WAVELET-BASED SPATIO-TEMPORAL FUSION OF OBSERVED RAINFALL WITH NDVI IN SRI LANKA

Yann Chemin

International Water Management Institute, Pelawatte, Sri Lanka

Keywords: Wavelet, Fourier, fusion, spatio-temporal, NDVI, MODIS, rainfall, daily, Sri Lanka

Abstract: Availability of rainfall time-series is limited in many parts of the World, and the continuity of such records is variable. This research endeavors to extend actual daily rainfall observations to ungauged areas, taking into account events of rainfall as well as cumulative total daily rainfall, over a period of 11 years. Results show that rainfall events histograms can be reconstructed, and that total cumulative rainfall is estimated with 85% accuracy, using a surrounding network of rain gauges at 30-50 Km of distance from the point of study. This research can strengthen various types of research and applications such as ungauged basins research, regional climate modeling, food security early warning systems, agricultural insurance systems, etc.

1. Introduction

Continuous long and reliable rainfall records at any location are the prerequisite for informed water management decisions and water resources planning: they tell how much water actually is available and how this availability varies with time. Ground observational network on rainfall is inevitably limited to a few locations, and is very sparse in many regions of the world. It is unlikely that traditional methods of rainfall data acquisition can help improve the situation in the future, but continuously advancing satellite-based approaches can.

To augment rainfall data availability, various forms of re-analyses are used, and subsequently, the synthetic rainfall data sets are fed into hydrological models. An initial review of remote sensing for hydrology is found in Schultz (1988) who showed (among other cases) how to use radar rainfall measurements in a distributed hydrological model for real-time flood forecasting. Later reviews, as found in Engman and Gurney (1991), Pietroniro and Prowse (2002), Schmugge et al (2002), Pietroniro and Leconte (2005), Wagner et al (2009). cover a range of possibilities of detection, classification, evaluation and management of water in its various phases using the large palette of space-borne sensors.

One possible avenue to significantly enhance rainfall data availability is to derive rainfall from remotely sensed (RS) vegetation data. Various RS products are available in public domain since late 1970s. Their spatial and temporal resolutions are orders of magnitude better than any ground observation network will ever be able to achieve. The condition of the ground vegetation, reflected by the widely used Normalized Difference Vegetation Index (NDVI) is related to the antecedent precipitation at any particular point and many attempts in the past were made to establish an explicit relationship between RS-derived vegetation indices and rainfall. Initial attempts to compare NDVI response to rainfall can be found in the research of Nicholson et al. (1990) who found a 60 days' time lag between the Vegetation Index and rainfall patterns and stronger correlation of inter-annual patterns in the Sahel than in East Africa. They also found that annual integration of NDVI correlates well with total rainfall. Later, Santos and Negri (1997) found that Vegetation Index and rainfall are uncorrelated in the Amazon Basin. They subsequently found that in the Northern part of Brazil, under a drier climate, correlation is very clear, and even sensitive to rainfall regimes less than 100 mm/month. Immerzeel et al (2005) used a signal processing method to correlate total rainfall to signal parameters in Tibet and found highest correlation of Vegetation Index with pre-monsoonal rains. They conclude that signal processing methods generate stronger correlations than standard time-series analysis.

The temporal daily rainfall reconstruction technique that this paper proposes to extend in the combined dimension of space and time is found initially in Yarleque et al (2005a, 2005b, 2006, 2007a, 2007b) and later the same in Quiroz et al (2011). The methodology found in these articles was restricted to the temporal dimension, using a signal processing technique called wavelet decomposition. It decomposed the rainfall signal in the high frequency domain and satellite vegetation index in the low frequency domain. Once those frequency signals were isolated, they were recomposed together creating a new merged signal of vegetation and rainfall that would fit temporally the initial rainfall signal from the meteorological station. This effectively fused heterogeneous physical sources to generate a mixed source synthetic time-series.

We propose to expand this technique by a wavelet-based spatio-temporal fusion technique, involving the interpolation of wavelet decomposition levels from daily rainfall stations and the recomposition with individual NDVI pixels. The original

method used wavelet and Fourier transforms within the temporal dimension. We used a convolution approach within the spectral dimension of wavelet levels to spatially interpolate the different locations of observed rainfall. Recomposition of mixed sources, rainfall and MODIS wavelet spectral dimension parameters produced a spatio-temporal dataset of daily interpolated rainfall over our study area, the island of Sri Lanka, for the period 2000-2010.

Objectives

This research aims to expand the initial results of a new spatio-temporal fusion algorithm dedicated to extend the time-series of rainfall events from meteorological stations and rain gauges across areas where there are no rainfall time-series. This is achieved by "encoding" rainfall time-series available and by spatially bridging these encodings together. Spatial and temporal Vegetation Index from satellite imagery is also encoded, and both types of encoding (i.e. rainfall and vegetation) are fused together to synthetically recreate rainfall time-series were none was before observed.

2. **METHODOLOGY**

Original daily rainfall data as received from the meteorological department of Sri Lanka had 106 meteorological stations or rain gauge data from year 2000 to year 2010 included. Parsing the files was complex and two sets of outputs were generated, their spatial coordinates in one file for use in the interpolation at the bottom of figure 1 and the rainfall data in a separated set of temporal files as used at the top of figure 1 (right side).
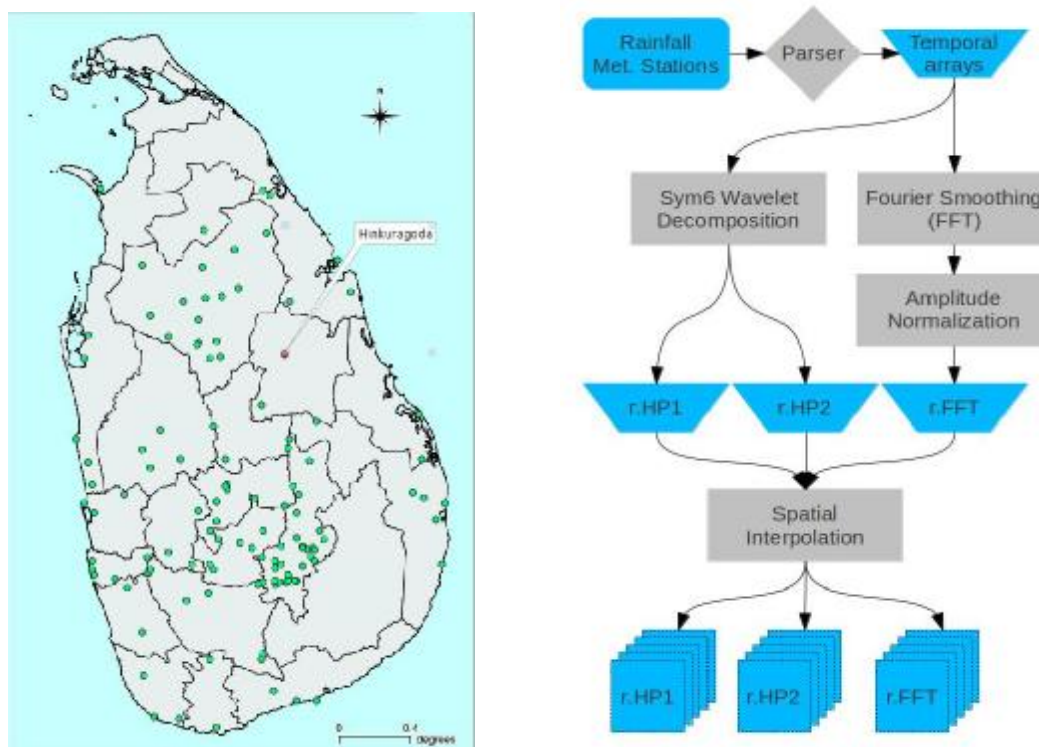

Figure 1: Rainfall ground data location and the decomposition/interpolation method
Note: Hinkuragoda meteorological station will be used in the experiments

The rainfall temporal arrays are decomposed down to level 2 using sym6 wavelets signatures. Only the high-pass information is retained for later, level 1 is named *r.HP1* and level 2 is named *r.HP2* in figure 1, a temporal signature of the two high-pass levels is seen in (Figure 3). Additionally, the temporal arrays are smoothed by a Fast Fourier Transform (FFT) with less than 10 harmonics, an amplitude normalization is applied to range the output signal from 0.0 to 1.0. This signal (*r.FFT*) will be used in the second part of the methodology (figure 2).

Using the geolocation information extracted from the Meteorological Department of Sri Lanka daily rainfall dataset (Fig. 1 left side), it is now possible to interpolate by convolution the wavelet decomposition high-pass levels (*r.HP1* and *r.HP2*)

every day for eleven years, thus generating ~4018/[HPlevelNo^2] images of Sri Lanka for each level (Figure 2).
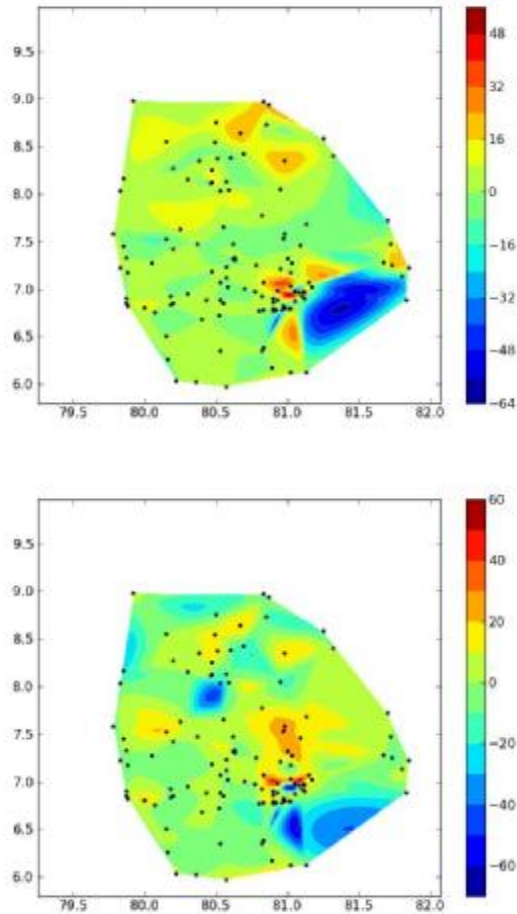

Figure 2: spatial interpolation of r.HP1 and r.HP2 at a given frequency slice

This will conclude the first part of the methodology (figure 1), whereby the daily rainfall data is temporally decomposed and its high-pass levels are interpolated daily, so as to generate a daily high-frequency wavelet decomposed rainfall footprint for all Sri Lanka from recorded ground data.

The second part of the methodology (figure 4) is using the Normalized Difference Vegetation Index (NDVI) from the MODIS sensor on Terra satellite platform at a 16-days composite period for consistency of BRDF correction and also because only low-pass information are going to be used from this dataset.
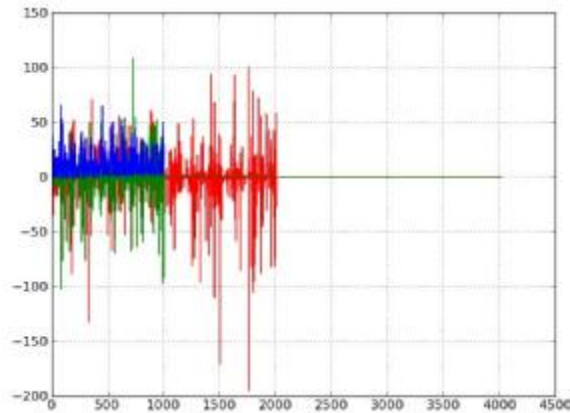
Figure 3: Frequency data: Red = r.HP1, Green = r.HP2, Blue = n.LP2
(note the ½ and ¼ signal dimensions from 4018)

The NDVI data (figure 4) is vacuumed from NODATA stamps with a moving average which is passed temporally to fill in these stamps. For each pixel, a temporal Fourier smoothing is applied with a low harmonic number input in a FFT as discussed in Yarleque et al (2006).

To fasten the matching with *r.FFT*, an amplitude normalization of NDVI data with bounds of 0.0 and 1.0 is also applied and named *n.FFT (*figure 4*)*. Both *r.FFT* and *n.FFT* are introduced into a phase-shift absorption type algorithm that is iteratively changing the phase delay of *n.FFT*, eventually reducing the phase difference between both of them. The phase-shift value iteratively identified is then retrofitted into the *n.FFT* signal. Nicholson et al. (1990) already found a 60 days' time lag between NDVI and rainfall patterns in Africa, this research found similar range of shifts (~65 days).
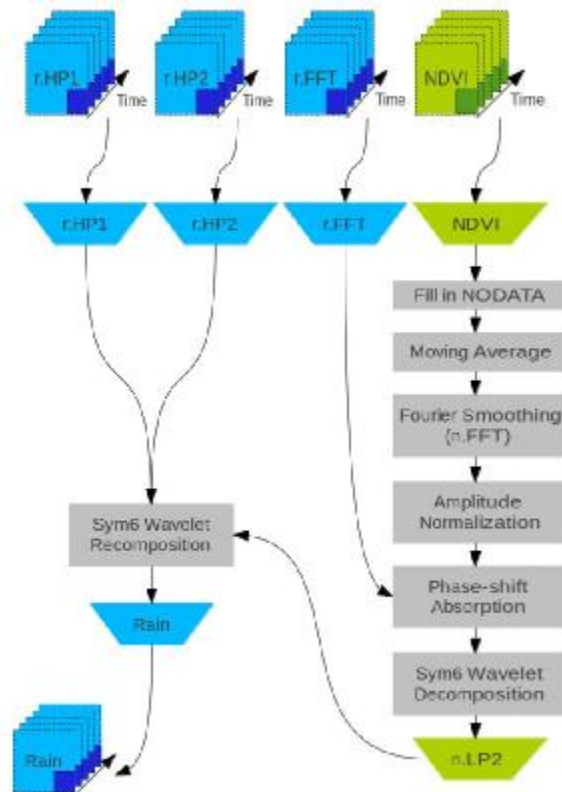

Figure 4: spatio-temporal fusion of frequency rainfall and NDVI data

The *n.FFT* is then decomposed down to level 2 using sym6 wavelets signatures for each pixel location in the stack of NDVI images (Figure 3). Only the low-pass at level 2 is kept (*n.LP2*) for use in the temporal wavelet recomposition using the same sym6 family member for each of the land pixels available in Sri Lanka.

Merging back each temporal set into the stack of contiguous temporal pixels permits daily images of Sri Lanka's rainfall to be regenerated. For a detailed desrcription of the methods used, see Appendix 2. Results were calculated for around 94,000 pixels (covering Sri Lanka at 1x1 km) propagated in time 4018 days (2000 – 2010).

3. **RESULTS**

Experiment 1

The first experiment aims at proving that the rainfall quantity found at a random meteorological station is conserved throughout the algorithmic transformations. The meteorological station is Hingurakoda (see Fig. 1) located in North Central Sri Lanka (lat: 8.05° ; long: 80.95°).
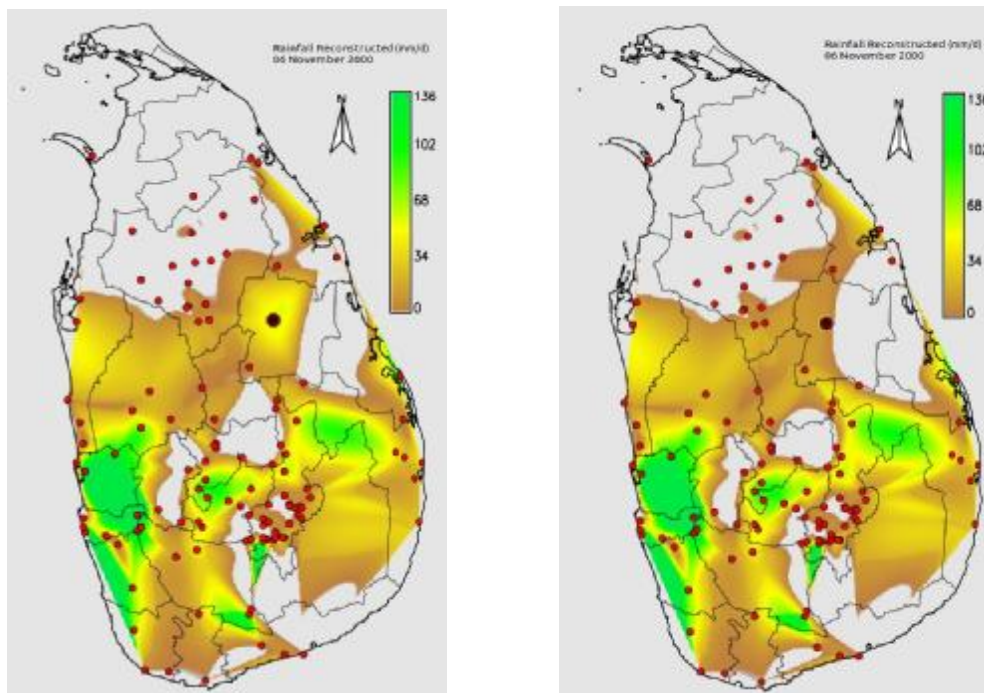

Fig 5: Sri Lanka map of reconstructed rainfall exp.1 and exp.2 (06/Nov/2000)
Note: the black circle indicates Hinkuragoda meteorological station

A first estimation of the results can be seen in Fig. 5 for a given day (6 November 2000). Results for the single pixel at the location of the meteorological station at Hinkuragoda (Fig. 1) are shown in Figure 6.
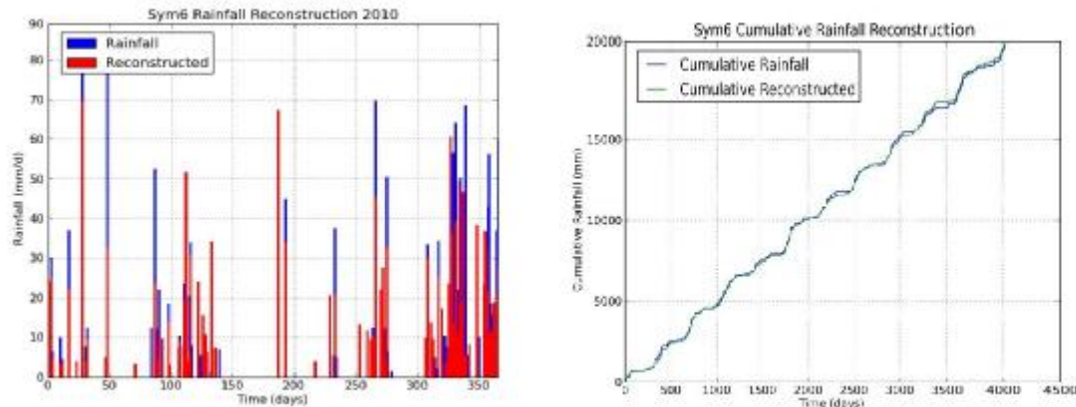
Fig 6: Hingurakoda meteorological station a) observed daily rainfall and its reconstruction for the year 2010 and b) cumulative daily rainfall and its reconstruction for the period 2000 - 2010.

Figure 6a shows the detail of reconstruction of individual rainfall events for a single year (2010); Figure 6b compares results for cumulative daily rainfall over 11 years (2000 – 2010)

Experiment 2

To reinforce the validation procedure, calculations were repeated with the Hingurakoda meteorological station was removed. The aim was to assess the reconstruction of a hypothetical rainfall gauge without initial knowledge of its temporal variability, except what is available from neighboring stations.
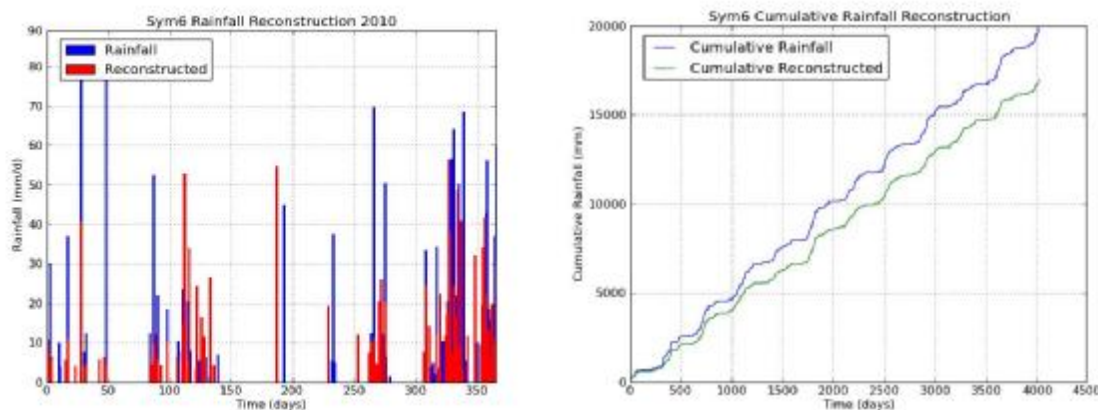

Figure 7: Rainfall observed Vs reconstructed (without original rainfall data) at Hingurakoda a) observed daily rainfall and its reconstruction for the year 2010 and b) cumulative daily rainfall and its reconstruction for the period 2000 - 2010.

The results of this second experiment (Figure 7a) found that some of the biggest events of rainfall are more underestimated compared to the first experiment. From the cumulative rainfall reconstruction (Figure 7 right), a constant deviation of reconstruction from real data was observed. The total difference yielded 85% accuracy of cumulative rainfall reconstruction in absence of the original data from the Hingurakoda meteorological station for a 11 years daily rainfall period.

From Figure 7b one can also observe a linear relationship between the original and estimated data with very few days where the overall pattern of cumulative rainfall are different. Thus, the contribution of rainfall events from neighbouring meteorological stations has a strong importance in recovering this pseudo-unknown rainfall information from the Hingurakoda meteorological station. In this context, the homogeneity of the climatic zone should be considered when reducing the number of rainfall gauges density to apply such method. The closest rainfall station is 33Km away in the North direction, then comes 35Km SW, 40Km W, 45Km SE and 55Km NE.
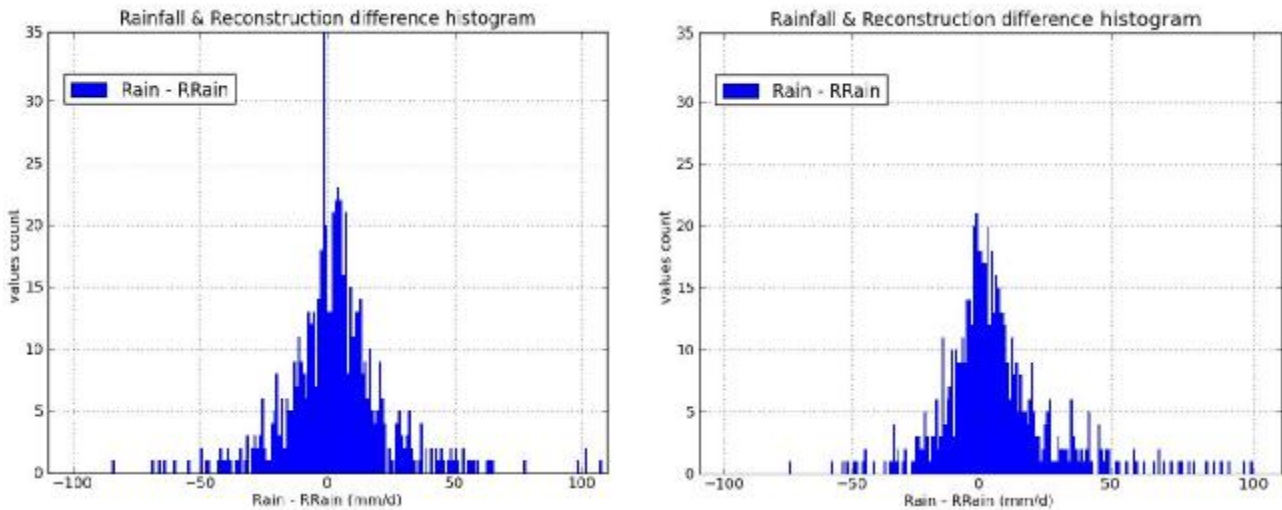
Figure 8: Hinkuragoda difference histograms for a) Experiment 1 and b) Experiment 2
Note: Removed all data where the difference equals 0 mm/d

Figure 8a and b show histograms of the difference between original rain gauges data and reconstructed rainfall data, for Experiment 1 and Experiment 2 respectively. In the y-axis is the number of days where the difference is recorded and in the x-axis, centered to zero value, is the difference in mm/day of rainfall values. The overall shape and limits (+/- 50 mm/d) are conserved throughout both experiments, but both histograms are slightly shifted towards the positive side. Differences in single large (events between +50 and +100 mm/d seen in Figure 8a) are in themselves the source of the reduction in accuracy found in Figure 7b.

Table 1: Descriptive statistics of the difference histograms

	Exp1*	Exp2*	Exp1	Exp2
Sample size	627	618	4018	4018
Minimum			-106.39	-112.23
Maximum			107.79	110.20
Arithmetic mean	3.08	6.08	0.02	0.72
Unbiased variance			120.87	131.86
Biased skewness	0.61	0.91	0.58	1.63
Biased kurtosis	4.40	3.07	21.20	21.03
* Removed diff=0				

Table 1 features basic descriptive statistics run on the histograms with (Exp1 and Exp2) and without (Exp1* and Exp2*) the zero value samples, translating in a large difference in samples and changes in descriptive statistics. The arithmetic mean is slightly positive, more for the experiment 2 than 1 in both cases. Variance is high, yet similar in both experiments. Skewness is different when using all data. Kurtosis does not shown significant changes in all cases.

Figure 7a shows a non-reconstructed rainfall event at around day 180. This can be attributed either as missing or as a lag between observed and reconstructed event, the lag maybe of a few days. Either way, this is worth investigating as events of neighboring stations influence largely the reconstruction. A preferential direction of interpolation following the rainfall path maybe investigated. This can be introduced by finding the rainfall events motion vectors using remote sensing rainfall intensity datasets such as the Tropical Rainfall Measurement Mission (NASA, 2012). This will also provide background information (rainfall intensity) to remove dependency from meteorological station measurements' density.

4. **CONCLUSIONS**

By extending the already existing wavelet temporal relationship between NDVI and daily rainfall into the spatio-temporal dimensions, this research aimed at exploiting the combined benefits of the spatial information in NDVI images and the temporal information of rainfall. Eleven years of cumulative rainfall were reconstructed without its original meteorological

station data with an accuracy of 85.34%. Descriptive statistics of the difference histograms for experiment 1 (using Hinkuragoda meteorological station data as input to the algorithm) and for experiment 2 (without the Hinkuragoda data) do not show much change.

Additional work will also involve scaling the method to larger spatio-temporal dimensions, implying the adaptation of the methodology into High-Performance Computing frameworks as found in Akhter et al. (2006, 2007, 2008), Chemin (2012).

Finally, direct impact of this research can strengthen ungauged basins research, regional climate modeling, food security early warning systems, agricultural insurance systems, etc.

Acknowledgements

References:

Akhter, S., Aida, K., Chemin, Y., 2008. Asynchronous Grid LMF Processing of Vegetation Index for Asia. International Journal of GeoInformatics. (4)4: 39-45pp.

Akhter, S., Jangjaimon I., Chemin Y., Uthayopas P., Honda, K., 2006. Development of a GRIDRPC Tool for Satellite Images Parallel Data Assimilation in Agricultural Monitoring. International Journal of GeoInformatics, (2)3: 29-33pp..

Akhter, S., Sarkar, I., Rabbany, K.G., Akter, N., Akhter, S., Chemin, Y., Honda, K., 2007. Adapting the LMF Temporal Splining Procedure From Serial to MPI/Linux Clusters. Journal of Computer Science 3(3): 130-133pp.

Chemin, Y., 2012. A distributed benchmarking framework for Actual ET models. In Evapotranspiration – Remote Sensing and Modeling, ISBN 978-953-307-216-6, January 2012, Chapter 19, 421-436pp.

Daubechies, I., 1992. Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), 61:341pp.

Engman, E.T., and Gurney, R.J., 1991. Remote Sensing in Hydrology. Chapman and Hall, New York. 1991. 225p.

Fourier, M., 1808. Memoire sur la propagation de la chaleur dans les corps solides. Nouveau Bulletin des Sciences de la Societe Philomathique de Paris, 1(6):102-116pp. (A scanned copy is available on the Internet at: http://gallica.bnf.fr/ark:/12148/bpt6k33707/f220.image.r=Oeuvres+de+Fourier.langFR)

Haar, A., 1910. "Zur Theorie der orthogonalen Funktionensysteme". Mathematische Annalen, 69(3): 331-371pp. doi:10.1007/BF01456326

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.

NASA, 2012. Website of the Tropical Rainfall Measurement Mission. Goddard Space Flight Center, National Atmospheric and Oceanic Administration. http://trmm.gsfc.nasa.gov/

Nicholson, S.E., Davenport, M.L., and 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.

Osgood, B., 2011. Lecture notes for EE261: The Fourier Transform and its Applications. Electrical Engineering Department, Stanford University. 432pp. Available online.

Pietroniro, A., and Prowse, T.D., 2002. Applications of Remote Sensing in hydrology. Hydrological Processes, 16:1537-1541.

Pietroniro, A., and Leconte, R., 2005. A review of Canadian remote sensing and hydrology: 1999-2003. Hydrological processes, 19:285-301.

Quiroz, R.A., Yarlequé, C., Posadas, A., Mares, V., Immerzeel, W.W., 2011. Improving daily rainfall estimation from NDVI using wavelet transform. Environmental Modeling and Software, 26:201-209.

Santos, P., and Negri, A.J., 1997. A comparison of the Normalized Difference Vegetation Index and Rainfall for the Amazon and Northeastern Brazil. Journal of Applied Meteorology, 36(7):958-965.

Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J., Rango, A., 2002. Remote Sensing in Hydrology. Advances in Water Resources, 25:1367-1385.

Schultz, G.A., 1988. Remote sensing in hydrology, Journal of Hydrology, 100(1-3):239-265.

Wagner, W., Verhoest, N.E.C., Ludwig, R., Tedesco, M., 2009. Remote sensing in hydrological sciences. Hydrological Earth Systems Sciences, 13:813-817.

Yarlequé, C., Posadas, A., Quiroz, R., 2005a. Reconstrucción de lluvia en series de tiempo, en el Altiplano peruano mediante Transformadas de Wavelet con dos niveles de descomposición. XV SIMPOSIO PERUANO DE FISICA. Sociedad Peruana de Física y la Facultad de Ciencias Físicas de la Universidad Nacional Mayor de San Marcos.

Yarlequé, C., Posadas, A., Quiroz, R., 2005b. Reconstrucción de Lluvia en series de Tiempo, en el Altiplano Peruano mediante Transformada de Wavelet con dos niveles de descomposición. IX Simposio Nacional de Estudiantes de Física y la V Jornada Nacional de Estudiantes de Física (IX SNEF Y V JNEF), Ica, Perú.

Yarlequé, C., Posadas, A., Quiroz., R., 2006. Estimating Rainfall Precipitation from Normalized Vegetation Index Using Wavelet Transform. International Congress on Mathematical Physics, 6-11 August. INPA, Rio de Janeiro, Brazil

Yarlequé, C., Posadas, A., Quiroz, R., 2007a. Reconstruccion de datos de precipitation pluvial en series de tiempo mediante transformadas de wavelet con dos niveles de descomposicion. Centro International de la Papa, Division de Manejo de Recursos Naturales, Documento de Trabajo No 2007-2. ISBN 978-92-9060-315-3

Yarlequé, C., Posadas, A., Quiroz, R., 2007b. Estimating rainfall from the normalized difference vegetation index using wavelet transform. Presented at ECI07 Encuentro Científico Internacional "Alberto Cazorla Talleri", Enero 2-5, 2007, Lima, Perú.

Appendix 1: **Theoretical background**

Dimensionality is the basis of most of the research on this Earth. Space and time dimensions, most ubiquitous to human senses, are but a basis to understand spatial and temporal events. The additional frequency dimension is briefly expanded below by two analytical bodies of research represented by the Fourier (Initial article: Fourier, 1808) and the Wavelet (Initial article: Haar, 1910) fields of sciences.

There are countless sources of Fourier Series and Transforms, we have used information from Osgood (2011). The underlying concept in Fourier analysis is that real data has intertwined repeating patterns with a quantifiable number of periods. A Fourier analysis will analyze the contribution of certain harmonics (periods) to the full signal, and summarize the information as fingerprints of the harmonics in a dimension of its own, the spectral dimension. A function t is periodic if for all T>0:

$$f(t+T) = f(t) \tag{1}$$

A finite Fourier series is a Fourier-type trigonometric sum of a periodic signal of period 1 can be expanded by:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos(2\pi n t) + b_n \sin(2\pi n t)) = \sum_{n=-N}^{N} c_n e^{2\pi i n t} \tag{2}$$

with c_n, the Fourier coefficient being a complex number satisfying the following:

$$c_{-n} = \bar{c}_n \quad also\ for\ n = 0 \quad we\ have\ c_0 = \bar{c}_0 = \frac{a_0}{2} \tag{3}$$

It turns out that one can go back and forth from f(t) and C_n. Let f(t) be in $L^2([0,1])$ and let:

$$c_n = \hat{f}(n) = \int_0^1 e^{-2\pi i n t} f(t) dt, \quad n \tag{4}$$

with C_n be its Fourier coefficients, within the spectral dimension, with the assumption of convergence of the partial sums. In this article, we will use the conversion of a time-series from real to spectral and from its n (n<10) first levels (harmonics) back to a degraded data in real dimension. This is called Fourier smoothing.

Haar (1910), rediscovered and expanded by Daubechies (1992) developed an understanding that periodicity of a given event may have a non-constant return of event and actually found an analytical way to quantify those changes in the dimensionality observed (usually time or space) as it unrolls. The complementarity of Wavelet analysis to Fourier analysis is that Fourier will not "remember" the uniqueness of some event in the real dimension (space or time), but will proportionally return its event at all frequencies involved in the analysis. The result will be the appearance of traces of the same event more than once, after Fourier will have returned the information from the frequency domain. Wavelet-based analysis will be able to change the parameterization of the frequency analysis in order to retain unique events, while keeping recurring events fully developed back into the real dimension. This is the major difference between Fourier and Wavelet analyses: Fourier excels at stationary events, Wavelets at non-stationary ones.

To analyze such changes and retain critical information throughout the dimensional changes, wavelets translate a window of analysis at a given scale of observation and store that information in the frequency domain. Thus wavelet analysis is a multi-location/resolution analysis of a signal. The Continuous Wavelet Transform (CWT) at a given location translation on the signal τ and amount of the signal analyzed at a given time (scale) s is given by the inner product of two functions. One function, $f(t)$ will be the signal we want to analyze with another function $\psi_{\tau,s}(t)$. The parameters of the functional ψ are explained below.

$$CWT_f^\psi(\tau,s) = \Psi_f^\psi(\tau,s) = \frac{1}{\sqrt{s}}\int f(t)\,\psi\left(\frac{t-\tau}{s}\right)dt$$

(5)

The signal x is said to be transformed by the transforming function ψ at $t-\tau$ location-translation on the signal and on $1/s$ size of the same signal starting on that location-translation point. In turn, s is a frequency and its unit can be [1/time] if the signal is running on a temporal dimension. The transforming function ψ, also called Mother Wavelet can be of many sorts, literally called families of Mother Wavelets, according to their mathematical properties. In this research, we are going to use the same as in Quiroz et al (2011), the symlet at level 6.