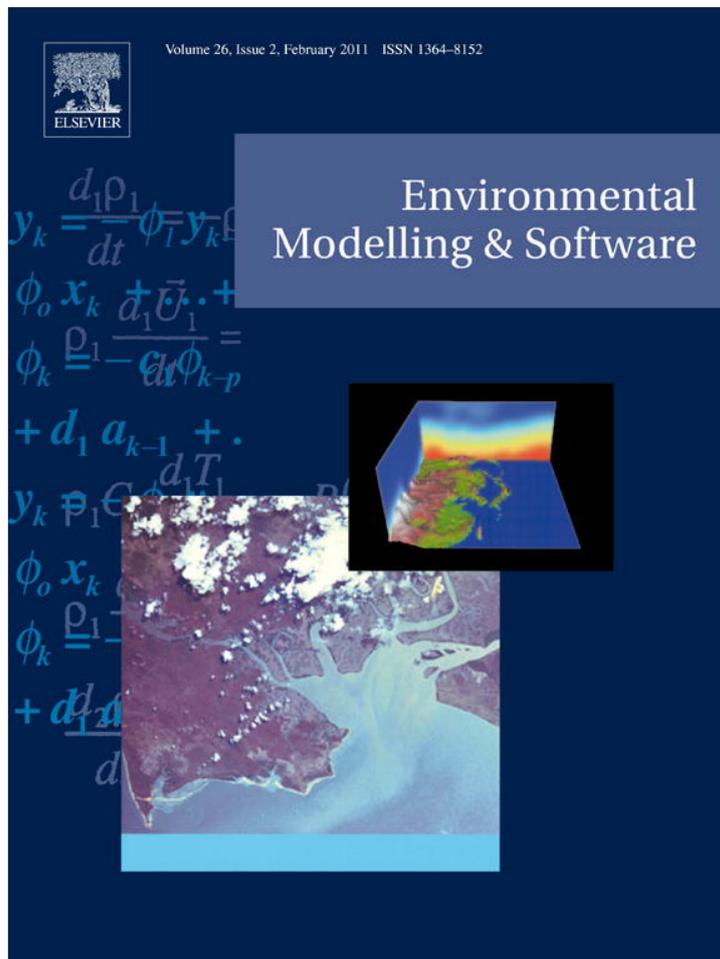


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

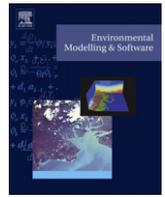
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Environmental Modelling &amp; Software

journal homepage: [www.elsevier.com/locate/envsoft](http://www.elsevier.com/locate/envsoft)

## Improving daily rainfall estimation from NDVI using a wavelet transform

Roberto Quiroz<sup>a,\*</sup>, Christian Yarlequé<sup>a</sup>, Adolfo Posadas<sup>a</sup>, Víctor Mares<sup>a</sup>, Walter W. Immerzeel<sup>b</sup><sup>a</sup> International Potato Centre, Lima, Peru<sup>b</sup> FutureWater, Wageningen, The Netherlands

## ARTICLE INFO

## Article history:

Received 18 May 2010

Received in revised form

7 July 2010

Accepted 24 July 2010

Available online 25 August 2010

## Keywords:

Rainfall

NDVI

Transforms

Wavelets

Fourier

Reconstruction

## ABSTRACT

Quantifying rainfall at spatial and temporal scales in regions where meteorological stations are scarce is important for agriculture, natural resource management and land-atmosphere interactions science. We describe a new approach to reconstruct daily rainfall from rain gauge data and the normalized difference vegetation index (NDVI) based on the fact that both signals are periodic and proportional. The procedure combines the Fourier Transform (FT) and the Wavelet Transform (WT). FT was used to estimate the lag time between rainfall and the vegetation response. Subsequently, third level decompositions of both signals with WT were used for the reconstruction process, determined by the entropy difference between levels and  $R^2$ . The low-frequency NDVI data signal, to which the high frequency signal (noise) extracted from the rainfall data was added, was the base for the reconstruction. The reconstructed and the measured rainfall showed similar entropy levels and better determination coefficients ( $>0.81$ ) than the estimates with conventional statistical relations reported in the literature where this level of precision is only found for comparisons at the seasonal levels. Cross-validation resulted in  $\leq 10\%$  entropy differences, compared to more than 45% obtained for the standard method when the NDVI was used to estimate the rainfall in the same pixel where the weather station was located. This methodology based on high resolution NDVI fields and data from a limited number of meteorological stations improves spatial reconstruction of rainfall.

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

Rainfall is a major driving force in the water cycle and the most important factor in promoting vegetation growth in rain-fed agriculture and natural grasslands and forests of the world. Accurate rainfall data with sufficient spatial resolution are of key importance in assessing basin scale hydrology but in many developing countries, adequate gauged data is seldom available.

Remote sensing can provide spatial precipitation patterns. Ground radar systems can also provide spatial precipitation information but validation of its data products is a major challenge for general hydrologic applications (Krajewski and Smith, 2002). Ground radar systems also have a limited range and are generally aimed at monitoring extreme events over limited time spans, making their use less suitable for long term assessments. Satellite remote sensing is a better source of spatial precipitation data, which are generally readily available over longer periods and large

areas. Many different algorithms and types of sensors aboard a variety of satellites exist. Adler et al. (2001) provide an extensive overview and inter-compare 25 satellite based products to four model based, and to two climatological products. As many of these products have either a poor spatial resolution ( $\sim 100$  km) or a poor temporal resolution ( $\sim 1$  month), there is a need to develop a robust downscaling methodology for precipitation.

Many studies have used the intuitive correlation between rainfall and plant biomass, particularly in arid and semi-arid environments, to fill in this rainfall data gap (see Richard and Pocard, 1998; Kawabata et al., 2001; Lotsch et al., 2003; Nicholson and Farrar, 1994; Farrar et al., 1994; Nicholson et al., 1990; Eklundh, 1998; Martiny et al., 2006; Chamaille-Jammes et al., 2006; Dinku et al., 2008). However, the vegetation response to precipitation is highly variable in space, mainly due to soil and other influencing factors. There is also a delayed response in time, termed lag time (Farrar et al., 1994), which is defined as the time required for a volume of water equal to the annual mean of exchangeable soil moisture to be depleted by the combined processes of runoff and evapotranspiration. This lag time varies for different agro-ecologies; in semi-arid regions it is usually of the order of 2–3 months (Nicholson and Lare, 1990). This process is also described by Entekhabi et al. (1996) and Brunsell and Young (2008),

\* Corresponding author. Tel.: +511 3175312; fax: +511 3175329.

E-mail addresses: [r.quiroz@cgiar.org](mailto:r.quiroz@cgiar.org) (R. Quiroz), [c.yarleque@cgiar.org](mailto:c.yarleque@cgiar.org) (C. Yarlequé), [a.posadas@cgiar.org](mailto:a.posadas@cgiar.org) (A. Posadas), [v.mares@cgiar.org](mailto:v.mares@cgiar.org) (V. Mares), [w.immerzeel@futurewater.nl](mailto:w.immerzeel@futurewater.nl) (W.W. Immerzeel).

who stated that the surface affects the role of soil moisture by acting as a low-pass filter to the high frequency rainfall signal into a near immediate moisture effect and then slowly into a vegetation effect.

Most studies have sought statistical relationships between rainfall and NDVI but they seldom go beyond simple correlation analysis (Brunsell and Young, 2008). A linear relationship between rainfall and NDVI has been reported for areas with precipitation ranging from 200 to 1200 mm per year (Nicholson et al., 1996). In other areas the upper limit is attained at lower annual precipitations (Martiny et al., 2006). Above the upper threshold, the index “saturates” and NDVI increases only very slowly with increasing rainfall or becomes constant. Statistical relationships to estimate rainfall seem to be meaningless when applied to dekadal or daily data, especially when autocorrelations are removed (Eklundh, 1998). In spite of the lack of accuracy of the statistical relationship between these two signals, this is the standard method of using satellite information to estimate rainfall in arid and semi-arid regions. Thus, current procedures for estimating rainfall from NDVI are of limited use in applications in modeling agricultural production and land-atmosphere interactions studies, where dekadal or daily rainfall is required.

The present study aimed at 1) developing a procedure for generating daily precipitation data, using reconstructions that combine wavelet-filtered signals containing the low-frequency lag-corrected vegetation greening signal, extracted from NDVI time series, and the high frequency portion of the daily rainfall data; and, 2) test whether daily rainfall events can be approximated for neighboring areas where only NDVI data is available.

## 2. Materials and methods

### 2.1. Study area

The Altiplano is a high Andean plateau centered around Lake Titicaca in the Peruvian-Bolivian border. The plateau rises from the lake level at 3800 meters (m) to over 4500 m altitude. The rainfall varies from less than 400–600 mm yr<sup>-1</sup>; average minimum temperature drops to -10 °C, droughts can last up to 150 d yr<sup>-1</sup>, while frost-free days are around 150. The dominant vegetation is natural grasslands with cultivated areas mainly near the lake (Quiroz et al., 2003).

Convective activity and precipitation in the Altiplano occur almost exclusively during the austral summer and are associated with the seasonal expansion of the upper-air easterlies and related near-surface moisture influx from continental lowlands to the east. Precipitation from the west is rare because Pacific moisture is trapped vertically by large-scale subsidence and a stable low-level inversion at 900 hPa, and laterally by the coastal escarpment (Vuille and Keimig, 2004).

### 2.2. Climate data

Rain-gauge daily data from 10 weather stations (Fig. 1) were obtained from the Peruvian national meteorology and hydrology service (SENAMHI). Data for the period from January 1st 1999 through December 31st 2002 was included in the analysis. The raw data (Fig. 2) was checked for inconsistency and outliers. The analysis was conducted for the ten sites where the weather stations were located.

### 2.3. NDVI data

A dataset containing 180 10-day (dekad) composite NDVI images derived from the SPOT-4 and SPOT-5 VEGETATION instruments was used, spanning the period January 1999–December 2003. The VGT1 sensor aboard the SPOT-4 satellite provided the data for the January 1999–January 2003 period whereas the remaining period was covered with data from the VGT2 sensor aboard the SPOT-5 satellite. Both sensors have the same spectral and spatial resolution. The red spectral band (0.61–0.68 mm) and the near-infrared (NIR) spectral band (0.78–0.89 mm) were used to calculate the NDVI (NIR - RED/NIR + RED) and the imagery had a spatial resolution of 1 km. The synthesized preprocessed S10 NDVI product, which is a geometrically and radiometrically corrected 10-day composite image (Immerzeel et al., 2005), was used. The periods were defined according to the civil calendar: from the 1st day to the 10th; from the 11th to the 20th; and from the 21st to the end of each month.

The GPS coordinates of the weather stations were co-registered with the NDVI imagery data corresponding to each site. Therefore, for each location a vector

containing 180 NDVI values, one for each civil calendar dekad, was extracted and used in the analysis. The dekadal NDVI value was repeated for each day within the respective dekad to match the daily observations in the rainfall data, generating 1826 NDVI daily values for the entire period. Given the difference in magnitude of the two signals, and for visual comparison purposes, the NDVI values were multiplied by the ratio of the mean value of both signals to generate magnitudes comparable to those registered for rainfall (Yarlequé, 2009).

### 2.4. Data pre-processing

#### 2.4.1. Fourier analysis

Fourier or harmonic analysis is a mathematical technique used to decompose a complex static signal into a series of individual cosine waves, each characterized by a specific amplitude and phase angle. Several authors have successfully applied Fourier analysis in analyzing time series of NDVI imagery (e.g. Azzali and Menenti, 2000; Roerink and Menenti, 2000; Jakubauskas et al., 2001, 2002; Moody and Johnson, 2001; Immerzeel et al., 2005).

A stationary process can be represented by a series of harmonic functions, whose frequencies are multiples of a base frequency. This series of harmonic functions is called a Fourier series. Assuming that the process can be described by a function  $S$ , the usual form of the Fourier series is (Pipes and Harvill, 1971):

$$S(t) = \frac{A_0}{2} + \sum_{n=1}^{n=\infty} A_n \cos(n\omega t) + \sum_{n=1}^{n=\infty} B_n \sin(n\omega t) \quad (1)$$

Where  $t$  = time and  $A$  and  $B$  are the Fourier coefficients.

The constant term in Eq. (1) is always equal to the mean value of the  $S(t)$ , (the mean NDVI value in a series of satellite imagery) and  $\omega = 2\pi f_0$ , where  $f_0$  is the base frequency. Eq. (1) can be written in different forms, following basic mathematical laws (Pipes and Harvill, 1971). In this research it was decided to transfer Eq. (1) to a form that only contains cosine terms, which facilitates interpretation. Eq. (1) can also be written as

$$S(t) = \frac{A_0}{2} + \sum_{n=1}^{n=\infty} C_n \cos(n\omega t - \theta_n) \quad (2)$$

Eq. (2) now has a convenient form with only cosine terms. The signal is decomposed in a series of cosine terms, each with its own amplitude ( $C_n$ ) and phase angle ( $\theta_n$ ), and a constant term ( $A_0/2$ ). When a signal is described using Fourier analysis the values for the coefficients  $C_n$  need to be found. An algorithm to recover those coefficients from a discrete signal is the Fast Fourier Transform (FFT). In this case we analyzed a signal comprised of 1826 discrete NDVI daily values to estimate the Fourier coefficients  $C_n$ . The result of the FFT is a complex vector, with a real part containing the  $A$  coefficients and an imaginary part containing the  $B$  coefficients of Eq. (1). The coefficients  $C$  of Eq. (2) can be derived from  $A$  and  $B$  by calculating the length of the vector. There are a few limitations to the FFT related to the underlying mathematics. Firstly, to correctly recover a signal from the Fourier transform of its samples, the signal must be sampled with a frequency of at least twice its bandwidth (Nyquist frequency). Secondly, the signal needs to be sufficiently static to permit the analysis under the assumption that the wave is static (intrinsic assumption of the FFT) which means that both amplitude and phase of the individual terms should not vary significantly over time.

#### 2.4.2. Determination of the lag time

The lag time between the onset of the rainy season and the greening of the vegetation was assessed with the Fourier analysis. Both rainfall ( $S_{\text{Rain}}$ ) and NDVI ( $S_{\text{NDVI}}$ ) signals were reconstructed with the six first harmonic components ( $n = 1$  to 6 in Eq. (2)) of the Fourier series, with sizes  $N$  and  $M$ , respectively. By including six harmonics in the simulation of rainfall and NDVI signals, most of the variance in the original data is explained (Immerzeel et al., 2005). These smoothed Fourier transform (SFT) signals were used to estimate the lag time between the two primary signals. A new independent variable was generated through the simulation of the SFT for different periods  $T$  (where  $T \in Z^+$ ). Out of all possible periods,  $T = 15, 30, 91, 121, 182,$  and  $365$  d were used for the analysis. Partitions  $P_T = \{0, T, 2T, \dots, kT\}$ ,  $k \in Z^+$ , with respect to  $T$  and  $kT < N, M$ , were defined. Each partition divided both signals ( $S_{\text{Rain}}$  and  $S_{\text{NDVI}}$ ) into several sub-intervals. These intervals were used to search for the lags. Fig. 3 shows this concept with a sample of two sub-intervals;  $i$  and  $i + 1$ . For each sub-interval the time difference between the peaks of the signals  $S_{\text{Rain}}$  and  $S_{\text{NDVI}}$  were registered as the lag time for that sub-interval ( $lag_i$  and  $lag_{i+1}$ ). Then the average lag time over the  $k$ -sub-intervals was obtained as:

$$Lag(T) = \langle lag_k \rangle = \langle \Delta t_k \rangle \quad (3)$$

where the  $\langle \rangle$  symbolizes average over  $k$ . Thus, we are estimating the lag time as a new function  $Lag(T)$  (Eq. (3)), of the period  $T$ . The best coefficient of determination was used for estimating the lag time for each meteorological station. Once the lag time was considered, only four complete raining seasons spanning five years were suitable for the analysis, comprising a data set of 1421 daily data pairs (NDVI, rainfall)

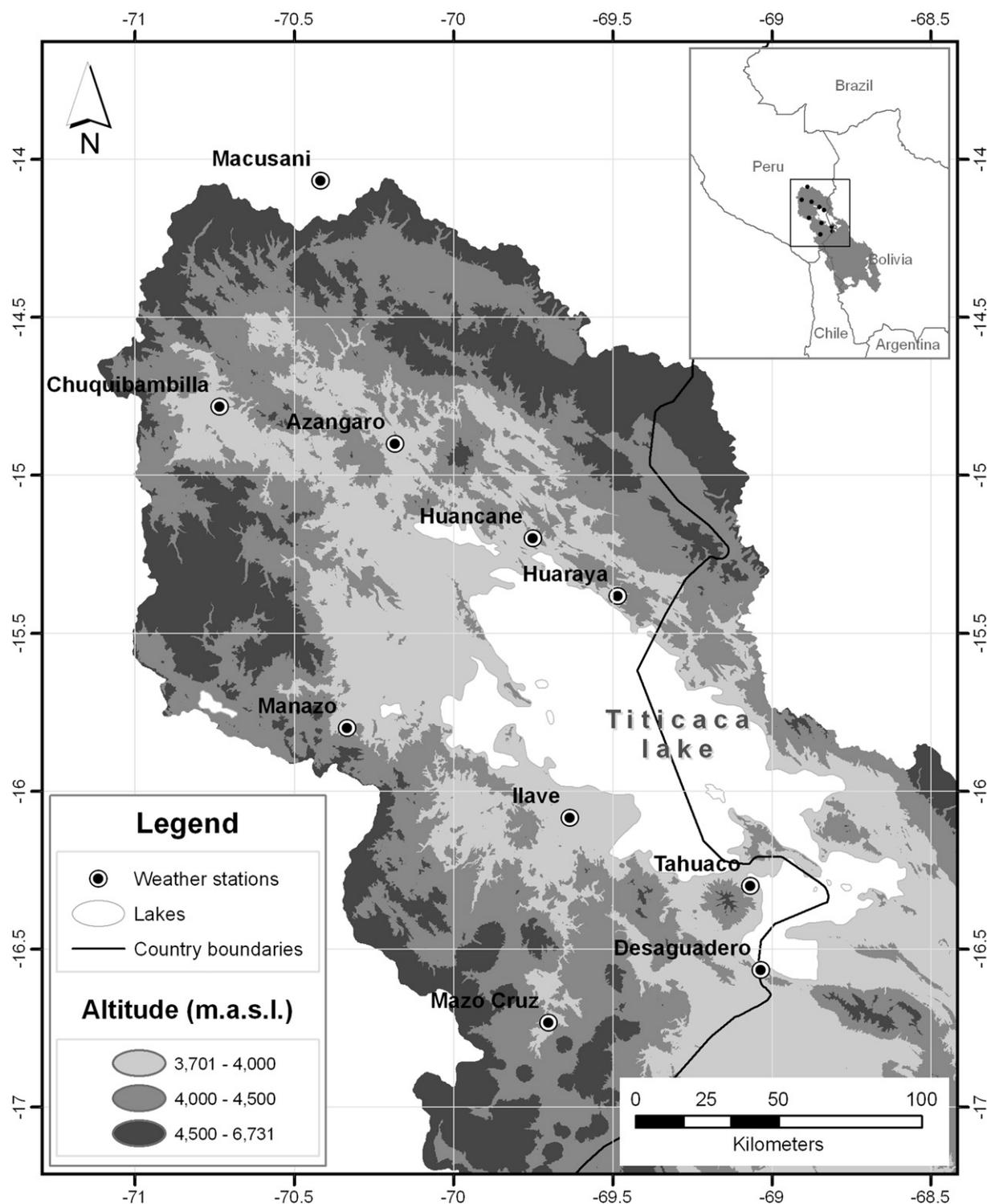


Fig. 1. Distribution of weather stations in the high plateau of Peru.

## 2.5. Wavelet analysis

### 2.5.1. The wavelet transform (WT)

The wavelet transform (WT) is a mathematical technique introduced in signal analysis in the early 1980s (Goupillaud et al., 1984; Grossmann and Morlet, 1984). It is a method based on expressing functions or signals as sums of 'little waves'. These waves are used like the sines and/or cosines in a Fourier series. Contrary to the Fourier transform, the wavelet transform is localized both in time and frequency and it has compact support. This property of wavelets is called time-frequency

localization (Foufoula-Georgiou and Kumar, 1994). It enables one to study features on the spatial series locally with a detail matched to their scale, i.e., broad features on a large scale and fine features on a small scale. This characteristic is especially useful for spatial variations that are significantly non-stationary, have short-lived transient components and features at different scales, or have singularities, which is perfectly suitable for displaying small fluctuations in signals. Due to this property, wavelet analysis has wide applications, from fluid dynamics (Farge, 1992; Gao and Li, 1993; Liu, 1994; Katul and Vidakovic, 1996) to geophysics or hydrology (Kumar and Foufoula-Georgiou, 1993; Labat et al., 2000).

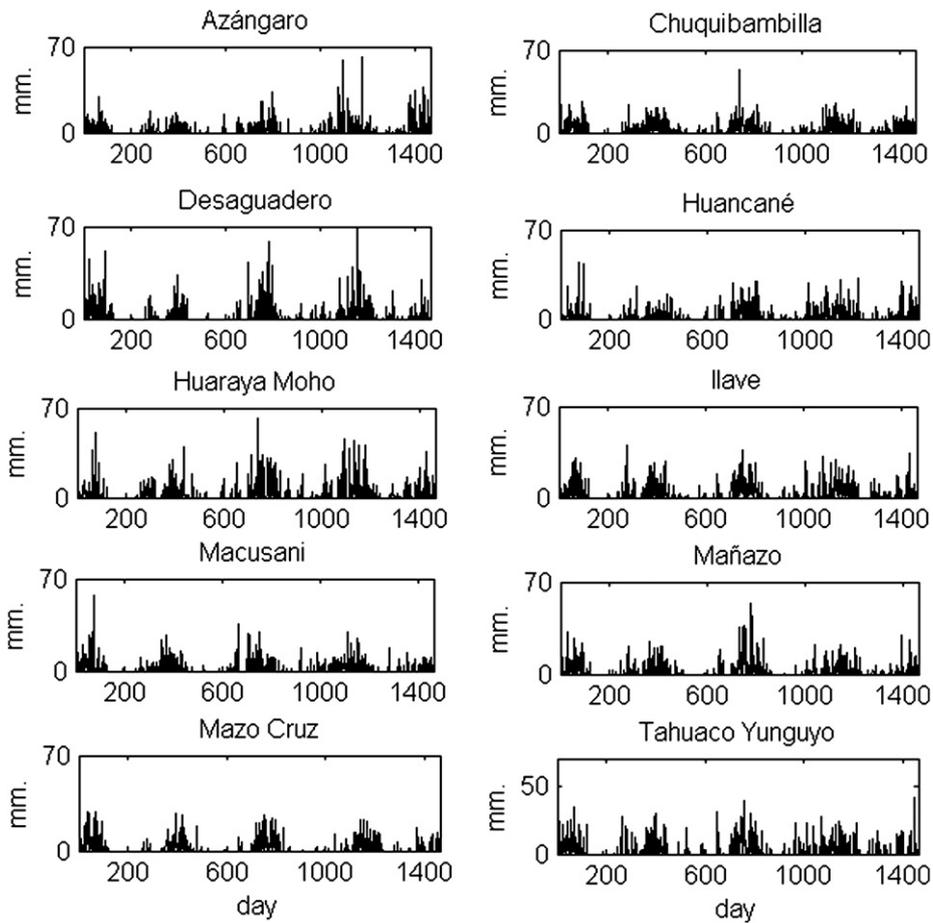


Fig. 2. Gauged rainfall signals in the ten weather stations used in the study.

A family of wavelets can be constructed from a mother wavelet function, which is ideally confined in a finite interval. Thus, we have small support for high-frequency features and large support for low-frequency or large wavelength features. This property enables one to zoom in on the irregularities of a function and characterize them locally in either spatial or frequency domains. This analysis is performed with the Wavelet Transform (WT), defined as (Foufoula-Georgiou and Kumar, 1994):

$$WTS(\lambda, \tau) = \int_{-\infty}^{\infty} S(t)\psi_{\lambda, \tau}(t)dt, \quad (4)$$

where,

$$\psi_{\lambda, \tau}(t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t-\tau}{\lambda}\right), \quad (5)$$

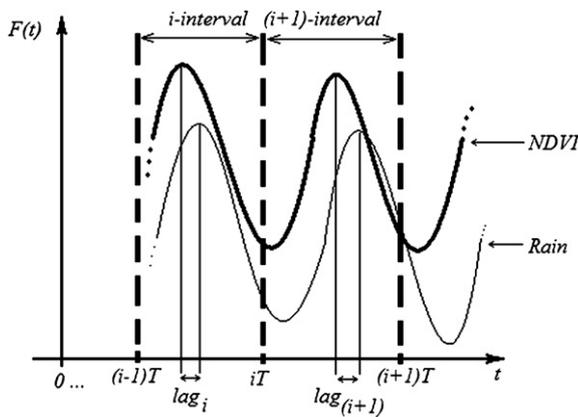


Fig. 3. Description of the Lag model for some *i*-intervals in one fixed *T*-period.

Here  $\lambda > 0$  represents the scaling factor (the wavelet's width) and  $\tau$  the shifting factor (the wavelet's position). The mother wavelet function ( $\psi(t)$ ) is generally chosen to be well localized in space (or time) and frequency (or scale). Not every function can qualify to be a mother wavelet (Mallat, 1999); it must meet the admissibility condition, described by Foufoula-Georgiou and Kumar, 1994.

2.5.2. The inverse wavelet transform (IWT)

The continuous wavelet transform is reversible due to the *admissibility condition*, which is satisfied even though the basis functions in general may not be orthonormal (Mallat, 1999). The reconstruction or inverse wavelet transform (IWT) is possible by using the following reconstruction formula on the signal transformed  $WTS(\lambda, \tau)$  to obtain the original signal  $S(t)$  (Prasad and Iyengar, 1997):

$$S(t) = IWT(S) = \frac{1}{\sigma_{\psi}} \int_{-\infty}^{\infty} \int_0^{\infty} \lambda^{-2} WTS(\lambda, \tau) \psi_{\lambda, \tau}(t) d\lambda d\tau, \quad (6)$$

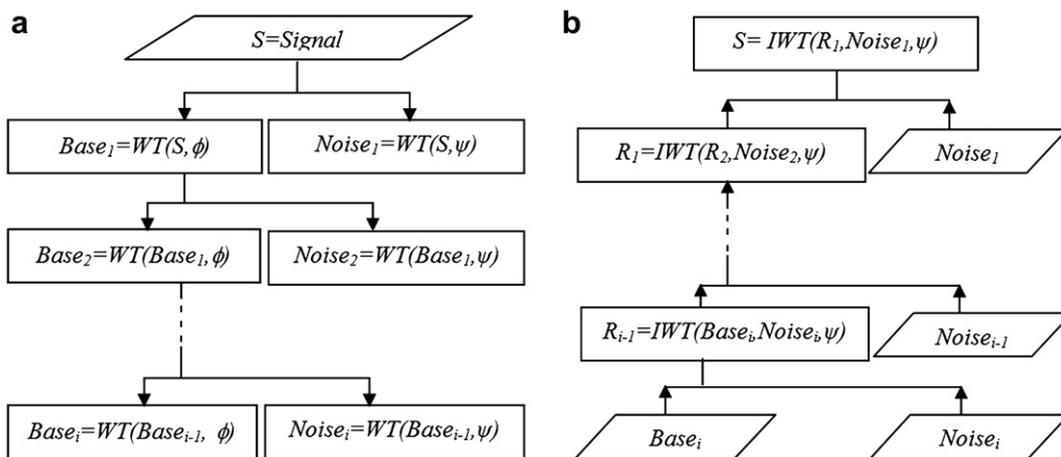
and  $\psi$  satisfies the admissibility condition:

$$\sigma_{\psi} = 2\pi \int_{-\infty}^{\infty} \frac{|F_{\psi}(w)|^2}{|w|} dw < \infty, \quad (7)$$

where  $F_{\psi}$  is the FT of  $\psi$ .

2.5.3. Multi-resolution analysis with wavelet (MRA)

Multi-resolution analysis (Mallat and Zhong, 1992), as implied by its name, comprises the evaluation of the signal at different frequencies with different resolutions. The MRA allows the decomposition of a signal into various resolution levels which retain the main features of the original signal. The filtering approach to multi-resolution WT is to form a series of half-band filters that divide a spectrum into a high-frequency band (retain information about the higher-frequency components) and a low-frequency band (contain information about lower-frequency components). It is formulated on a scaling function or low-pass filter (LP) and a wavelet function or high-pass filter (HP). These filters initially act on the entire signal band at the high frequencies (small-scale) filters and gradually reduce the signal band at each stage (see Fig. 4).



**Fig. 4.** Multi-resolution analysis processes. a) Decomposition or “up-scaling” process. b) Reconstruction or “down-scaling” process. The algorithm and the mother wavelet ( $\psi$ ) for both processes is the same.  $i = 0,1,2,\dots$  is the finite process level (decomposition (a) and reconstruction (b)) which define the scale as  $\lambda = 2^i$  and  $\phi$  is the scaling function associated to  $\psi$ .

In the present work *orthogonal wavelets*, which permit to separate effectively the low- and high-frequency signal contents at each resolution level, are used. These particular wavelet transforms re-express a correlated series in terms of some combination of uncorrelated variables (Jaffard, 2004).

The non redundant representation and perfect reconstruction of the original signal can be performed only through compactly supported wavelets. Thus, using the orthonormal conditions of the wavelet function  $\psi$ , it can be related to the **scaling function**  $\phi$  (Daubechies, 1990; Fofoula–Georgiou and Kumar, 1994), which generates a functions' family  $\phi$ . When this scaling function, where  $\psi$  and  $\phi$  are in the same vector sub-space is used, the orthogonal wavelet transform conditions on vectorial space  $L^2(\mathbf{R})$  are satisfied (Jaffard, 2004). There are some techniques to obtain  $\phi$  –functions from the wavelet  $\psi$  in order to implement the wavelet transform on sampled signals (see Fofoula–Georgiou and Kumar, 1994 for details).

Thus, *WT* is expressed through all the finite functions  $S(t)$  as a linear combination of the orthonormal wavelet base functions (Daubechies, 1990; Fofoula–Georgiou and Kumar, 1994):

$$S(t) = E(t) + N(t) \tag{8}$$

where:

$$E(t) = \sum_{k=-\infty}^{\infty} \rho_{m_0,k} \phi_{m_0,k}(t), \quad N(t) = \sum_{m=-\infty}^{m_0} \sum_{k=-\infty}^{\infty} \xi_{m,k} \psi_{m,k}(t) \tag{9}$$

and,

$$\rho_{m_0,k} = \int f(t) \phi_{m_0,k}(t) dt, \quad \xi_{m,k} = \int f(t) \psi_{m,k}(t) dt \tag{10}$$

Where  $m$  and  $k$  are integer numbers related to the scaling and shifting factors as:  $\lambda = \lambda_0^m$  and  $\tau = \tau_0^k$ . The  $E(t)$  and the  $N(t)$  signals will be called “*Base or Trend*” and “*Noise*” respectively. The *Base* represents all the characteristics of the  $S(t)$  process which are approximated through a linear combination of the  $k$  translations of the scaling function  $\phi(t)$  with the fix scale in  $2^{m_0}$ . The *Noise* conveys the details to be added as the scale is reduced to values less than  $2^{m_0}$  in the analysis process and the wavelet  $\psi(t)$  translations (Kumar and Fofoula–Georgiou, 1993). Eq. (8) represents the signal decomposition and its inverse, the reconstruction (Fig. 4).

The process described above for FFT and MRA were implemented and run, first in MATLAB, and then programmed in C++. The software can be downloaded from <http://inrm.cip.cgiar.org/home/downmod.htm>.

## 2.6. Validation methods

The information entropy has been extensively studied in communication theory (Andraud et al., 1994). Shannon and Weaver (1949) defined the information of the system by its entropy. If a system has  $Z$  different possible states with probability of occurrence  $p_i$ ,  $i = 1, 2, \dots, Z$ , then the gain in information from observing the occurrence of the event ( $i$ ) is defined as:

$$I(p_i) = \ln \frac{1}{p_i} \tag{11}$$

This definition follows from the fact that for two independent events with probabilities  $p, q$  we have  $I(pq) = I(p) + I(q)$ ; from the fact that for certain event with  $p = 1$  we have  $I(1) = 0$ ; and from the requirement  $I \geq 0$ .

The expected value of such a gain in information is defined as the entropy ( $H$ ) of the system:

$$H = \sum_{i=1}^N p_i I(p_i) \Rightarrow H = - \sum_{i=1}^N p_i \ln p_i \tag{12}$$

where  $p_i$  is the probability that the system assumes its  $i$ th possible outcome.

Entropy concepts were used for 1) helping decide at which decomposition level to stop and to assess at which level the reconstruction should start, 2) comparing similarities between gauged and reconstructed daily rainfall signals. For the first aim, entropy differences between the bases,  $\Delta H = H(\text{NDVI Base}_i) - H(\text{RAIN Base}_i)$  such that  $\Delta H \rightarrow 0$  was the criteria used. That is, when the internal information of the NDVI base is similar to the rainfall base. For the second aim, the entropy differences between the gauged and the reconstructed signals either from the conventional statistical method or the IWT were compared. The entropy metric was used because systems with identical values are in a certain sense isomorphic to each other and thus expected to have identical statistical laws of motion (Feng and Tse, 2008).

Several other metrics were also used to complement entropy and as reference to readers who use them: cumulative probability functions,  $R^2$ , the relative mean absolute error (MAE) and bias (Dinku et al., 2008).

## 3. Results and discussion

### 3.1. Decomposition with MRA

Fig. 5 shows an example of rainfall (panel a) and NDVI signal decomposition duly de-lagged. On the left hand column the low-frequency pass signals (low-pass), generated by the scaling function of the Symlet2 wavelet (Graps, 1995) are shown. They are labeled as RAIN Base $_i$  and NDVI Base $_i$  for rainfall and NDVI, respectively, for each decomposition level  $i = 1, 2, 3, 4$  and 5. These signals correspond to the trend at each level of decomposition or resolution. On the right hand column, the high-frequency pass signals (high-pass) for both series (RAIN Noise $_i$  and NDVI Noise $_i$ ) are also shown. These signals provide information on the noise or variance contribution at each resolution  $i$ .

### 3.2. Multilevel wavelet decomposition

Table 1 presents different metrics for describing the degree of association between the bases of NDVI and rainfall signals at different levels of wavelet decomposition (as described in Section 3.1). The rightmost column contains the coefficient of determination. Based on this metric, a decomposition level 4 or 5 is needed to attain an acceptable determination coefficient.

Entropy and entropy differences were also used to determine the most suitable decomposition level. There was a steep decline

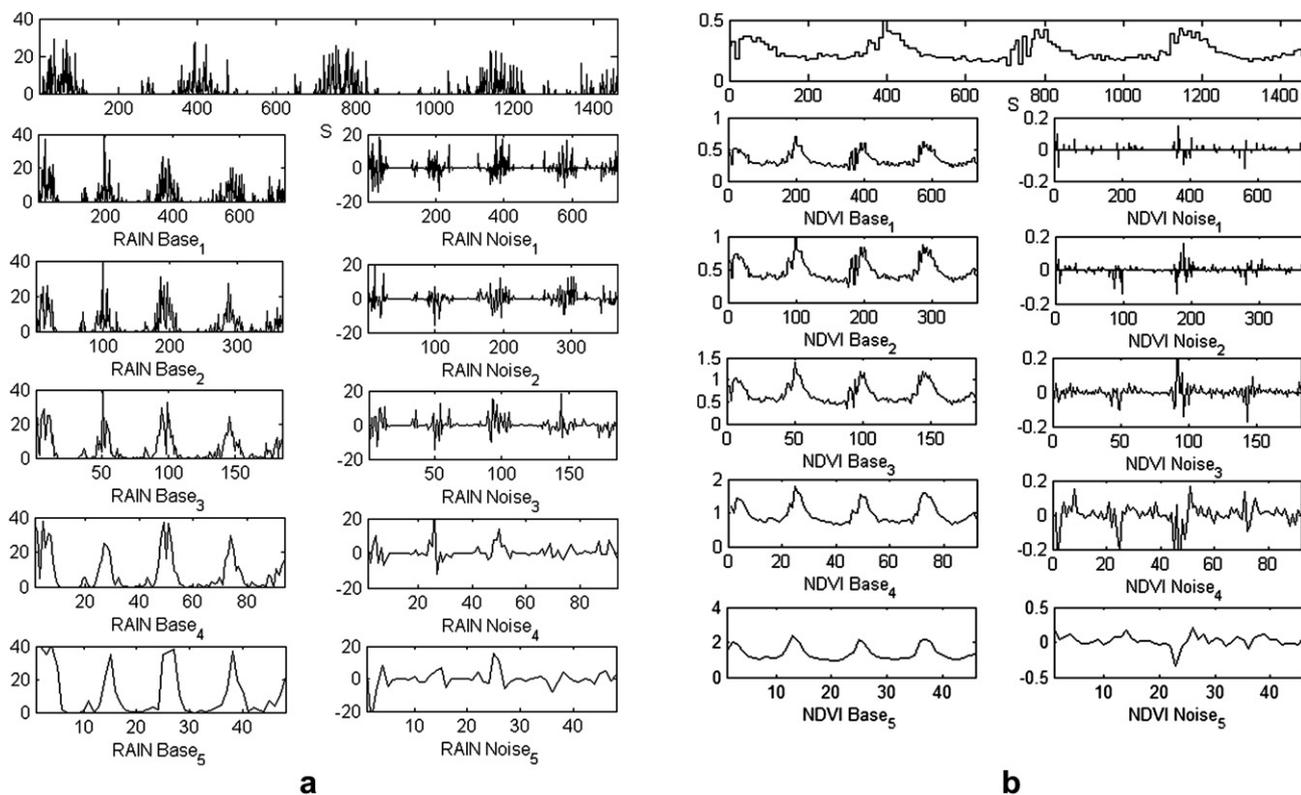


Fig. 5. Signal decomposition at 5 levels, using the MRA technique. (a) Rainfall (b) NDVI.

in  $\Delta H$  until the third level of decomposition. The entropy difference from this level onwards seems to level off (Table 1).

The two metrics (entropy and  $R^2$ ) do not corroborate each other to make a clear-cut level of decomposition. On the first hand, entropy metrics seems to suggest that when the NDVI and rainfall signals are broken down to a third level of a wavelet decomposition, at this lower resolution, the bases or low-pass signals are similar to each other. This degree of similarity seems sufficient to initiate the reconstruction process, a process described below. The scale parameter at the third decomposition level equals 8 days, which is a time interval that corresponds well to the rain characteristics of the Altiplano (slow convective systems). On the other hand, the coefficient of determination increases substantially even at the fifth level of decomposition. The coefficient of determination assesses the pair-wise comparison between the low-frequency decomposition levels of rain and NDVI.

### 3.3. Rainfall reconstruction

An inverse wavelet transform can accurately reconstruct the original signal since all the information is contained in the base and

**Table 1**  
Entropy, entropy difference, and determination coefficients for the NDVI and rainfall bases for different decomposition levels.

Level (i)	Scale ( $\lambda$ )	H NDVI Base <sub>i</sub>	H RAIN Base <sub>i</sub>	$\Delta H$	$[\Delta H/\max(\Delta H)] \times 100\%$	$R^2$ from Base NDVI <sub>i</sub> vs Base RAIN <sub>i</sub>
0	1 day	149.41	273.82	124.4	100	0.18
1	2 days	46.37	129.67	83.30	66.95	0.26
2	4 days	8.36	50.67	42.30	34	0.36
3	8 days	5.22	16.12	10.90	8.76	0.45
4	16 days	4.54	5.55	1.01	0.81	0.58
5	32 days	3.87	3.97	0.10	0.08	0.64

noise vectors at each decomposition level (see Fig. 4b). Since our interest is to combine the vectors from two different signals (NDVI and rainfall) to reconstruct daily rainfall events, it is desirable to use the minimum level of decomposition where both base signals show similarities while avoiding losing the level of details provided by the noise signal, which is flattened at higher decomposition levels. That is, a level of decomposition where the stationary or quasi-stationary properties of the base signals are maintained, while the noise signal maintains cyclic variations and other regular and irregular random events that differentiate one site from other.

Based on entropy and  $R^2$  metrics (Table 1), used to assess the decomposition levels, the reconstruction started from the third level upwards.

Fig. 6 graphically portrays an example of how the reconstruction looks like – using the inverse wavelet transform function (IWT, Eq. (6) and the Symmetlet 2 wavelet) – when the process is initiated at level 3. The low-pass signal from the third decomposition level of the NDVI (NDVI Base<sub>3</sub>, in Fig. 5b), is combined with the high-pass signal from the same level of the rainfall (RAIN Noise<sub>3</sub>, in Fig. 5a). This combination produces the signal labeled  $R_2$ . A second level reconstruction follows; for this step the reconstructed low-pass signal ( $R_2$ ) is then combined with the high-pass signal from the rainfall (RAIN Noise<sub>2</sub>, in Fig. 5a) to produce the  $R_1$  signal. The same procedure is repeated in level one to produce the reconstructed rainfall signal (S). To make sure no artificial rainfall is added to the reconstructed signal, during the periods where there is no rain, whenever the high-pass signal from the decomposition of the gauged signal was equal to zero, the reconstructed signal in the process described above (IWT) was also zero. This was based on the fact that when there is no rainfall there is no perturbation, so there is no high frequency signal.

The reconstructed rainfall signal is overlaid with the gauged rainfall signal. A visual inspection indicates a good match between reconstructed and measured daily rainfall. The determination

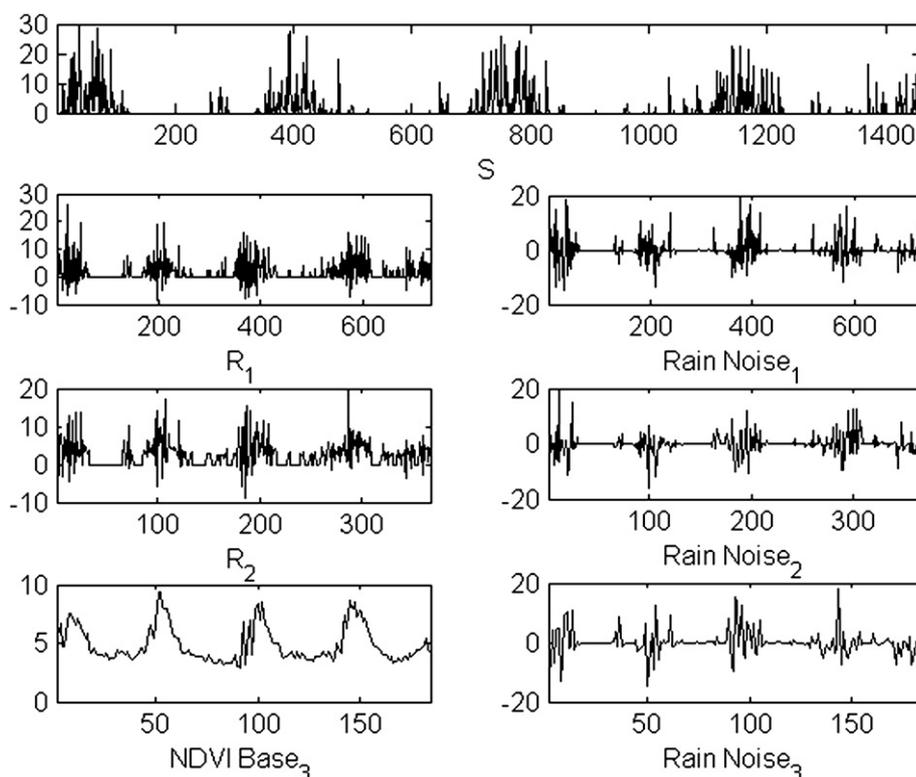


Fig. 6. Rainfall reconstruction process initiated at level 3, using the data shown in Fig. 5.

coefficient and the entropy difference were used as a metrics to ascertain the adequacy of the reconstruction. The underlying assumption for  $R^2$  is that the generated data should correspond to precisely the same gauged event. This is not necessarily true since the reconstruction process generates rainfall data containing the distributional parameters obtained from the gauged rainfall events. Therefore, a pair-wise comparison where the simulated events for  $t_i$  do not correspond to the measured event for that particular time, might generate a low determination coefficient in spite of the similarities encountered for the original and reconstructed distributions using other tests such as the Kolmogorov–Smirnov or the multifractal spectrum. On the other hand, the entropy is estimated over the entire distribution and as such seems to be a more appropriate metric to define the reconstruction level to be used.

As explained above, the entropy analysis suggested that the level three was the minimum level recommended to obtain an acceptable reconstruction. The reconstructions were conducted from levels one through four. The increments in the proportion of the variance in measured daily rainfall explained by the reconstructed signal for each reconstruction were assessed (Table 2). The table shows both the determination coefficient and the additional explanation ( $\Delta R^2 = [(R^2_{(i+1)} - R^2_i) / R^2_i] * 100\%$ ) produced when the decomposition level started at a higher level ( $i = 1$  through 4).

Table 2

Changes in the determination coefficient as affected by the level where rainfall reconstruction starts: Mazo Cruz, with  $Lag(T) = 53$  and  $T = 121$ .

Level where the reconstruction started ( $i$ )	Reconstruction vs rainfall $r^2$	$\Delta R^2$ (%)
1	0.57	—
2	0.73	28.06
3	0.82	11.21
4	0.85	4.42

As expected,  $R^2$  increases as the level of reconstruction ( $i$ ) moves from 1 to 4. It can be seen that when the reconstruction starts at level two  $R^2$  increases in 28%, compared to the reconstruction starting in level 1. This  $\Delta R^2$  substantially decreases when the reconstruction starts at levels 4 or higher (not shown). Level 3 is an appropriate starting point for reconstruction since the gain by starting in level 4 is marginal and the quality of the reconstruction is better than any estimation of daily rainfall from NDVI found in the literature. As a matter of fact, the robustness for estimating daily rainfall with this procedure is similar or better than monthly and seasonal estimations based on conventional statistical relationships, reported in the literature. The comparative analysis between the reconstructed signal -combining NDVI Base and RAIN Noise – with the gauge data is shown in Table 3. NDVI bases were estimated from the pixel where the weather stations were located. Over 89% of the gauged signal variance was explained by the reconstructed signals and the errors were relatively low. The entropy difference was also low, as will be explained below. It is important to highlight that the estimation of daily rainfall using the

Table 3

Determination coefficient for reconstructed versus gauged rainfall data and lag time for different sites in the high plateau.

Station	$R^2$	$T$ (days)	Lag( $T$ ) (days)	MAE <sup>a</sup>	Bias
Azangaro	0.89	365	19	0.87	1.12
Chuquibambilla	0.90	121	82	0.85	1.13
Desaguadero	0.87	121	57	1.22	1.23
Huancané	0.90	121	74	0.86	1.11
Huaraya Moho	0.88	121	86	1.16	1.11
Ilave	0.86	121	76	1.16	1.18
Macusani	0.93	91	84	0.58	1.10
Mañazo	0.87	365	47	0.98	1.19
Mazo Cruz	0.86	121	56	0.83	1.25
Tahuaco Yunguyo	0.85	121	43	0.99	1.19

<sup>a</sup> MAE = relative mean absolute error.

**Table 4**  
Cross-validation of daily rainfall reconstruction using percentage entropy values.

Rain Noise From Station $x^a$	NDVI Base From Station $y^a$										Classic Model Results
	1	2	3	4	5	6	7	8	9	10	
1	<b>8.31</b>	10.12	5.22	9.92	8.53	10.80	5.76	6.13	5.01	6.23	59.71
2	8.37	<b>10.02</b>	3.92	9.28	7.72	10.17	4.30	5.40	3.34	4.37	63.81
3	1.03	3.24	<b>-2.19</b>	2.95	1.53	3.10	-2.08	-1.72	-2.73	-2.00	45.69
4	5.45	6.94	1.90	<b>6.26</b>	5.57	6.80	2.34	2.73	1.54	2.73	54.67
5	5.96	7.85	2.70	6.89	<b>7.36</b>	8.06	2.92	3.31	2.10	2.89	61.65
6	7.68	9.85	3.98	8.85	8.05	<b>9.13</b>	4.42	5.36	3.67	4.52	56.50
7	3.89	5.39	-0.27	5.09	4.23	6.25	<b>0.66</b>	1.19	-0.81	0.05	57.38
8	4.40	7.42	1.02	6.16	4.52	6.52	1.34	<b>1.70</b>	0.56	1.66	57.06
9	4.83	7.78	1.34	6.04	5.07	7.03	1.67	2.15	<b>0.89</b>	2.01	45.44
10	2.49	5.04	-1.72	3.24	2.43	4.88	-1.18	-0.41	-2.03	<b>-0.57</b>	58.97

The diagonal portrayed by the values in bold represent the NDVI base and the rain noise corresponding to the same station.

<sup>a</sup> Station  $x$  or  $y$ : 1 = Azangaro; 2 = Chuquibambilla; 3 = Desaguadero; 4 = Huancane; 5 = Huaraya Moho; 6 = Ilave; 7 = Macusani; 8 = Mañazo; 9 = Mazo Cruz; 10 = Tahuaco Yunguyo.

statistical conventional method, explained less than 20% of the variance (data not shown).

### 3.4. Lag time

Table 3 shows the lag times estimated for  $S_{\text{Rain}}$  and  $S_{\text{NDVI}}$  time series using different periods  $T$ . It also presents the correlations between the measured rainfall time series and the reconstructed one using three decomposition levels.  $T = 121$  d was the best time resolution for analyzing the lag time across most sites. Only three of the sites showed better fit for other periods, Mañazo, Azangaro and Macusani ( $T = 365$  d for the first two and 91 d for the latter). Nonetheless, the coefficients of determination for  $T = 121$  were also high and significant; 0.82, 0.84, and 0.90 for the three respective sites.

The lag time or Lag ( $T$ ) was analyzed for a range of periods  $T$ . This lag time varied from 42 to 86 d. It was shown that there was a good correlation between the reconstructed and gauged daily rainfall ( $0.91 > R^2 > 0.81$ ). The MAE is lower than values reported for monthly data (Dinku et al., 2008) and the bias show a slight underestimation

Sites with shorter lag times were associated with sandy loam or loam soils; whereas longer lag times coincided with sandy clay loam soils. Lag time did not vary significantly with topography, mean temperature, and land use.

Similar lag times were found in semi-arid regions in Africa with similar rainfall patterns (Farrar et al., 1994; Nicholson and Lare, 1990; Entekhabi et al., 1996 and Brunsell and Young, 2008).

### 3.5. Cross-validation

It is important to highlight the diversity in daily rainfall events when all the seasons and sites are combined (See Fig. 2). Notwithstanding, the Kolmogorov–Smirnov test showed that all the rainfall events registered across gauge stations were samples of the same distribution, thus allowing us to “extrapolate” the noise from the gauged data to areas without physical measurements. Our comparison across 10 sites in the high plateau of Peru, using the entropy difference between the gauged and the reconstructed signals, is shown in Table 4. The comparison is based on reconstruction initiated at level 3. Several aspects of the comparison are presented. The rightmost column depicts the entropy difference (as percent) between the gauged signal and the one estimated using the conventional statistical relationship between gauged rainfall and NDVI of the pixel where the weather station is located. Note that the entropy difference ranged from 45 to 64%, indicating that the signals compared were different. The rest of the table contains the results of the cross-validation analyses, with the exception of

the main diagonal which portrays the results where the NDVI Base comes from the pixel where the weather station was located. The entropy difference in this main diagonal is attributed to the use of the NDVI Base, rather than a RAIN Base, which would provide a reconstruction of the original signal.

The entropy difference values off-main diagonal are the cross-validation, where the RAIN Noise is taken from the vertical stations and the NDVI Base from the horizontal stations. The matrix is non-symmetrical since neither the bases nor the noises are similar. For instance, taking the RAIN Noise from Desaguadero to reconstruct Azangaro (NDVI Base from Azangaro) generated an entropy difference of 1.03% (value in row 3, column 1). On the other hand, using the RAIN Noise from Azangaro to reconstruct Desaguadero, generates an entropy difference of 5.22% (row 1, column 3). All in all, entropy differences between signals reconstructed with detailed information given by any other weather station in the Altiplano, even when the two sites were over 200 km apart, were less than 11 percent. It is hypothesized that the gauged and reconstruct signals with the IWT may be controlled by similar statistical laws of motion (Feng and Tse, 2008). For this particular case this might be associated to the fact that the rainfall in the entire Altiplano and surroundings is controlled by the Bolivian high (Vuille and Keimig, 2004).

## 4. Conclusions

In this paper we showed a new reconstruction tool to generate daily rainfall from NDVI data. The tool is supported by the Wavelet Transform that maintains the same distributional properties of the measured events. The results obtained for the highly variable rainfall of the Andean highlands were superior to reconstructions based on conventional statistical relations reported in the literature. Actually the explanatory power of the reconstructed signal is comparable to exercises conducted at the seasonal level, using conventional statistical relationships between the two data sets.

Entropy analysis of the signals was a good metric to select the level of wavelet decomposition needed to maintain the distinguishing feature of rainfall events across space (point estimates within a region in this exercise) and time thus assuring a better representation of the phenomena being reconstructed. Entropy was also successfully used for assessing similarities between gauged and reconstructed rainfall signals.

Wavelet analysis proved to be a very flexible technique. It allowed the viewing of a data series as it would be observed at different resolutions, referred to as “levels of decomposition” (Brunsell and Young, 2008). The analysis provided the mechanism for separating the trend from the noise thus providing information on when an event occurs and its relative importance or variance.

The method described in this paper is suitable for reconstructing daily rainfall events through combining two signals, the trend from NDVI time series and the detailed variance of daily rainfall events, from gauge measurement in neighboring sites. This is suggested by the fact that it was feasible to reconstruct rainfall signals using the high frequency detail obtained from the gauged information extracted from any weather station even when this site was over 200 km apart from the site where the NDVI was extracted. The extrapolation domain for weather stations was not formally addressed in this research, since the spatial anisotropy of rainfall events must be analyzed. This investigation is being conducted in our laboratory using multifractal analysis.

## References

- Adler, R.F., Kidd, C., Petty, G., Morissey, M., Goodman, H.M., 2001. Intercomparison of global precipitation products: the third precipitation intercomparison project (PIP-3). *Bulletin of the American Meteorological Society* 82 (7), 1377–1396.
- Andraud, C., Beghdadi, A., Lafait, J., 1994. Entropic analysis of random morphologies. *Physica A: Statistical and Theoretical Physics* 207 (1–3), 208–212.
- Azzali, S., Menenti, V., 2000. Mapping vegetation-soil-climate complexes in southern Africa using temporal Fourier analysis of NOAA-AVHRR NDVI data. *International Journal of Remote Sensing* 21 (5), 973–996.
- Brunsell, N.A., Young, C.B., 2008. Land surface response to precipitation events using MODIS and NEXRAD data. *International Journal of Remote Sensing* 29 (7), 1965–1982.
- Chamaille-Jammes, S., Fritz, H., Murindagomo, F., 2006. Spatial patterns of the NDVI-rainfall relationship at the seasonal and interannual time scales in an African savanna. *International Journal of Remote Sensing* 27 (23), 5185–5200.
- Daubechies, I., 1990. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory* 36 (5), 961–1005.
- Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S.J., Ropelewski, C.F., 2008. Validation of high-resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing* 29 (14), 4097–4110.
- Eklundh, L., 1998. Estimating relations between AVHRR NDVI and rainfall in East Africa at 10-day and monthly time scales. *International Journal of Remote Sensing* 19 (03), 563–570.
- Entekhabi, D., Rodríguez-Iturbe, I., Castelli, F., 1996. Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology* 184 (1–2), 3–17.
- Farge, M., 1992. Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics* 24, 395–458.
- Farrar, T.J., Nicholson, S.E., Lare, A.R., 1994. The influence of soil type on the relationships between NDVI, rainfall, and soil moisture in semiarid Botswana. II. NDVI response to soil moisture. *Remote Sensing of Environment* 50 (2), 121–133.
- 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.
- Foufoula-Georgiou, E., Kumar, P. (Eds.), 1994. *Wavelets in Geophysics*. Academic Press, p. 373.
- Gao, W., Li, B.L., 1993. Wavelet analysis of coherent structure at the atmosphere-forest interface. *Journal of Applied Meteorology* 32 (11), 1717–1725.
- Goupillaud, P., Grossmann, A., Morlet, J., 1984. Cycle-octave and related transforms in seismic signal analysis. *Geophysical Research Letters* 11, 85–102.
- 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.
- Immerzeel, W.W., Quiroz, R.A., De Jong, S.M., 2005. Understanding precipitation patterns and land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series. *International Journal of Remote Sensing* 26 (11), 2281–2296.
- Jaffard, S., 2004. Wavelet techniques in multifractal analysis, fractal geometry and applications. In: *Proceedings of Symposia in Pure Mathematics*, AMS, Providence, RI.
- Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering & Remote Sensing* 67 (4), 461–470.
- Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture* 37 (1–3), 127–139.
- Katul, G., Vidakovic, B., 1996. The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. *Boundary Layer Meteorology* 77 (2), 153–172.
- Kawabata, A., Ichii, K., Yamaguchi, Y., 2001. Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *International Journal of Remote Sensing* 22 (7), 1377–1382.
- Krajewski, W.F., Smith, J.A., 2002. Radar hydrology: rainfall estimation. *Advances in Water Resources* 25 (8–12), 1387–1394.
- 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.
- 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., Kumar, P. (Eds.), *Wavelets in Geophysics*. Academic Press, New York, pp. 151–166.
- Lotsch, A., Friedl, M.A., Anderson, B.T., Tucker, C.J., 2003. Coupled vegetation-precipitation variability observed from satellite and climate records. *Geophysical Research Letters* 30 (14), 1774.
- Mallat, S., 1999. *A Wavelet Tour of Signal Processing*, second ed.. Academic Press, p. 637.
- Mallat, S., Zhong, S., 1992. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (7), 710–732.
- Martiny, N., Camberlin, P., Richard, Y., Philippon, N., 2006. Compared regimes of NDVI and rainfall in semi-arid regions of Africa. *International Journal of Remote Sensing* 27 (23), 5201–5223.
- Moody, A., Johnson, D.M., 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment* 75 (3), 305–323.
- Nicholson, S.E., Davenport, M.L., Malo, A.R., 1990. A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic Change* 17 (2–3), 209–241.
- Nicholson, S.E., Farrar, T.J., 1994. The influence of soil type on the relationships between NDVI, rainfall, and soil moisture in semiarid Botswana. I. NDVI response to rainfall. *Remote Sensing of Environment* 50 (2), 107–120.
- Nicholson, S.E., Lare, A.R., Marengo, J.A., Santos, P., 1996. A revised version of Lettau's evapoclimatology model. *Journal of Applied Meteorology* 35 (4), 549–561.
- Nicholson, S.E., Lare, A.R., 1990. A climatonic description of the surface energy balance in the central Sahel. Part II: the evapoclimatology submodel. *Journal of Applied Meteorology* 29 (2), 138–146.
- Pipes, L.A., Harvill, L.R., 1971. *Applied Mathematics for Engineers and Physicists*, third ed.. McGraw-Hill, Singapore, p. 1015.
- Prasad, L., Iyengar, S.S., 1997. *Wavelet Analysis with Applications to Image Processing*. CRC Press, p. 279.
- 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.
- Richard, Y., Poccard, I., 1998. A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing* 19 (15), 2907–2920.
- Roerink, G.J., Menenti, M., Verhoef, W., 2000. Reconstructing cloud free NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing* 21 (9), 1911–1917.
- Shannon, C.E., Weaver, W., 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, p. 125.
- 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.
- 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. pp. 202.