



## Reconstrucción de datos faltantes de precipitación pluvial diaria mediante la Transformada Wavelet

### Daily rainfall data-gap filling using a Wavelet Transform-based methodology

MARIELLA CARBAJAL CARRASCO<sup>1,2</sup>, CHRISTIAN YARLEQUÉ<sup>1</sup>, ADOLFO POSADAS<sup>1,3</sup>, ELIZABETH SILVESTRE<sup>4</sup>, ABEL MEJÍA<sup>2</sup> & ROBERTO QUIROZ<sup>1,\*</sup>

<sup>1</sup>División de Sistemas de Producción y Ambiente, Centro Internacional de la Papa, Av. La Molina 1895, Lima 12, Perú

<sup>2</sup>Universidad Nacional Agraria La Molina, MRH-EPG-UNALM, Av. La Molina s/n, La Molina, Lima 12, Perú

<sup>3</sup>EMBRAPA, Rua XV de Novembro, 1452, Sao Carlos, SP, CEP 13560-970, Brasil

<sup>4</sup>Servicio Nacional de Meteorología e Hidrología, SENAMHI, Castilla 1308, Lima 11, Perú.

\*r.quiroz@cgiar.org

#### Resumen

Se describe la aplicación de un nuevo método para la completación de datos faltantes de precipitación diaria utilizando la "Transformada de Wavelets" (TW). Series completas de datos de precipitación diaria para los periodos 1999-2000, 2003-2004, 2004-2005, seis estaciones meteorológicas del altiplano peruano fueron utilizadas en el análisis. El proceso consistió en la descomposición y reconstrucción de la señal, a dos niveles, usando el Wavelet Haar. El programa desarrollado utiliza la señal que resulta del filtro de baja frecuencia, proveniente de la estación a ser corregida y el detalle o señal de alta frecuencia proveniente de una estación vecina, la que proporciona el detalle aleatorio que se usa en el llenado de datos faltantes. La selección de la estación vecina se basó en la similitud de la distribución acumulada de ambas señales previamente estandarizadas. El procedimiento se validó generando vacíos aleatorios, llenando éstos con la transformada de Wavelets o el generador de datos climáticos ClimGen y comparando las señales reconstruidas con la señal original sin alteración en todos los casos. Son comprobados en todos los casos, el método basado en la TW produjo mejores coeficientes de determinación que ClimGen. La prueba Kolmogorov-Smirnov constató que las señales reconstruidas y originales son similares. Otros estimadores estadísticos de las distribuciones también fueron similares. El método basado en la TW, parece ser el más apropiado para la completación de datos diarios de precipitación ya que, además de producir datos muy próximos a los reales, la generación de los parámetros para la reconstrucción requiere de pocos datos en comparación con los métodos convencionales.

**Palabras claves:** Precipitación Pluvial; descomposición; reconstrucción; Transformada Wavelet; ClimGen.

#### Abstract

This paper describes the application of a new wavelet transform-based method to fill data gaps in daily rainfall signals. Complete rainfall gauge data from six weather stations located in the high plateau of Peru, spanning the periods: 1999-2000, 2003-2004, 2004-2005 were used for the study. The process entailed the decomposition and reconstruction of the signal at two levels using the wavelet Haar. The software developed utilizes the low-pass signal from the target weather station and the high-pass signal from a neighbor weather station which confer the random element to the reconstruction of a daily signal to fill in the data gaps. The selection of the neighbor weather station was based on the similarity of the standardized cumulative distribution function of both signals. The procedure was validated by generating random data gaps, filling the gaps with the wavelet transformed (WT) method or the standard ClimGen software and comparing the reconstructed signals to the full signal prior to the generation of the data gaps. In all cases, the WT method produced slightly better determination coefficients than the standard ClimGen method. The Kolmogorov-Smirnov test confirmed that the original and reconstructed signals were similar; as other statistical estimators of the time series did too. The WT method proposed in this paper seems to be a more suitable alternative for filling daily rainfall data gaps since it provided a closer approximation to the original data with the plus that it requires much less data than standard methods to estimate the required generating parameters.

**Keywords:** Rainfall; Decomposition; Reconstruction; Wavelet Transform; ClimGen.

## INTRODUCCIÓN

La precipitación pluvial es una variable climática de gran importancia para los sistemas hidrológico, agrícola, industrial y energético. El entendimiento de su comportamiento temporal y espacial es de sumo interés, especialmente en los estudios de riesgos climáticos, donde la disponibilidad de información de alta resolución y de buena calidad es esencial.

En el Perú, la información sobre precipitación pluvial diaria es escasa e incompleta (Giddings & Soto, 2006), lo que limita la capacidad de análisis de riesgos de déficit hídricos en agricultura de secano y a su vez constituye una fuente adicional de error cuando se modelan los balances hidrológicos (Giddings & Soto, 2006).

Este inconveniente nos lleva a buscar alternativas de solución en muchos modelos de la literatura, los cuales según sus autores son capaces de generar, completar o interpolar datos de precipitación y otras variables climáticas, utilizando para ello diferentes métodos como los modelos regresivos univariados y multivariados (Sveinsson et al., 2007), la interpolación con Kriging (Wackernagel, 2003), etc. Cada uno de estos modelos tiene sus limitaciones según el método que aplique y el número de entradas disponibles.

En este contexto, el presente trabajo propone la aplicación de un nuevo método para completar datos medidos de lluvia diaria, en base a la transformada Wavelet (TW). El método de reconstrucción de precipitación diaria con la TW ha sido desarrollado y validado por Quiroz et al. (2010), Yarlequé (2009) y Yarlequé et al. (2007). Las pruebas de reconstrucción de datos climáticos mostradas se realizaron con la información de seis estaciones del Altiplano peruano, localizadas en la provincia de Puno, definiéndose el término "reconstrucción" como el proceso de obtener una nueva señal y completar con

ella datos ausentes en la serie real. La validación cruzada de la técnica propuesta para la reconstrucción de datos faltantes se realizó con las mismas seis estaciones del Altiplano y se comparó con los resultados obtenidos con el generador de datos climáticos ClimGen (Nelson, 2002).

## MATERIALES

### Área de Estudio

El Altiplano peruano está ubicado alrededor del Lago Titicaca, entre las coordenadas 15°13'59'' S; 70°2'15'' O; 68°34'4'' E y 16° 58' 18'' S (Ver Fig. 1). La altitud de esta región varía entre los 3800 y los 4500 metros (m) sobre el nivel del mar. La precipitación pluvial varía de <400 hasta 600 mm por año; en promedio la temperatura mínima desciende hasta 10°C y las sequías pueden durar hasta 150 días (d) por año. El uso predominante de la tierra es de pastizales con áreas cultivadas, principalmente cerca al lago (Quiroz et al., 2003).

La actividad convectiva y la precipitación en el Altiplano ocurren casi exclusivamente durante el verano austral (diciembre-marzo) y están asociadas con la expansión estacional de los vientos del Este superior y relacionadas con la afluencia de la humedad cercana a la superficie continental de las tierras bajas del Este (Vuille & Keiming, 2004).

La variación de precipitación está también relacionada con la proximidad al Lago Titicaca encontrándose los valores más altos cerca a éste. Los estudios corroboran que más lluvia cae sobre el lago Titicaca que sobre cualquier otro lugar del Altiplano. Durante la época de lluvias el flujo atmosférico general es del sureste y las áreas viento arriba del lago Titicaca (sureste del lago) tienden a ser más secas que las áreas viento abajo (Grace, 1985).

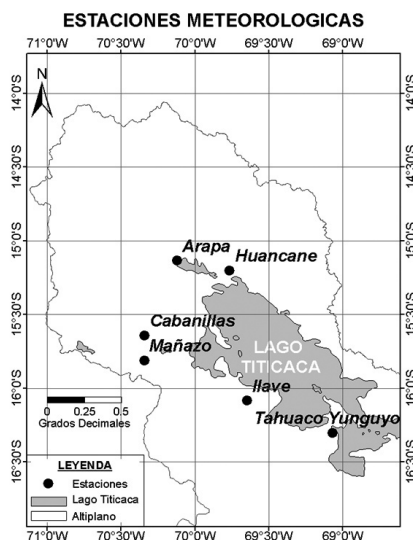


Fig. 1. Estaciones meteorológicas empleadas en el Estudio.

**Data Meteorológica**

La data pluviométrica diaria de seis estaciones meteorológicas fue obtenida del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). La ubicación geográfica de las estaciones climáticas se muestra en la Fig. 1, y los periodos de análisis de los datos observados se detallan en la Tabla 1. La data fue pre procesada (análisis no mostrado) para verificar su consistencia mediante la técnica de doble masa (Unesco/Rostlac, 1982; Dunne & Leopold, 1978).

Cabe mencionar que para la reconstrucción o completación de datos con el método de la Transformada Wavelet el número de datos utilizado fue el mismo que el periodo de análisis. Por otro lado, con el generador ClimGen se utilizó un registro de 27 a 42 años dependiendo de la data disponible de cada estación para la

parametrización del generador. Se procedió luego a generar la data de toda la serie de tiempo. Finalmente se extrajo la data del período de análisis (Tabla 1) para su posterior comparación con los resultados obtenidos con el método de la Transformada Wavelet.

**Transformada Wavelet (TW)**

La TW es una técnica matemática introducida en el análisis de señales en los primeros años de la década de 1980 (Mallat 1999; Prasad & Iyengar, 1997). La TW se encuentra definida en un dominio de tiempo y escala (frecuencia), y posee un soporte compacto (Foufoula-Georgiou & Kumar, 1994). Esto permite estudiar series de tiempo con una resolución definida por la escala; esto es, estructuras generales en la señal en una escala grande (baja resolución) y finas estructuras en escalas pequeñas (alta resolución). Esta característica es especialmente útil para

Tabla 1. Ubicación y período de análisis de las estaciones meteorológicas en el Altiplano peruano.

Nº	Estación	Latitud S	Longitud O	Altitud (msnm)	Periodo de análisis
1	Cabanillas	15°38'38"	70°20'19"	3890	01-12-1999 /02-02-2000 (64 días)
2	Mañazo	15°48'48"	70°20'20"	3920	
3	Ilave	16°05'05"	69°38'37"	3850	11-11-2003 /17-03-2004 (128 días)
4	Tahuaco	16°18'18"	69°04'03"	3860	
5	Arapa	15°08'08"	70°07'07"	3920	11-11-2004/ 17-03-2005 (128 días)
6	Huancané	15°12'12"	69°45'45"	3880	

variaciones espaciales que son no-estacionarias, y que poseen componentes que trascienden con una corta duración, con estructuras en diferentes escalas, o que poseen singularidades, las cuales son ideales para mostrar pequeñas fluctuaciones en las señales (Posadas et al., 2005). Debido a esta propiedad, el análisis con la TW tiene muchas aplicaciones, en dinámica de fluidos (Farge, 1992; Gao & Li, 1993; Liu, 1994, Katul & Vidakovic, 1996), geofísica o hidrología (Kumar & Foufoula-Georgiou, 1993; Labat et al., 2000).

La TW es una técnica física-matemática (Graps, 1995), que describe la señal en estudio por intermedio de sus coeficientes, los que a su vez describen la diversidad de cambios ocasionados por eventos (singularidades) locales (Kumar & Foufoula, 1997). La TW no es una técnica meramente estadística, que requiera un gran número de datos, sino que ella representa la descomposición de una señal en diferentes niveles de resolución, proceso conocido como multiresolución. El uso de una representación multiresolución ofrece un marco jerárquico simple para interpretar la información contenida en la señal. A diferentes resoluciones, los detalles de una señal generalmente caracterizan diferentes estructuras físicas de ella. A una resolución más gruesa, estos detalles generalmente caracterizan las grandes estructuras que su entorno contiene. Con un aumento de la resolución obtenemos detalles más finos. Así, la TW puede ser discreta o continua:

La TW se define como la convolución de una función  $f$ , con una función de onda (Wavelet) como Kernel de la transformada (Foufoula-Georgiou & Kumar, 1994):

$$T_f(\lambda, t) = \int_{-\infty}^{\infty} f(\mu) \psi_{\lambda, \mu}(\mu) d\mu \quad (1)$$

donde:

$$\psi_{\lambda, \mu}(t) = \frac{1}{\sqrt{\lambda}} \psi\left(\frac{t - \mu}{\lambda}\right) \quad (2)$$

Aquí  $\lambda > 0$  representa la variable de escala (o ancho de la función Wavelet) y  $\mu$  la variable de translación (ubicación de la función Wavelet). La función  $\psi(t)$  es la conocida Wavelet Madre o Wavelet, que es utilizada para que sean óptimamente localizados los valores de la transformada en el tiempo (espacio) y en distintas escalas (frecuencias).

El operador Inverso de la Transforma Wavelet (IWT) es definido como (Foufoula-Georgiou & Kumar, 1994):

$$f(t) = ITW(TW[f(\lambda, t)]) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} \lambda^{-2} TW[f(\lambda, \mu)] \psi_{\lambda, \mu}(t) d\lambda d\mu, \quad (3)$$

$$C_\psi = 2\pi \int_0^{\infty} \frac{|\hat{\psi}(w)|^2}{w} dw < \infty, \quad (4)$$

Para mayor información acerca de las propiedades de la TW y el ITW, el lector puede revisar los trabajos recopilados en Foufoula-Georgiou & Kumar (1994).

Con el fin de construir algoritmos eficientes para calcular la TW, se realiza la discretización del parámetro de escala  $\lambda$  y del parámetro de translación  $t$ , lo que produce la Transformada de Wavelet Discreta (y la inversa en su forma discreta). La selección de  $\lambda = 2m$  y  $t = n2m$ , resulta en una representación discreta ortogonal (Foufoula-Georgiou & Kumar, 1994; Daubechies, 1988).

### Análisis de Multiresolución (AMR) con la TW

El AMR (Mallat, 1989) proporciona un tipo de descomposición y reconstrucción de una señal en base a las propiedades de la TW (e ITW), en su forma discreta. Para una información detallada, los autores recomiendan revisar la siguiente documentación (Mallat, 1999; Foufoula-Georgiou & Kumar, 1994).

En términos generales, seleccionando una cierta función Wavelet Ortogonal (Foufoula-Georgiou & Kumar, 1994), es posible obtener una descomposición de la función  $f$  como:

$$f(t) = A_N(t) + D_N(t) \quad (5)$$

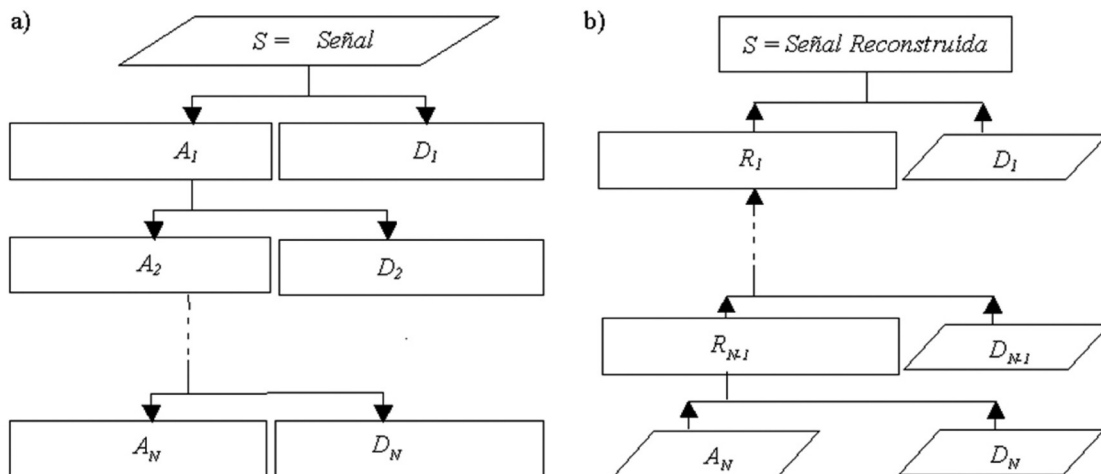
Aquí  $A_N$  es llamado el coeficiente de aproximación o señal de tendencia;  $D_N$  es el coeficiente de detalle o señal de detalle; y  $N$  indica el nivel del proceso de descomposición (o reconstrucción), donde la escala varía creciente o decrecientemente, respectivamente (Foufoula-Georgiou & Kumar, 1994). Las características y diferencias entre estas dos nuevas señales resultan del tipo de frecuencias que contiene cada una de ellas. En  $A$  se tienen las frecuencias más bajas de  $f$ , donde aparecen las frecuencias de los componentes estacionarios, periódicos y cíclicos, mientras que en  $D$  aparecen las señales de alta frecuencia existentes en la señal  $f$ ; frecuencias que describen la variabilidad, perturbación y el ruido existente en la data. En el presente trabajo las señales  $A$  y  $D$ , son nombrados como Tendencia y Ruido. Posteriormente, se puede realizar la misma descomposición para la señal  $A_1$ , obteniendo:

$$f(t) = A_1(t) + D_1(t) = A_2(t) + D_2(t) + D_1(t) \quad (6)$$

y así es posible obtener características singulares en diferentes escalas de estudio. En general, la relación entre los coeficientes de tendencia y detalle en dos niveles de descomposición (o reconstrucción) está definido como:

$$A_{N-1} = A_N(t) + D_N(t) \quad (7)$$

donde:  $N$  indica el nivel de descomposición (o reconstrucción) o escala del proceso. Esta ecuación (7) de AMR muestra que para obtener la señal  $A_N$ , a una escala menor (mayor desagregación)  $N-1$ , es necesaria la información extra ( $D_N$ ). Así  $A_{N-1}$  viene a ser la nueva señal  $A$  en la escala  $N-1$ . Este proceso es conocido como Downscaling y es equivalente al proceso de reconstrucción de AMR (Figura 2b), y el proceso inverso es equivalente al proceso de descomposición de AMR (Figura 2a). En la figura 2,  $R$  es equivalente a  $A$ , la única diferencia es que  $A$  se obtiene de un proceso de descomposición y  $R$  de un proceso de reconstrucción.



**Fig. 2.** Proceso AMR. a) Descomposición o proceso de Up-scaling. b) Reconstrucción o proceso de Downscaling. La función Wavalet Madre ( $\psi$ ) para ambos procesos es la misma.  $N=0, 1, 2, \dots$  es el nivel del proceso (aquí se define la escala como  $\lambda=2N$  y  $f$  es la escala asociada a la función  $\psi$ ).

## Generador de datos climáticos ClimGem

El ClimGem es un generador de datos climáticos a escala diaria, basado en correlaciones y regresiones (Nelson, 2002). Este generador fue desarrollado por el Departamento de Ingeniería de Sistemas Biológicos de la Universidad Estatal de Washington, y es un software de libre acceso en la Web ([www.bsyse.wsu.edu/climgen/ClimGem/index.html](http://www.bsyse.wsu.edu/climgen/ClimGem/index.html)), requiere de series diarias para calcular los parámetros a utilizar en el proceso de generación. Las variables climáticas que genera son precipitación pluvial, temperatura, radiación solar, entre otras variables.

En el caso de precipitación diaria, la reconstrucción está basada en dos supuestos: primero, que la condición de lluvia en un cierto día  $i$ , se relaciona con la condición de lluvia en el día  $i-1$  (Cadena Markov); segundo, que la cantidad de lluvia en los días lluviosos está descrita por una función de distribución adecuada, en cada mes. Los valores diarios de dichas probabilidades son interpolados usando funciones spline, siendo el requerimiento mínimo para poder generar datos con el ClimGem un registro histórico de por lo menos 25 años (continuos, no necesariamente completos).

El ClimGem es uno de los programas computacionales más empleado y conocido para generar datos diarios, por ello ha sido utilizado como referente para evaluar la metodología propuesta para la reconstrucción de datos de precipitación diaria en las estaciones y periodos de estudio. En este caso, se generaron parámetros y nuevas series en cada estación a partir de registros pluviométricos (incompletos) de entre 27 y 42 años en función de la data existente y luego se completó la serie incompleta con los valores generados para los días sin información.

## MÉTODOS

### Generación de vacíos de información aleatorios

Para simular la ausencia de datos observados (vacíos) en las estaciones N° 1, 3 y 5 (Tabla 1) se generaron vacíos en posiciones aleatorias, según el porcentaje de vacíos deseado. En cada caso (Cabanillas, Ilave y Arapa) se generó un total de 25% de datos faltantes en relación con el total de datos.

### Estudio de la información de vacíos

El estudio de vacíos consiste en determinar el porcentaje de datos faltantes en las series temporales estudiadas, para así seleccionar el nivel de descomposición que se requiere para iniciar el proceso de reconstrucción. Dicho porcentaje es calculado en distintas escalas (o niveles de descomposición de la base de la función Wavelet de Haar ( $n$ )) y posiciones. El análisis de ausencia de información en una vecindad de cada  $2N$  datos, dado por  $(\text{Número de Vacíos}) \cdot 100 / 2N$ , determinó que el nivel más adecuado para la reconstrucción era  $N = 2$ . Por lo cual, de cada 4 datos analizados el porcentaje de vacíos puede ser 0, 25, 50, 75 y 100% dependiendo si hay 0, 1, 2, 3 y 4 datos faltantes. Este nivel fue seleccionado considerando un umbral no mayor a 25% para la mayoría de vecindades, es decir de donde se pueda obtener una tendencia representativa para la posterior reconstrucción, asimismo es importante tratar de preservar el nivel de detalle y trabajar con escalas menores.

### Modelo de reconstrucción de la función Wavelet

Una vez determinado el nivel  $N$ , donde se cumple la condición de minimizar la pérdida de información en las series no completas en una cierta escala o intervalos de análisis en toda la serie, se procede a reconstruir con la Transformada Wavelet.

El modelo de reconstrucción se inicia con la normalización o estandarización de las señales (ecuación 8, Nakken, 1999) con el objetivo de convertir los datos de las señales en adimensionales y llevarlos a una misma

escala (valores entre  $-1$  y  $1$ , con media  $0$  y desviación estándar  $1$ ) para poder trabajar con ellos. Este procedimiento es realizado solo para los valores correspondientes a condiciones de lluvia (mayores a  $0$ ).

$$f(t) = \left( \frac{F(t) - \mu}{\sigma} \right) \quad (8)$$

Con las señales estandarizadas se procede a determinar la tendencia de la estación a completar y a extraer la variabilidad a heredar de la estación de apoyo en la escala  $N$ . La tendencia se obtiene dividiendo la señal en particiones de  $2^N$  y calculando el promedio de los valores observados en cada partición (sin considerar en el promedio los vacíos) resultando una nueva señal del tipo de las señales de tendencia  $A_{1N}$  (Fig. 2), cuyo número de datos será: (longitud de la señal original)/ $2^N$ . La variabilidad es extraída con la Transformada Wavelet (filtro Haar, ya que reduce la distorsión de los datos, según Das & Mohanty, 2008) al descomponer la señal completa hasta el nivel  $N$ , generando las señales  $A_{2N}$  y  $D_{2N}$  para todo  $N=1,2,\dots$ . Luego se procede a reconstruir como se muestra en la Fig. 2, pero a partir del nivel  $N$ , se toma la tendencia  $A_{1N}$  y los detalles  $D_{2N}$  con  $N=N, (N-1), \dots, 1$ . Posteriormente, se realiza el proceso contrario a la estandarización de los datos correspondientes a días lluviosos, y se eliminan los posibles errores por negatividad en la señal reconstruida e inexistencia de ruido en la señal vecina. Los errores en el primer caso son debido a la matemática utilizada en la obtención de tendencias y ruidos a diferentes escalas, y los errores en el segundo caso se dan cuando la señal vecina registró un día sin lluvia, pero la tendencia en la escala de descomposición presenta valor positivo producto de su cálculo. Ambos errores son reemplazados con cero. Finalmente, se completan los datos faltantes en la señal original, reemplazándolos con los valores recíprocos de la señal reconstruida.

El método propuesto para reconstruir datos faltantes en una señal de precipitación pluvial diaria consiste en combinar información de la estación incompleta (que se requiere completar) con información de otra estación cercana (estación de apoyo),

comprobando previamente la similitud entre sus señales estandarizadas. Dicho modelo de reconstrucción así como otros procedimientos útiles para el análisis de datos fueron programados en el software de cálculo matemático Matlab.

A continuación se describe el algoritmo de reconstrucción:

1. Seleccionar dos estaciones vecinas, una con vacíos ( $f$ ) y la otra completa ( $g$ ), de una misma familia de distribución y con indicadores climáticos similares.
2. Estandarizar las señales  $f$  y  $g$  siguiendo la metodología descrita por Nakken (1999).
3. Descomponer la señal como en las ecuaciones 3 y 4, hasta un cierto nivel de descomposición  $N$ , escogido según la cantidad de vacíos existentes en  $f$ .
4. Promediar los valores de  $f$  en el nivel  $N$ , cada  $2^N$  datos. Esto debe corresponder a una aproximación de  $A$ , por lo que se toma este resultado como la tendencia en la reconstrucción (en lugar de  $A$  utilizamos el ponderado de  $f$  cada  $2^N$  vecindades).
5. Tomar la tendencia de  $f$  (obtenida en 4) en el nivel  $N$  y todos los ruidos de  $g$  de todos los niveles  $N, (N-1), \dots, 1$ .
6. Realizar el proceso inverso de la estandarización de la parte dos y obtener la señal  $h$ .
7. Los valores faltantes de la señal  $f$  son reemplazados por los valores que se generaron en el proceso de reconstrucción de la señal  $h$ .

### Métodos de Validación

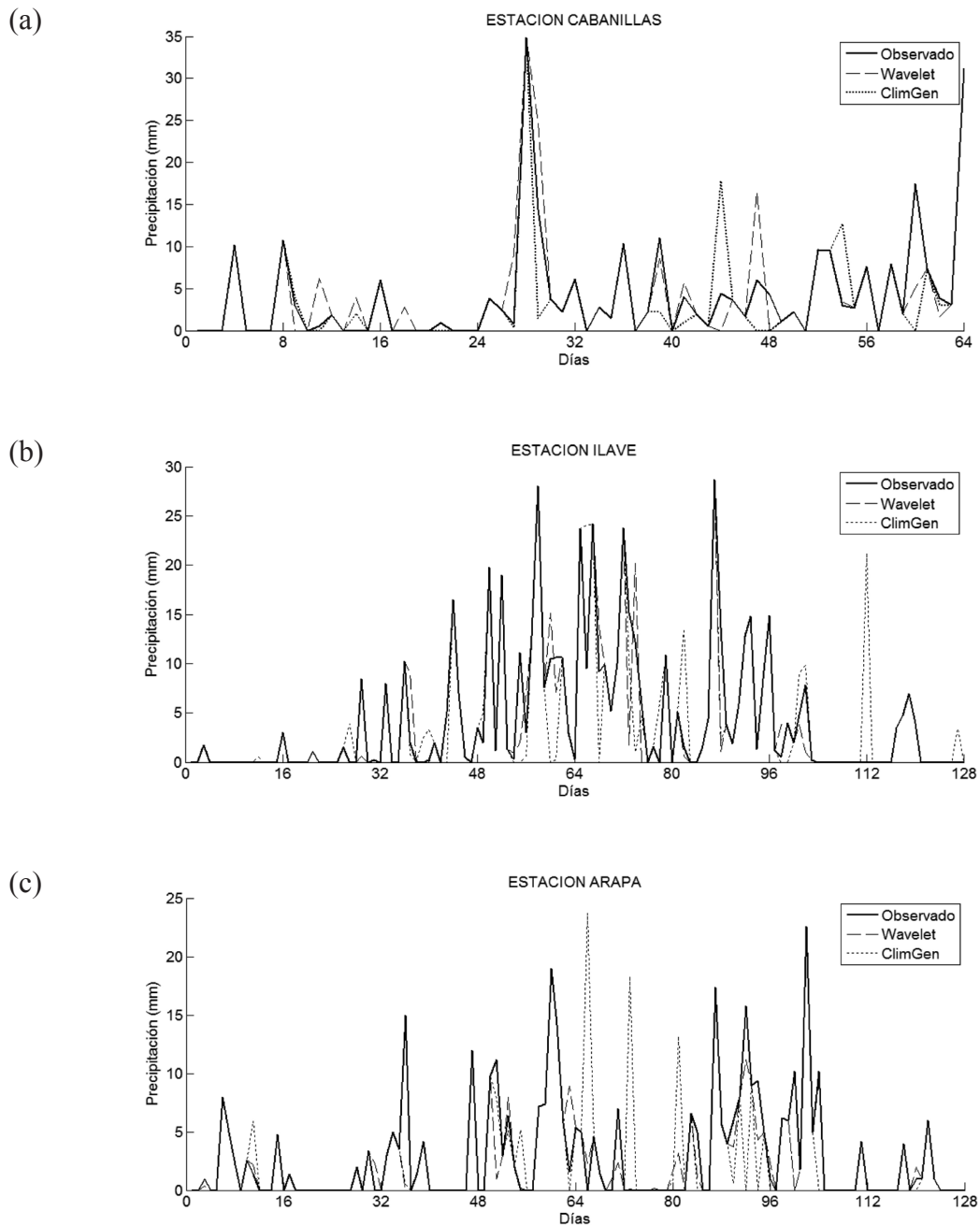
Las pruebas comparativas con los datos observados completos son importantes para la verificación de los datos modelados, tanto con el método de Transformada Wavelet como los modelados con el ClimGen. Se verificó la similitud entre estadígrafos de las señales reconstruidas con la señal original (análisis no mostrado). Se evaluó la bondad de ajuste entre cada señal modelada y la señal completa, mediante el coeficiente de determinación ( $R^2$ ) y la prueba de Kolmogorov-Smirnov; asimismo se utilizó el RMSE (Raíz del error medio cuadrático) y el

MAE (Error Absoluto Medio), indicadores no probabilísticos de predicciones para variables continuas (Wilks, 2006). Adicionalmente, se analizó la diferencia porcentual de entropías entre reconstrucciones con la hipótesis de que cuando la diferencia tiende a cero las señales se rigen por la misma ley dinámica (Feng & Tse, 2008). Finalmente, se realizó un análisis residual por cada reconstrucción.

## RESULTADOS

### Similitud de Patrones en el Altiplano

Las reconstrucciones de las señales de lluvia diaria en el Altiplano peruano se realizaron luego de estudiar y encontrar un patrón similar de comportamiento a diferentes magnitudes y dispersiones en la mayoría de estaciones. Dicho análisis se realizó mediante



**Fig. 3.** Datos observados (línea continua) y datos reconstruidos con las técnicas de reconstrucción de la TW (línea discontinua) y generador de datos ClimGen (línea punteada) para a) Cabanillas, b) Ilave y c) Arapa.



la prueba de Kolmogorov-Smirnov, comparando las señales diarias estandarizadas (análisis de 28 estaciones, no mostrado), habiéndose encontrado mayor similitud entre estaciones cercanas.

**Reconstrucción de Señales**

Luego de aplicado el modelo Wavelet y el Generador ClimGen se completaron los datos ausentes (producto de la generación de vacíos aleatorios) y se obtuvieron las siguientes señales completadas (Ver Fig. 3).

En todos los casos los valores de coeficiente de determinación (0.82, 0.86 y 0.77), correspondientes a las series reconstruidas con la técnica de la TW, fueron superiores ( $p < 5\%$ ) a los obtenidos con las reconstrucciones con el generador ClimGen (0.70, 0.71 y 0.40). Los resultados estadísticos de la comparación entre datos reales y datos reconstruidos se muestran en la Tabla 2. Los estadígrafos de las señales reconstruidas en ambos casos (media, desviación estándar, coeficiente de asimetría, curtosis) no muestran diferencias estadísticamente significativas (Tabla 2).

Las medidas de precisión de los modelos (MAE y RMSE) indican que en todos los casos el modelo Wavelet completó los datos ausentes con menor error, sin embargo según el BIAS se sobreestimaron ligeramente algunos de los valores reales, mientras que la completación con ClimGen subestimó en menor medida dichos valores.

La prueba de Kolmogorov-Smirnov muestra que las distribuciones de los datos reconstruidos con TW y ClimGen son equivalentes a la distribución de los datos observados (Tabla 2. D: máxima desviación vertical entre distribuciones acumuladas).

La diferencia porcentual de entropía de las señales indica que ambos métodos de reconstrucción dan resultados que se asemejan a los medidos. Cuando la diferencia entre los métodos fue mayor al 5%, el método de la reconstrucción con la TW presentó una menor diferencia con la señal medida.

**Análisis residual**

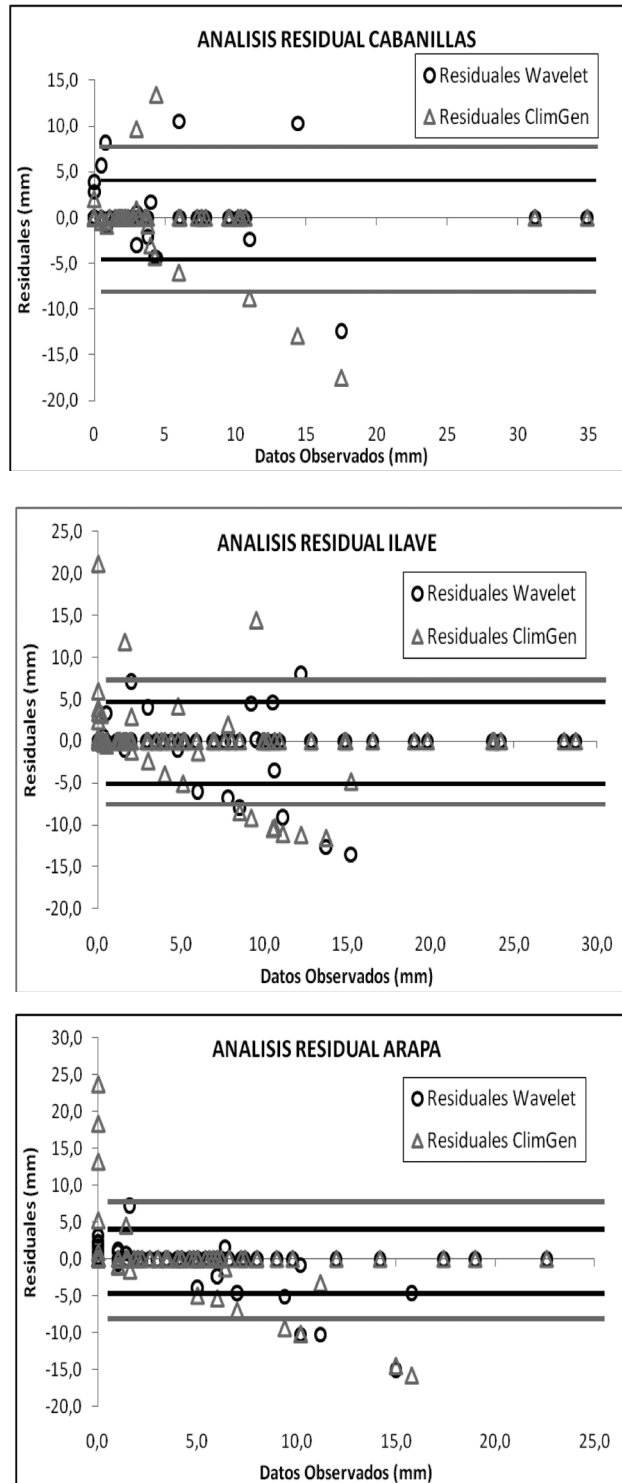
Los estadígrafos calculados y los test estadísticos realizados entre los valores observados y los valores reconstruidos con la técnica de la TW y el generador ClimGen muestran que las señales reconstruidas son similares a la señal observada. Sin embargo, es importante conocer la proporción de los errores respecto a las observaciones mediante un análisis residual. Se determinaron los residuales para los tres casos (Fig. 4), encontrándose en todos los casos que la reconstrucción con ClimGen tiene mayor cantidad de residuales que se encuentran fuera del intervalo de confianza y que en su mayoría son de mayor magnitud. El 95.5% de los residuales están comprendidos dentro del intervalo  $[\mu + 2s, \mu - 2s]$ , donde  $\mu$  es la media del residual y  $s$  es la desviación estándar residual.

**Tabla 2.** Comparación Estadística de las Reconstrucciones.

Estación Reconstruida		R <sup>2</sup>	MAE	RMSE	BIAS	p-value*	D*	ΔH%
Cabanillas	TW	0.82	1.14	2.90	0.05	1.00	0.08	-3.94
	ClimGen	0.70	1.27	3.70	-0.11	1.00	0.08	-10.57
Ilave	TW	0.86	0.74	2.40	-0.06	0.99	0.07	-0.55
	ClimGen	0.71	1.31	3.70	-0.03	1.00	0.00	1.10
Arapa	TW	0.77	0.63	2.20	-0.10	1.00	0.07	3.12
	ClimGen	0.40	1.19	3.90	-0.06	1.00	0.06	2.33

**Tabla 3.** Valores de estadígrafos principales del análisis residual.

	Cabanillas		Ilave		Arapa	
	TW	ClimGen	TW	ClimGen	TW	ClimGen
Media residual	0.22	-0.46	-0.23	-0.13	-0.27	-0.16
s residual	2.96	3.74	2.44	3.70	2.17	3.96



**Fig. 4.** Análisis residual de las reconstrucciones con la TW (o) y ClimGen ( $\Delta$ ), de: a) Cabanillas, b) Ilave y c) Arapa. Se muestran intervalos de confianza del 95.5% de los datos residuales, con la técnica de la TW (línea horizontal negra) y ClimGen (línea horizontal gris).

Los valores residuales de las reconstrucciones se encuentran distribuidos normalmente con media cercana a cero para todas las reconstrucciones realizadas,

sin embargo las desviaciones estándares reflejan una mayor dispersión del error e intervalos de confianza más amplios en las reconstrucciones con ClimGen.

## DISCUSIÓN DE RESULTADOS

La TW está basada en principios físico-matemáticos que están asociados al proceso de la señal, siendo por lo tanto independiente de la cantidad de información utilizada en el proceso de análisis (reconstrucción). Esto no ocurre con una técnica estadística clásica, donde se infieren los parámetros probabilísticamente de una muestra representativa del proceso. Esto convierte a la TW en una herramienta muy útil para este tipo de procesos. Esta ventaja permitió las reconstrucciones a partir de pocas estaciones y un número reducido de datos.

El período de análisis de las tres señales de lluvia diaria de las estaciones en estudio se seleccionó teniendo en cuenta diferentes períodos y cantidad de datos, y la presencia de eventos extremos, para tratar de observar diversos escenarios. Por ejemplo, la Fig. 3a muestra 64 datos de la señal diaria de Cabanillas, que es una señal de bajas frecuencias o baja variabilidad temporal (debido principalmente al poco número de datos). Por otro lado, las estaciones de Ilave (Fig. 3b) y Arapa (Fig. 3c), de 128 datos cada una, tienen un comportamiento de mayor variabilidad temporal (de altas frecuencias). En el caso de Ilave, la señal muestra propiedades estacionales con una variabilidad temporal creciente al inicio y decreciente en la parte final; es decir, la señal sigue una tendencia con pendiente positiva y luego negativa. Lo contrario se observa en Arapa, que tiene una señal con mucho ruido y sin tendencia. Es así que el ruido en las señales supone una mayor complejidad para su modelamiento. En estos casos las señales diarias o con eventos climáticos extremos (como El Niño), son poco modelables o los resultados suelen ser poco reales y requieren de registros prolongados. Esto último es mostrado en la gráfica de residuales y en los valores del coeficiente de determinación ( $R^2$ ), obtenidos al comparar los valores observados versus los reconstruidos con ClimGen, los que son menores a los de la reconstrucción con la TW. Para el caso de Arapa (señal con mucho ruido), el valor de  $R^2$  es mucho menor que en

los demás casos. Sin embargo, la reconstrucción con la TW no se ve afectada en ninguno de los casos indicados anteriormente, ya que el método TW propuesto utiliza la variabilidad de otra estación que tiene patrones similares (misma aleatoriedad), en un mismo tiempo (Ver Fig. 3).

En el presente trabajo se utilizaron series de precipitación diarias que presentan una alta variabilidad lo que las hace difícil de modelar. Esta variable, a nivel diario en comparación con otras variables de menor ruido o mayor escala temporal, no ha sido bien estudiada. También, las reconstrucciones a escalas de tiempo mayores (como las mensuales) se van haciendo más sencillas debido a que la precipitación mensual (u otra variable climática) es de carácter estacional y además la correlación entre estaciones meteorológicas cercanas es por lo general muy alta.

## CONCLUSIONES

La metodología desarrollada permitió completar series de precipitación diaria para tres estaciones meteorológicas, obteniéndose mejores resultados que con el generador ClimGen, según los indicadores estadísticos, no probabilísticos y físicos utilizados. En consecuencia, este modelo de reconstrucción es una alternativa eficiente para completar datos faltantes, cuando se cuenta con información del mismo periodo, en la misma escala temporal y con similares probabilidades de ocurrencia, proveniente de otra estación cercana.

La reconstrucción es posible siempre y cuando se demuestre que tanto la señal incompleta como la señal de apoyo estandarizadas tienen la misma probabilidad acumulada de ocurrencia para un mismo intervalo de tiempo.

La técnica de reconstrucción descrita es aplicada a datos de lluvia diaria, no obstante su aplicación es de uso genérico para otras variables climáticas a distintas escalas temporales (horas, cada 2 días, semanal, quincenal, mensual, anual, etc.). Esta

aplicación genérica se basa en que el análisis con la TW está influenciado por los eventos locales, preservándose el aspecto temporal.

Una ventaja de utilizar la técnica de la TW es que no requiere periodos de información pasados para su desarrollo (como sí lo requiere ClimGen y otros modelos); sin embargo, su eficiencia está asociada directamente a la relación que existe entre la variabilidad de la estación a reconstruir con la de alguna otra estación vecina y al porcentaje de datos faltantes en cada partición de  $2^N$  datos, y no al tamaño del registro histórico.

El análisis conjunto de la TW y el AMR es una técnica flexible que permite estudiar las series de datos y su variabilidad en diferentes resoluciones de manera dinámica. Es así como en las reconstrucciones con este método la presencia de eventos extremos y singularidades se mantiene, al ser heredados de la estación cercana, como se verificó en el análisis espectral.

#### Agradecimientos

A la División de Sistemas de Producción y Ambiente del Centro Internacional de la Papa y al Servicio Nacional de Meteorología e Hidrología. Este trabajo fue financiado por el Proyecto ALTAGRO con fondos del CIDA-Canadá.

#### REFERENCIAS

**Das, N.N. & Mohanty, B.P. (2008).** Temporal dynamics of PSR-based soil moisture across spatial scales in an agricultural landscape during SMEX02: A wavelet approach. *Remote Sensing of Environment*, 112(2), 522-534.

**Dunne, T. & Leopold, L.B. (1978).** *Water in Environmental Planning*: W.H. Freeman & Company, 1st Ed., 818 p.

**Farge, M. (1992).** Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics* 24, 395-458.

**Feng, J.C. & Tse, C.K. (2008).** Reconstruction of chaotic signals with applications to chaos-based communications.

Tsinghua University Press and World Scientific Publishing Co. Pte. Ltd., 218p.

**Foufoula-Georgiou, E. & Kumar, P. (Eds.). (1994).** *Wavelets in geophysics*. Academic Press, 373p.

**Gao, W. & Li, B.L. (1993).** Wavelet analysis of coherent structure at the atmosphere-forest interface. *Journal of Applied Meteorology* 32 (11), 1717-1725.

**Giddings, L. & Soto, M. (2006).** Participación de América Latina y del Caribe en los Registros Globales Climatológicos, GHCN. *Interciencia: Revista de Ciencia y Tecnología de América*, 31(9), 632-637.

**Goupillaud, P., Grossmann, A. & Morlet, J. (1984).** Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* 23 (1), 85-102.

**Grace, B. (1985).** El clima del altiplano, departamento de Puno, Perú. INIPA CIPA XV. Estación Experimental Salcedo.

**Graps, A. (1995).** An introduction to wavelets. *IEEE Computational Science and Engineering* 2 (2), 50-61.

**Grossmann, A. & Morlet, J. (1984).** Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis* 15 (4), 723-736.

**Jaffard, S. (2004).** Wavelet techniques in multifractal analysis, fractal geometry and applications, In: AMS (Ed.), *Proceedings of Symposia in Pure Mathematics*, Providence, RI.

**Katul, G. & Vidakovic, B. (1996).** The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. *Boundary Layer Meteorol*, 77(2), 153-72.

**Kumar, P. & Foufoula-Georgiou, E. (1993).** A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large-and small-scale features using wavelet transform. *Water Resources Research*, 29(8), 2515-2532.

**Kumar, P. & Foufoula, E. (1997).** Wavelet Analysis for Geophysical Applications. *Reviews of Geophysics* 35(4), 385-412.

**Labat, D., Ababou, R. & Mangin, A. (2000).** Rainfall-runoff relations for karstic springs. Part. II: Continuous wavelet and discrete

- orthogonal multiresolution analyses. *Journal of Hydrology* 238(3-4), 149-178.
- Liu, P.C. (1994).** Wavelet spectrum analysis and ocean wind waves. In: Foufoula-Georgiou, E. and Kumar, P. (Eds.), *Wavelets in Geophysics*. Academic Press, New York, pp. 151-166.
- Nakken M. (1999).** Wavelet analysis of rainfall-runoff variability isolating climatic from anthropogenic patterns. *Environmental Modelling and Software*, 14(4), 283-295.
- Nelson, R. (2002).** *ClimGen – Climatic Data Generator User’s Manual*. Washington.
- Posadas, A.N.D., Quiroz, R., Zorogastúa, P. & León-Velarde, C. (2005).** Multifractal Characterization of the Spatial Distribution of Ulexite in a Bolivian Salt Flat. *Int J Remote Sens.* 26(3), 615-627.
- Quiroz, R., León-Velarde, C., Valdivia, R., Zorogastúa, P., Baigorria, G., Barreda, C., Reinoso, J., Holle, M. & Li Pun, H. (2003).** Making a difference to Andean livelihoods through an integrated research approach. In: Harwood, R.R., Kassam, A.H. (Eds.), *Research Towards Integrated Natural Resources Management*. CGIAR, Rome, pp. 111-122.
- Quiroz, R., Yarlequé, C., Posadas, A., Mares, V. & Immerzeel, W.W. (2010).** Improving daily rainfall estimation from NDVI using a wavelet transform. Submitted to *Environmental Modelling & Software*.
- Sveinsson, O.G.B., Salas, J.D, Lane, W.L. & Frevert, D.K. (2007).** Stochastic analysis, modeling, and simulation (SAMS). User’s Manual, Version 2007. ([http://www.engr.colostate.edu/Sams-CSU-USBR/SAMS/Resources/SAMS2007\\_User\\_Manual.pdf](http://www.engr.colostate.edu/Sams-CSU-USBR/SAMS/Resources/SAMS2007_User_Manual.pdf)). Accesado en febrero 08 de 2010.
- UNESCO/Rostlac. (1982).** Guía metodológica para la elaboración del balance hídrico de América del Sur.
- Vuille, M. & Keimig, F. (2004).** Interannual variability of summertime convective cloudiness and precipitation in the central Andes derived from ISCCP-B3 data. *Journal of Climate* 17(17), 3334-3348.
- Wackernagel, H. (2003).** *Multivariate geostatistics: An introduction with applications*. Third Edition. Springer-Verlag Berlin, Heidelberg, New York, 403 p.
- Yarlequé, C. (2009).** Análisis de campos de biomasa del altiplano usando wavelet y parámetros universales multifractales. Tesis de Licenciatura en Física. Universidad Nacional del Callao, Perú. p 202.
- Yarlequé, C., Posadas & A., Quiroz, R. (2007).** Reconstrucción de datos de precipitación pluvial en series de tiempo mediante transformadas de wavelet con dos niveles de descomposición. Centro Internacional de la Papa, Working Paper No. 2007-2, Lima, Perú, 17 p.
- Wilks, D. (2006).** *Statistical Methods in the Atmospheric Sciences*. Volume 91, Second Edition (International Geophysics). 627 p.