Estimating daily rainfall from NDVI using the wavelet transform

Christian Yarlequé[1]; Adolfo Posadas[1,2]; Víctor Mares[1]; Roberto Quiroz[1]

[1]Centro Internacional de la Papa, Apartado Postal 1558, Lima 12-Perú,
[2] Brazilian Agricultural Research Corporation (EMBRAPA)/ Embrapa Agricultural Instrumentation (CNPDIA)
P.O. Box. 741, 13560-970, São Carlos-SP, Brazil.
c.yarleque@cgiar.org; a.posadas@cgiar.org; v.mares@cgiar.org; r.quiroz@cgiar.org

Quantifying rainfall at spatial and temporal scales is a challenge posed to scientists in different disciplines given its importance in agriculture, natural resource management and land-atmosphere interactions. This paper describes a new approach to assess rainfall combining rain gauge data with the normalized difference vegetation index (NDVI) based on the fact that both events are periodic and proportional. The procedure developed to reconstruct the rainfall signal combining the Fourier Transform (FT) and the Wavelet Transform (WT) is described. FT was used to estimate the lag time between rainfall and the vegetation response. 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 signal from the NDVI data was used as the base signal for the reconstruction to which the high frequency signal (noise) extracted from the rainfall data was added. The reconstructed daily rainfall was compared to the measured one obtaining determination coefficients > 0.81. This finding is quite an improvement over the estimates reported in the literature where this level of precision is only found for comparisons at the seasonal levels. This methodology has clear scope to improve spatial interpolation of rainfall based on high-resolution NDVI fields and a limited number of meteorological stations.

Keywords: Rainfall, NDVI, transforms, wavelets, Fourier, and reconstruction.

Introduction

Numerous studies have used the intuitive correlation between rainfall and biomass, particularly in arid and semi-arid environments to fill in this rainfall data gap (see Richard and Poccard, 1998; Kawabata *et al*., 2001; Nicholson and Farrar, 1994; Farrar *et al*., 1994; Nicholson *et al.,* 1990; Eklundh, 1998; Martiny *et al.,* 2006; and Dinku *et al,.* 2008). The vegetation response to precipitation is highly variable in space, mainly due to soil and other factors influencing the vegetation response. The delayed response in time (lag) has been termed residence time (Farrar *et al.,* 1994) and defined as the time required for a volume of water equal to the annual mean of exchangeable soil moisture to be depleted by runoff and evapotranspiration. This lag time varies for different agroecologies; in semiarid regions it is usually on the order of 2-3 months (Nicholson and Lare, 1990). 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.,* 1994). Above the upper threshold, the index "saturates", and NDVI increases only very slowly with increasing rainfall or becomes constant. Actual 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 aims to develop a methodology to reconstruct daily precipitation based on NDVI and precipitation data, to further improve spatial precipitation fields with a high temporal resolution through a robust procedure. Secondly the focus is on the assessment of the lag time and a further analysis of vegetation response to rainfall.

Materials and methods

Study area

The Altiplano is a high Andean plateau centered geographically and socioeconomically on Lake Titicaca. The plateau rises from the lake level at 3,800 meters (m) to over 4,500 m altitude and is bisected by the international border between Peru and Bolivia. For more details see Quiroz *et al*., 2003. The analysis presented in this paper addresses the rainfall situation on the Peruvian side.

Climate data

Rain-gauge daily data from 10 weather stations were obtained from the Peruvian national meteorology and hydrology service (SENAMHI). The period January 1st 1999 through December 31st 2002 was used in the analysis.

The raw data was checked for consistency and outliers. The analysis was conducted for the ten sites where the weather stations were located.

NDVI data

A dataset containing 197 10-day (dekad) composite NDVI images derived from the SPOT-4 and SPOT-5 VEGETATION instruments was used, spanning the period January 1999 through December 2003. 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 GPS coordinates of the weather stations were co-registered with the NDVI imagery for the extraction of the data corresponding to each site. The dekadal NDVI value was repeated for each day within the respective dekad to match the daily observations in the rainfall data. These original NDVI values were multiplied by the ratio of the mean value of both signals to generate magnitudes comparable to those registered for rainfall.

Data pre-processing

Fourier analysis

For a rainfall process described by a function S(t), the Fourier series can be expressed as (Pipes and Harvill, 1971):

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

The constant term in equation (1) is always equal to the mean value of the equation, e.g. the mean NDVI value in a series of satellite imagery. The signal is decomposed in a series of cosine terms, each with its own amplitude (C_n) and phase angle (θ_n), and a constant term ($A_0/2$) and $\omega = 2\pi f_0$, where f_0 is de base frequency. When a signal is described using Fourier analysis the values for the coefficients C_n need to be found.

Determination of the time lag

The time lag between the onset of the rainy season and the greening of the vegetation was assessed with the Fourier analysis. Both rainfall (S_{Rain}) and NDVI (S_{NDVI}) signals were reconstructed with the six first harmonic components (n=0 to 6 in equation 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 between the two (Yarlequé *et al.*, 2007; Yarlequé, 2009). 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 $P_T = \{0, T, 2T, \ldots, kT\}$, $k \varepsilon Z^+$, with respect to T and $kT < N,M$, were defined. Each partition divided both signals (S_{Rain} and S_{NDVI}) into several sub-intervals. These intervals were used to search for the lags. Then the average lag over the k-sub-intervals was obtained as:

$$Lag(T) = \langle lag_k \rangle = \langle \Delta t_k \rangle , \qquad (2)$$

where the <> symbolizes average over k. Thus, we are estimating the lag as a new function $Lag(T)$ (equation 2), of the period T. The best coefficient of determination was used for determining the residence time for each meteorological station.

Wavelet analysis

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).

The Wavelet Transform (WT)

The Wavelet Transform (WT) is defined as (Foufoula–Georgiou and Kumar, 1994):

$$WTS(\lambda,\tau) = \int_{-\infty}^{\infty} S(t)\psi_{\lambda,\tau}(t)dt, \tag{3}$$

Where,

$$\psi_{\lambda,\tau}(t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t-\tau}{\lambda}\right), \tag{4}$$

Here λ>0 represents the scaling factor (the wavelet's width) and τ the shifting factor (the wavelet's position). The mother wavelet function (*ψ(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. The Inverse Wavelet Transform (IWT) is deduced from equation (3), i.e. the S(t) function can be reconstructed from the WTS (Prasad and Iyengar, 1997). The Multi-Resolution Analysis with Wavelet (MRA) is described in more details in the works of Mallat, 1999; Daubechies, 1990; Foufoula–Georgiou and Kumar, 1994. This technique is used to implement a decomposition (upscaling) and a reconstruction (downscaling) of the S(t) function (signal) in several scales (levels), realizing a cascade process (Yarlequé *et al.*, 2007; Yarlequé, 2009; Foufoula-Georgiou and Kumar, 1994; Mallat, 1999). This cascade process is illustrated in the decomposition and reconstruction process with MRA in the results section.

Validation methods

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

$$H = -\sum_{i=1}^{N} p_i \ln p_i \tag{5}$$

Where p_i is the probability that the system assumes its i^{th} possible outcome. Entropy concepts were also used for helping decide at which decomposition level to stop and to assess at which level the reconstruction should start. Entropy differences between the bases, ΔH = H(NDVI Base_i)- H(RAIN Base_i), such that ΔH →0 was the criteria used. That is, when the internal information of the NDVI base is similar to the rainfall base. The R^2, the relative mean absolute error (MAE) and Bias (Dinku *et al.,* 2008), were use to validate the reconstruction results.

Results and discussion

Figure 1 shows an example of NDVI (panel a) and rainfall 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. 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.

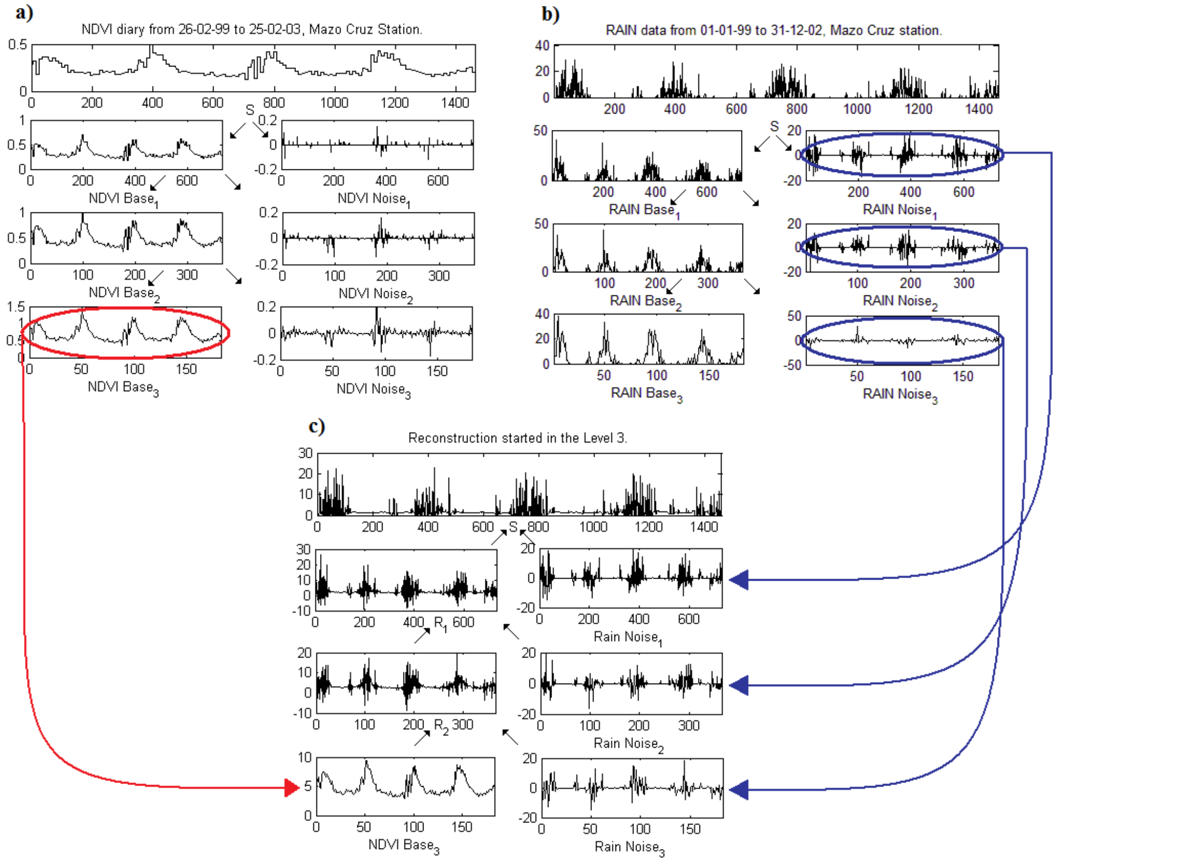

Figure 1. Signal decomposition at 3 levels, using the MRA technique for (a) NDVI (b) Rainfall data, following the arrows sense. c) Rainfall reconstruction process initiated at level 3, using the data shown in figures 1a and 1b, in the inverse arrows sense.

Table 1 presents different metrics for relating the degree of association between the bases of NDVI and rainfall signals at different levels of wavelet decomposition. The rightmost column contains the coefficient of determination. Based on this metric, a decomposition level 4 or 5 is needed to attain an acceptable R^2. Entropy and entropy differences were also used to determine the most suitable decomposition level. There was a steep decline in ΔH until the third level of decomposition. The entropy difference from this level onwards seems to level off (Table 1)

Table 1. Entropy, entropy difference, and R^2 values for the NDVI and rainfall bases for different decomposition levels

Level (*i*)	Scale ( )	H NDVI Base $_i$	H RAIN Base$_i$	H	[H/max(H)]*100%	R^2 from Base NDVI$_i$ vs Base RAIN$_i$
0	1day	149.41	273.82	124.4	100	0.18
1	2days	46.37	129.67	83.30	66.95	0.26
2	4days	8.36	50.67	42.30	34.00	0.36
3	8days	5.22	16.12	10.90	8.76	0.45
4	16days	4.54	5.55	1.01	0.81	0.58
5	32days	3.87	3.97	0.10	0.08	0.64

Rainfall reconstruction

An inverse wavelet transform can accurately reconstruct the original signal since all the information is contained in the base and noise vectors at each decomposition level (Figure 1c). Based on entropy (Table 1) and R^2, metrics used to assess the decomposition levels, the reconstruction started from the third level upwards.

Figure 1 graphically portrays an example of how the reconstruction looks like - using the inverse wavelet transform function (IWT, section 2.5.1 and the Symmlet 2 wavelet) - when the process is initiated at level 3. The low-pass signal from the third decomposition level of the NDVI (NDVI $Base_3$, in figure 1a), is combined with the high-pass signal from the same level of the rainfall (RAIN $Noise_3$, in figure 1b). 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 $Noise_2$, in figure 1b) to produce the R_1 signal. The same procedure is repeated in level one to produce the reconstructed rainfall signal (S).

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*)	R^2: Reconstruction vs rainfall	R^2 (%)
1	0.56	-----
2	0.72	29.21
3	0.82	13.16
4	0.86	4.96

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). As expected, R^2 increments 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 29 %, compared to the reconstruction starting in level 1. This ΔR^2 substantially decreases when the reconstruction starts at levels 4 or higher (not shown). Levels 3 or 4 can be the starting points for reconstruction and the quality of the reconstruction is better than any estimation of daily rainfall from NDVI found in the literature (Figure 2). As a matter of fact, the robustness for estimating daily rainfall with this procedure is similar or better than monthly and seasonal estimations reported in the literature.

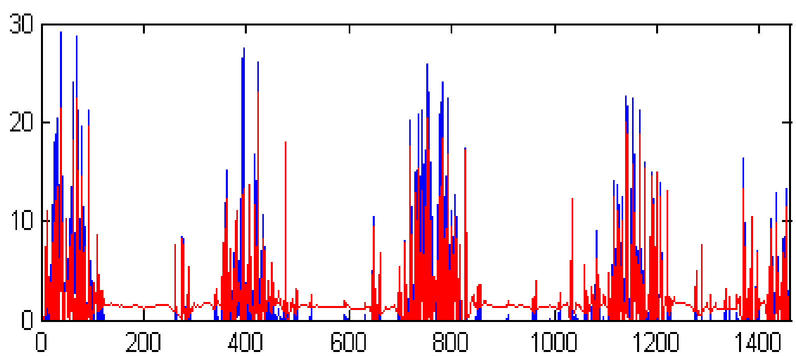

Figure 2. Rainfall reconstruction from NDVI trend and rain-gauge detail signal, using the inverse wavelet transform (blue=gauge; red=reconstructed)

Lag time

Table 3 shows the lag times estimated for S_{Rain} and S_{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 residence time across most ecozones. Only three of the ecozones showed better fit for other periods: Mañazo, Azángaro and Macusani (T= 365 d for the first two and 91 d for the latter).

Table 3. Determination coefficient for reconstructed versus gauged rainfall data and lag time for different sites in the high Andean plateau

Station	R^2	T (days)	Lag (T) (days)	MAE	Bias
Mazo Cruz	0.82	121	56	1.46	0.86
Mañazo	0.83	365	47	1.68	0.86
Huaraya Moho	0.85	121	86	0.9	0.84
Huancané	0.87	121	74	1.40	0.85
Azángaro	0.85	365	19	1.45	0.84
Macusani	0.91	91	84	1.01	0.87
Chuquibambilla	0.87	121	82	1.35	0.91
Desaguadero	0.82	121	57	2.12	0.84
Tahuaco Yunguyo	0.81	121	43	1.7	0.9
Ilave	0.81	121	76	1.93	0.9

MAE=relative mean absolute error.

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

Conclusions

In this paper we showed a new reconstruction tool to generate daily rainfall from NDVI data, with the support of the Wavelet Transform, that maintain the same distributional properties of the measured events. The results obtained for the highly variable Andean highlands were superior to similar data 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.

The methodology described in this paper is suitable for interpolating daily rainfall from gauge measurement in specific points in space to larger areas. This can be accomplished by defining extrapolation domains for the stations and the support of NDVI measurements within the extrapolation boundaries, an investigation being conducted in our laboratory.

References

Brunsell, N.A.; Young, C.B. 2008. Land surface response to precipitation events using MODIS and NEXRAD data. Int J Remote Sens., 29(7), 1965-1982.

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. Int J Remote Sens., 29(14), 4097-4110.

Eklundh, L. 1998. Estimating relations between AVHRR NDVI and rainfall in East Africa at 10-day and monthly time scales, Int J Remote Sens., 19(03), 563-570.

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 Sens. Environ. 50(2), 121-133.

Foufoula-Georgiou, E.; Kumar, P. (eds.). 1994. Wavelets in geophysics. Academic Press, p. 373.

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

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. Int J Remote Sens., 26(11), 2281-2296.

Kawabata, A.; Ichii, K.; Yamaguchi, Y. 2001. Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. Int J Remote Sens., 22(7), 1377-1382.

Mallat, S. 1999. A Wavelet Tour of Signal Processing. Academic Press; 2nd edition, p. 637.

Martiny, N.; Camberlin, P.; Richard, Y.; Philippon, N. 2006. Compared regimes of NDVI and rainfall in semi-arid regions of Africa. Int J Remote Sens., 27(23), 5201-5223.

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 Sens. Environ. 50(2), 107-120.

Nicholson, C.F.; Lee, D.R.; Boisvert, R.N.; Blake, C.I.; Urbina, I.C. 1994. An optimization model of the dual purpose cattle production system in the humid lowlands of Venezuela. Agric. Syst., 46(3), 311-334.

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.; Lare, A.R., 1990. A climatonomic description of the surface energy balance in the central Sahel. Part II: The evapoclimatonomy submodel. J Appl. Meteorol., 29(2), 138-146.

Pipes, L.A.; Harvill, L.R. 1971. Applied Mathematics for Engineers and Physicists. Singapore, McGraw-Hill; 3rd ed., 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: Examples of research problems, approaches and partnerships in action in the CGIAR, Rome, Italy, pp 111-122.

Richard, Y.; Poccard, I. 1998. A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. Int J Remote Sens., 19(15), 2907-2920.

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, p 27.

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.