitación, especialmente en regiones montañosas donde la orografía juega un papel crítico en los patrones de precipitación.

En conclusión, el enfoque propuesto representa una mejora significativa respecto a métodos de interpolación tradicionales, al no depender de la densidad de estaciones meteorológicas. Esto facilita la extensión del área de estimación más allá de las estaciones existentes, siempre que las características topográficas sean comparables a las utilizadas en el ajuste del modelo.

**Título de la publicación:** "Comparison of multiple linear regression and artificial neural network models for downscaling TRMM precipitation products using MODIS data"  
**Año de publicación:** 2016  
**Autores:** D.D. Alexakis, I.K. Tsanis

El estudio compara dos enfoques metodológicos, regresión lineal múltiple (MLR) y redes neuronales artificiales (ANN), para mejorar la resolución espacial de los datos de precipitación del producto TRMM 3B42 de 25 km a 1 km. Se incorporaron datos ambientales de sensores MODIS, incluyendo NDVI, albedo, índice de sequía y topografía, para el análisis en la isla de Creta, Grecia.

La metodología consistió en vincular estadísticamente las precipitaciones medidas por TRMM con parámetros ambientales derivados de MODIS. Además, los resultados se validaron con datos independientes de 20 estaciones de pluviómetros distribuidas en la región, excluyendo áreas urbanas y cuerpos de agua que no mostraban dependencia de precipitación.

La regresión lineal múltiple (MLR) se basó en un modelo de mínimos cuadrados para minimizar las desviaciones entre valores observados y predichos. Por otro lado, las redes neuronales artificiales (ANN) emplearon un modelo de perceptrón multicapa con algoritmos de retropropagación para identificar relaciones no lineales entre los datos.

Los indicadores de desempeño utilizados para evaluar los modelos incluyeron el error cuadrático medio (RMSE) y el sesgo (Bias). La validación mostró que la MLR tuvo mejor rendimiento en Creta occidental, mientras que las ANN fueron superiores en la región oriental, debido a las diferencias en la importancia del parámetro de albedo.

La MLR presentó valores de R² de 0.46 y 0.47 para Creta occidental y oriental, respectivamente. El modelo mostró una relación significativa entre la precipitación TRMM y los parámetros de albedo y NDVI, especialmente en la región oriental. Sin embargo, no se encontraron correlaciones significativas con la altitud.

Las redes neuronales demostraron mayor capacidad para manejar datos complejos y no lineales, logrando valores promedio de R² de 0.65 y 0.55 en Creta oriental y occidental, respectivamente. La importancia relativa de los predictores varió entre regiones, siendo el albedo el parámetro más influyente.

El análisis mostró que ambos métodos mejoraron significativamente la resolución espacial de los datos de precipitación. Sin embargo, las ANN demostraron una mayor capacidad para capturar la variabilidad de precipitación en regiones con valores extremos de albedo, como la desertificación en Creta oriental.

Se aplicaron pruebas estadísticas como la correlación de Pearson para evaluar las relaciones iniciales entre las precipitaciones TRMM y los datos ambientales. Además, la interpolación spline se utilizó para mapear las precipitaciones residuales, optimizando así la precisión del modelo.

En términos de validación, el RMSE de las ANN osciló entre 11 y 182 mm dependiendo del mes y la región, mientras que para la MLR fue ligeramente inferior en algunas estaciones. El sesgo mostró una tendencia general de subestimación en las ANN y valores mixtos en la MLR.

En conclusión, ambas metodologías ofrecen ventajas complementarias para mejorar la resolución de los datos de precipitación. Las ANN son más adecuadas para áreas con alta heterogeneidad ambiental, mientras que la MLR es más eficiente en regiones con menor variabilidad de parámetros. Estos resultados resaltan el potencial de datos satelitales gratuitos para estudios hidrológicos y planificación de recursos hídricos.

**Título de la publicación:** "Filling Gaps in Daily Precipitation Series Using Regression and Machine Learning in Inter-Andean Watersheds"  
**Año de publicación:** 2022  
**Autores:** Marcelo Portuguez-Maurtua, José Luis Arumi, Octavio Lagos, Alejandra Stehr, Nestor Montalvo Arquiñigo

Este estudio abordó el problema de completar series diarias de precipitación en las cuencas de los ríos Mala, Omas y Cañete en la vertiente del Pacífico de Perú. Se evaluaron métodos de regresión lineal (LRM y MRM) y técnicas de aprendizaje automático (ML), incluyendo K-vecinos más cercanos (KNN), árboles de decisión mejorados (GBT) y bosques aleatorios (RF), optimizados mediante Bayes.

La metodología siguió cuatro etapas principales: recolección de datos de precipitación diaria, análisis exploratorio y homogenización, regionalización, y aplicación de modelos para rellenar datos faltantes. Se implementaron pruebas de control de calidad, como el test de homogeneidad normal estándar (SNHT), para identificar inconsistencias en las series.

La regionalización utilizó métodos como Ward, K-means y el análisis vectorial regional para identificar zonas homogéneas de precipitación. Se determinaron tres regiones principales con patrones de variabilidad interanual similares, basándose en correlaciones y desviaciones estándar de las estaciones de monitoreo.

Los métodos de regresión lineal mostraron desempeño variable. El modelo MRM presentó coeficientes de determinación (R²) de hasta 0.49, mientras que los métodos ML, particularmente RF y GBT, superaron ampliamente los resultados, logrando R² de hasta 0.89 para los datos de entrenamiento.

El proceso de optimización basado en métodos bayesianos permitió mejorar el rendimiento de los modelos ML. Por ejemplo, los modelos optimizados (OML) presentaron errores cuadrados medios (RMSE) más bajos y valores más altos del coeficiente de Nash-Sutcliffe (NSE), con un mejor equilibrio entre sesgo y precisión.

En el análisis de validación, los modelos OML-GBT y OML-RF destacaron por su capacidad para capturar la variabilidad no lineal de los datos. Esto fue evidente en estaciones como Ayaviri, donde NSE alcanzó valores cercanos a 0.88 para los datos de entrenamiento y 0.71 para los datos de prueba.

Los resultados de las pruebas estadísticas incluyeron indicadores clave como R², RMSE, NSE y PBIAS. Los modelos ML-RF y OML-RF presentaron el mejor desempeño general, con bajos sesgos de predicción y altos niveles de correlación entre valores observados y predichos.

El análisis de outliers y valores faltantes identificó estaciones con más del 10% de datos incompletos, como Cañete, Socsi y Pacarán, que afectaron la homogenización y fueron excluidas de algunos análisis. Las restantes estaciones mostraron datos de alta calidad tras la homogenización.

Este trabajo demuestra la superioridad de los métodos ML sobre la regresión lineal tradicional en contextos de alta variabilidad temporal y espacial, como las cuencas interandinas. Los modelos optimizados fueron particularmente eficaces en regiones con condiciones hidroclimáticas secas.

En conclusión, el uso de ML, combinado con técnicas de regionalización y optimización, ofrece una solución robusta para llenar lagunas en series diarias de precipitación. Esto tiene implicaciones significativas para la modelación hidrológica y la gestión de recursos hídricos en regiones montañosas con datos escasos o incompletos.

**Título de la publicación:** "Filling Gaps in Daily Precipitation Series Using Regression and Machine Learning in Inter-Andean Watersheds"  
**Año de publicación:** 2022  
**Autores:** Marcelo Portuguez-Maurtua, José Luis Arumi, Octavio Lagos, Alejandra Stehr, Nestor Montalvo Arquiñigo

Este estudio abordó el problema de completar series diarias de precipitación en las cuencas de los ríos Mala, Omas y Cañete en la vertiente del Pacífico de Perú. Se evaluaron métodos de regresión lineal (LRM y MRM) y técnicas de aprendizaje automático (ML), incluyendo K-vecinos más cercanos (KNN), árboles de decisión mejorados (GBT) y bosques aleatorios (RF), optimizados mediante Bayes.

La metodología siguió cuatro etapas principales: recolección de datos de precipitación diaria, análisis exploratorio y homogenización, regionalización, y aplicación de modelos para rellenar datos faltantes. Se implementaron pruebas de control de calidad, como el test de homogeneidad normal estándar (SNHT), para identificar inconsistencias en las series.

La regionalización utilizó métodos como Ward, K-means y el análisis vectorial regional para identificar zonas homogéneas de precipitación. Se determinaron tres regiones principales con patrones de variabilidad interanual similares, basándose en correlaciones y desviaciones estándar de las estaciones de monitoreo.

Los métodos de regresión lineal mostraron desempeño variable. El modelo MRM presentó coeficientes de determinación (R²) de hasta 0.49, mientras que los métodos ML, particularmente RF y GBT, superaron ampliamente los resultados, logrando R² de hasta 0.89 para los datos de entrenamiento.

El proceso de optimización basado en métodos bayesianos permitió mejorar el rendimiento de los modelos ML. Por ejemplo, los modelos optimizados (OML) presentaron errores cuadrados medios (RMSE) más bajos y valores más altos del coeficiente de Nash-Sutcliffe (NSE), con un mejor equilibrio entre sesgo y precisión.

En el análisis de validación, los modelos OML-GBT y OML-RF destacaron por su capacidad para capturar la variabilidad no lineal de los datos. Esto fue evidente en estaciones como Ayaviri, donde NSE alcanzó valores cercanos a 0.88 para los datos de entrenamiento y 0.71 para los datos de prueba.

Los resultados de las pruebas estadísticas incluyeron indicadores clave como R², RMSE, NSE y PBIAS. Los modelos ML-RF y OML-RF presentaron el mejor desempeño general, con bajos sesgos de predicción y altos niveles de correlación entre valores observados y predichos.

El análisis de outliers y valores faltantes identificó estaciones con más del 10% de datos incompletos, como Cañete, Socsi y Pacarán, que afectaron la homogenización y fueron excluidas de algunos análisis. Las restantes estaciones mostraron datos de alta calidad tras la homogenización.

Este trabajo demuestra la superioridad de los métodos ML sobre la regresión lineal tradicional en contextos de alta variabilidad temporal y espacial, como las cuencas interandinas. Los modelos optimizados fueron particularmente eficaces en regiones con condiciones hidroclimáticas secas.

En conclusión, el uso de ML, combinado con técnicas de regionalización y optimización, ofrece una solución robusta para llenar lagunas en series diarias de precipitación. Esto tiene implicaciones significativas para la modelación hidrológica y la gestión de recursos hídricos en regiones montañosas con datos escasos o incompletos.

**Título de la publicación:** "Artificial Neural Networks and Multiple Linear Regression for Filling in Missing Daily Rainfall Data"  
**Año de publicación:** 2022  
**Autores:** Ioannis Papailiou, Fotios Spyropoulos, Ioannis Trichakis, George P. Karatzas

Este estudio aborda la imputación de datos diarios de lluvia faltantes en estaciones meteorológicas, utilizando redes neuronales artificiales (ANN) y regresión lineal múltiple (MLR). Se trabajó con datos de cinco estaciones en la región de Chania, Grecia, aplicando ambas metodologías para completar series temporales incompletas.

La metodología consistió en dividir los datos disponibles en conjuntos completos e incompletos. Los datos completos se utilizaron para entrenar y validar las ANN y el modelo MLR, mientras que los datos incompletos se imputaron utilizando las simulaciones generadas.

Se crearon ensambles de ANN con una capa oculta para modelar cada combinación de datos faltantes. En total, se entrenaron 10,000 redes para cada caso, seleccionando la mejor red según el error de prueba. Las ANN fueron implementadas en MATLAB, optimizando su estructura mediante un algoritmo competitivo.

Los indicadores de desempeño utilizados para evaluar los modelos incluyeron el error cuadrático medio (RMSE), el coeficiente de Nash-Sutcliffe (NSE) y el coeficiente de correlación (R). Las ANN superaron consistentemente al modelo MLR en todas las métricas, destacando especialmente en casos con múltiples datos faltantes.

Los resultados mostraron que el RMSE de las ANN varió entre 1.16 mm y 2.42 mm, mientras que para el modelo MLR osciló entre 2.37 mm y 6.43 mm. El mejor desempeño de las ANN se observó en casos donde solo una estación tenía datos faltantes.

El NSE para las ANN se mantuvo por encima de 0.91 en todos los casos, alcanzando un valor máximo de 0.989. En comparación, el NSE del modelo MLR presentó valores menores, cayendo por debajo de 0.71 en los casos más complejos.

El coeficiente de correlación (R) también favoreció a las ANN, con valores que oscilaron entre 0.94 y 0.99, mientras que el modelo MLR alcanzó un máximo de 0.93. Esto evidencia la capacidad superior de las ANN para capturar patrones complejos de precipitación.

Aunque las ANN mostraron un desempeño superior, su tiempo de entrenamiento fue significativamente mayor, alcanzando 36 horas en una computadora de especificaciones medias, mientras que el modelo MLR se ejecutó en solo minutos. Esto sugiere una compensación entre precisión y rapidez.

Los resultados destacan que las ANN son particularmente eficaces en casos de eventos extremos de lluvia, como los registrados en 2006 y 2019, donde los valores simulados por las ANN reflejaron mejor los patrones observados en las estaciones disponibles.

En conclusión, las ANN ofrecen una solución más precisa para imputar datos faltantes en series diarias de lluvia, especialmente en escenarios complejos. Sin embargo, el MLR sigue siendo una opción válida cuando el tiempo de procesamiento es una restricción. Este estudio proporciona una base para futuras investigaciones que integren enfoques híbridos o nuevos algoritmos de optimización.

**Título de la publicación:** "Orographic Precipitation Modeling with Multiple Linear Regression"  
**Año de publicación:** 2004  
**Autores:** S. Naoum, I. K. Tsanis

Este estudio desarrolla un modelo de regresión lineal múltiple (MLR) para estimar la precipitación anual en la isla de Creta, Grecia. La precipitación se relacionó con variables predictoras como la elevación, la latitud y la longitud, utilizando tecnología de sistemas de información geográfica (GIS). El enfoque permite explorar la influencia de los factores orográficos en diferentes escalas espaciales, desde la isla completa hasta subcuencas específicas.

Se recopilaron datos de 74 estaciones de precipitación ubicadas a diferentes elevaciones, con registros de 12 a 50 años. La precipitación anual promedio en años secos, normales y húmedos fue estimada en 700 ± 100 mm, 950 ± 150 mm, y 1,300 ± 200 mm, respectivamente. Se analizaron gradientes precipitación-elevación en rangos de 0.45 a 1.3 mm/m dependiendo del año.

El modelo utilizó técnicas avanzadas de regresión, incluidas pruebas estadísticas de significancia y análisis de residuales para validar la precisión de las predicciones. Los coeficientes de determinación (R²) para modelos basados en elevación alcanzaron hasta el 46% en años húmedos, aumentando significativamente al incluir variables como latitud y longitud.

Se aplicó el análisis de sumas de cuadrados extra para evaluar la importancia de términos adicionales en los modelos. Aunque se mejoró el ajuste al incluir términos de segundo orden e interacciones, esto resultó en pérdida de grados de libertad, especialmente en áreas con baja densidad de estaciones de monitoreo.

Los resultados mostraron que dividir la isla en regiones (norte, sur y este) mejoró la precisión de los modelos. Por ejemplo, el norte, que recibe el 45% de la precipitación total, presentó un aumento del 34% en R² en el modelo regional respecto al modelo general para la isla completa.

El análisis espacial reveló una clara influencia de la orografía en los patrones de precipitación, con mayores acumulaciones en el norte debido a la dirección predominante del viento noroeste. Las estaciones ubicadas en zonas de sombra de lluvia mostraron menores precipitaciones, destacando la importancia de factores topográficos.

Los modelos basados en elevación exclusivamente subestimaron la precipitación en áreas de alta complejidad topográfica. Al incluir latitud y longitud, los modelos lograron capturar mejor la variabilidad espacial, especialmente en años extremos como 1977-1978 (húmedo) y 1989-1990 (seco).

El uso de DEM (modelos digitales de elevación) permitió representar los efectos orográficos de manera más precisa, aunque se identificaron limitaciones relacionadas con la resolución de los datos y la representatividad de las estaciones en altitudes superiores a 905 m.

Los modelos desarrollados mostraron un balance entre ajuste y simplicidad, siendo útiles tanto para descripciones generales como para predicciones específicas. Sin embargo, en áreas como el este de Creta, la baja densidad de estaciones afectó la confiabilidad de los modelos más complejos.

En conclusión, el enfoque basado en MLR y GIS ofrece una herramienta robusta para modelar la precipitación orográfica, proporcionando estimaciones precisas en terrenos complejos. Los resultados destacan la importancia de considerar múltiples variables y escalas espaciales, lo que tiene implicaciones significativas para la gestión de recursos hídricos en áreas montañosas.