Entropy Analysis of Real Time Series

Susana Vanlesberg and Mario Silber

Abstract: In the paper Entropy Analysis of Synthetic Time Series (Silber, M. & Vanlesberg, S.) properties of the entropy of the information provided by synthetic series with different characteristics were demonstrated. In this opportunity, observed time series of precipitation on Mesopotamian and central region of Argentina were analyzed, with the following objectives: to verify the properties shown in the cases described in the theoretical work mentioned above, the entropy of geographically related information with its variability and, in some way, to relate the empirical entropy (independent of the parameters of the theoretical probability distribution of the variable) with some of the variables included in the water balance.

Riassunto: *Nel lavoro precedente Entropy Analysis of Synthetic Time Series (Silber, M. & Vanlesberg, S.) sono state dimostrate le proprietà dell'entropia sulla base delle informazioni provenienti da serie di dati sintetiche con caratteristiche differenti. In questo lavoro sono state analizzate le serie temporali di precipitazioni misurate in Mesopotamia e nella regione centrale dell'Argentina. Il lavoro ha avuto le seguenti finalità: verificare le proprietà riscontrate nell'analisi delle serie sintetiche, l'entropia delle informazioni locali con la sua variabilità e correlare l'entropia empirica (indipendente dai parametri della distribuzione di probabilità teorica della variabile) con alcune delle variabili che influenzano il bilancio idrico.*

Keywords: sample entropy, water balance, precipitation

Susana VANLESBERG
Mario SILBER ✉
Facultad de Ingenieria y Ciencias Hidricas
Universidad Nacional del Litoral
Ciuciad Universitaria
CC 217 - Ruta Nacional 168 - Km 472,4
(3000) - Santa Fe - Rep. Argentina
Tel: (54) (342) 457 5243/44 - Int. 160
Fax: (54) (342) 457 5224

Susana VANLESBERG
susvan@gmail.com

Mario SILBER
silber.mario@gmail.com

Received: 27 november 2012 / Accepted: 11 december 2012
Published online: 30 december 2012

Introduction

In the planning of water resources knowledge of both the quantity as well as their availability in time becomes a critical decision factor. The importance of rain and other hydroclimatic variables in the planning and climate studies is particularly important especially in subsistence agricultural activities in the economy of many countries including Argentina.

Changes in climate may affect the hydrological cycle elements such as the distribution of precipitation and/or its concentration. In recent times some phenomena, especially the occurrence of severe events occurred in parts of the province of Santa Fe, Argentina (rain recorded on 10/24/2011: 95.7mm, and 150.4mm accumulated since 10/22/2011 to 10/25/2011) suggest that the characteristics of this variables hould be the subject of more detailed and specific analysis. This is referred especially to the patterns of spatial and temporal distribution, as well as their general and particular behavior al characteristics.

Much has been said and studied about the worldwide climate change, so to study or analyze its possible causes again may be redundant. So the important thing is to focus on those variables, such as precipitation, that are of great importance in the development of agricultural activities in a region and the daily lives of the inhabitants of that region, since the activities are hindered by the occurrence of these concentrates and severe phenomena, both drought and rainfall

The annual and monthly variability of precipitation in different areas of the Pampas was investigated by Castañeda and Barros (1994), Minettia and Vargas (1997), Penalba and Vargas(1996), Rusticucci

DOI 10.4409/Am-049-12-0050

and Penalba(2000), Krepper et al.(1989), and Krepper and Sequeira (1998), among others.

Several recent studies reported large increases in precipitation and flows through the United States in the second half of the 20th century (eg, Karl and Knight, 1998; Lettenmaier et al, 1994) and low flows (LinsandSlack,1999,Douglas et al, 2000, Small et al., 2006). There was a significant increase in precipitation in a large number of stations (Lettenmaier et al., 1994), at the watershed level (Small et al., 2006), at regional level (Douglas et al, 2000) and on the national averages (Karl and Knight, 1998).

Another reason why the precipitation is analyzedis because it is one of the key variables on issues related to agricultural production, which is the basis of the economy of our country.

The variability of precipitation and its participation in the hydrological cycle and energy cycles are important for understanding the behavior and changes in the Earth's climate system.

With these underlying concepts is that the concept of entropy has been used to analyze seasonal and annual time series of precipitation, as well as the distribution of the amount of annual and seasonal precipitation and the number of days of annual and seasonal rainfall.

The time series considered in the paper are: annual series, which consists of the accumulated monthly rainfall on every month for the entire study period, seasonal time series, consisting of individual seasons each year throughout the length of the original series, and monthly time series, consisting of individual months of each year for the entire period of data.

By estimating the disorder in the annual and monthly distribution of rainfall is possible to assess the availability of this resource, especially in areas where its presence is vital to the development of both human activities and agricultural production.

Study area and used data

Daily precipitation data from 13 monitoring stations belonging to the National Weather Service (NWS), the Communications Division of the prov. Santa Fe and the National Institute of Agricultural Technology (INTA) was used. These stations were selected in the study region considering the existence of complete sets of precipitation (over statistically consistent data) during the study period, between January 1971 and December 2012 (Fig. 1), being this the common period of data over the stations taken.

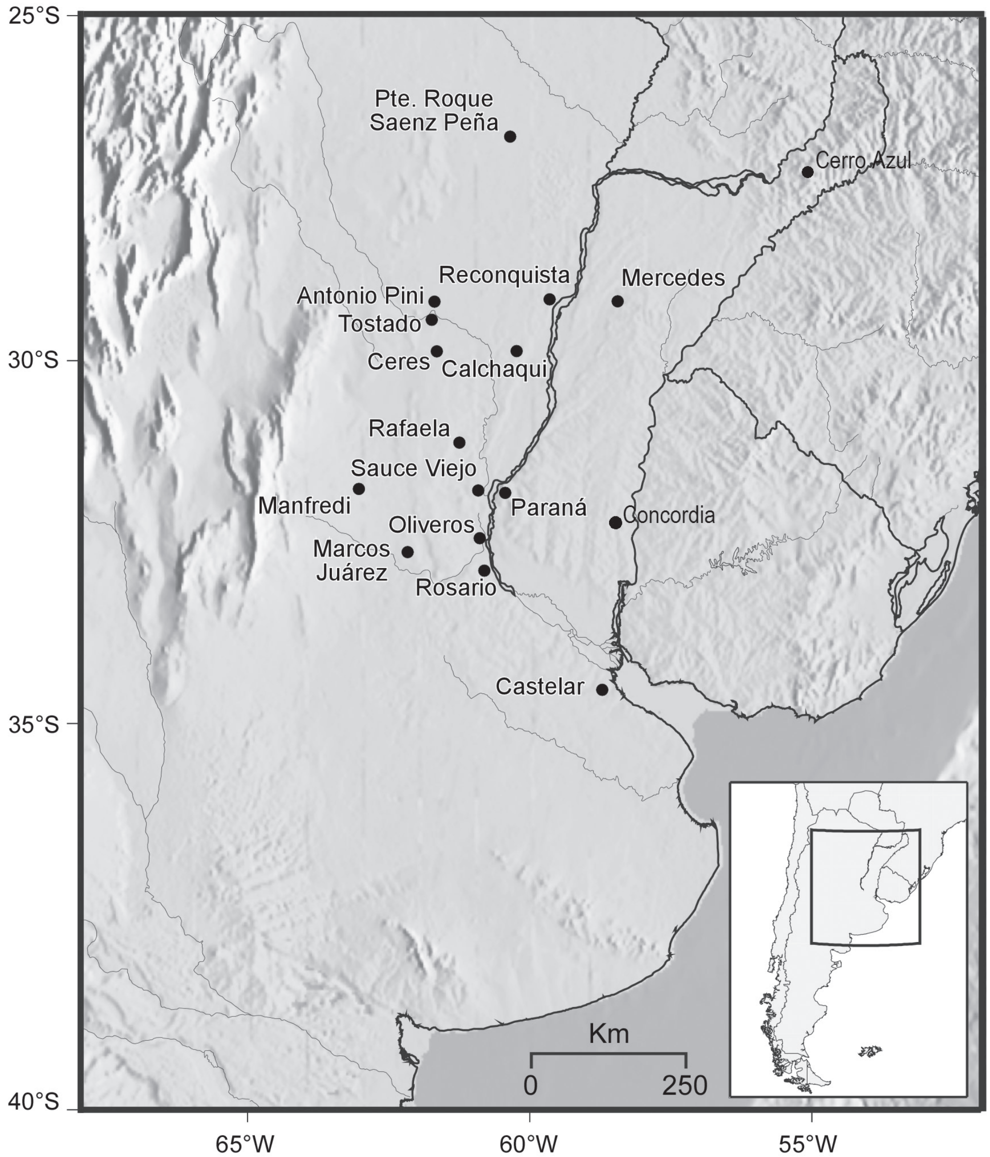

Fig. 1: *Study area*

Method

Analyzed cases

Case 1: Series drawn from a uniformly distributed population.

This is not a study case because the analyzed variable (precipitation) is not a uniformly distributed variable.

Case 2: Series drawn from normal populations with different coefficients of variability and described with uniform number of classes.

In this case different levels of aggregation were used in order to work under the assumption of normal distribution. The total annual precipitation at each measurement site (rainfall stations) and also annual quarterly totals for each site was obtained. Samples were described with the same number of bins (7) and the entropy of each of the samples according to this way was obtained (Figures 2, 3, 4, 5 and 6)

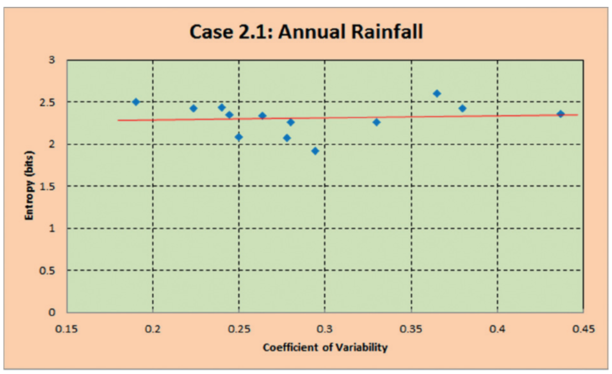

Fig. 2: *Case 2.1*

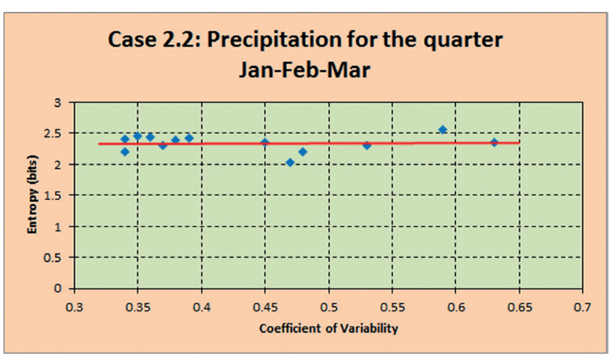

Fig. 3: *Case 2.2*

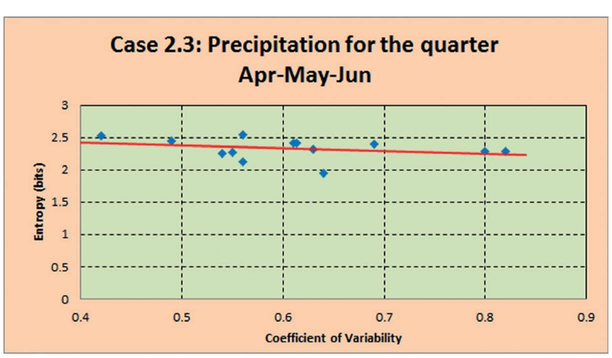

Fig. 4: *Case 2.3*

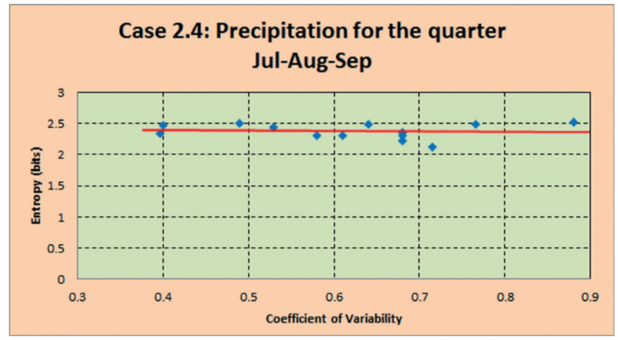

Fig. 5: *Case 2.4*

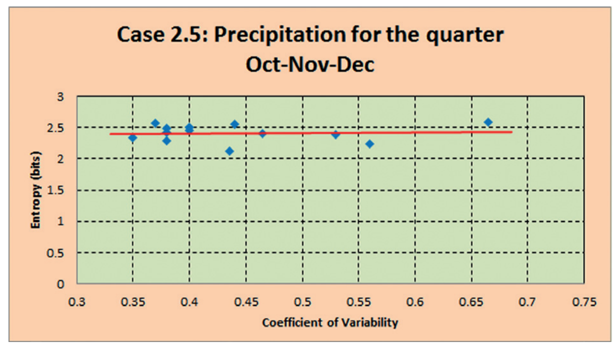

Fig. 6: *Case 2.5*

Case 3: Series drawn from normal populations, with different coefficients of variability, described with uniform class widths

In this case, the same level of aggregation was performed in order to work under the assumption of normal distribution. The sample was described with the same class width, and so the sample entropy was obtained (Figures 7, 8, 9, 10 and 11).

As was already demonstrated in the preceding paper, the entropy is a function of the variability of the sample. Therefore, by relating the entropy of the sample to the standard deviation the following relationships were obtained, confirming the theoretical hypothesis.

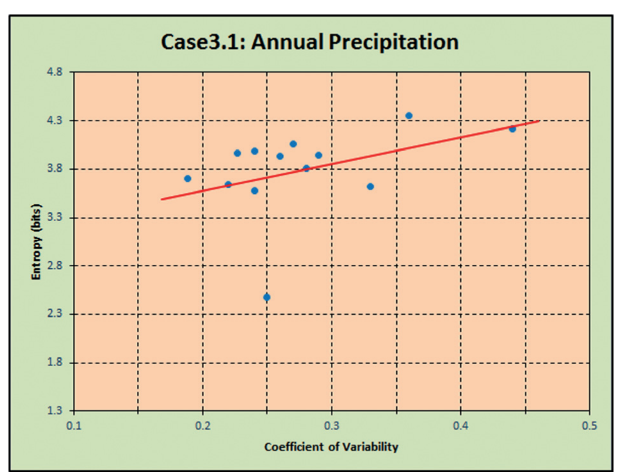

Fig.7 : *Case 3.1*

DOI 10.4409/Am-049-12-0050

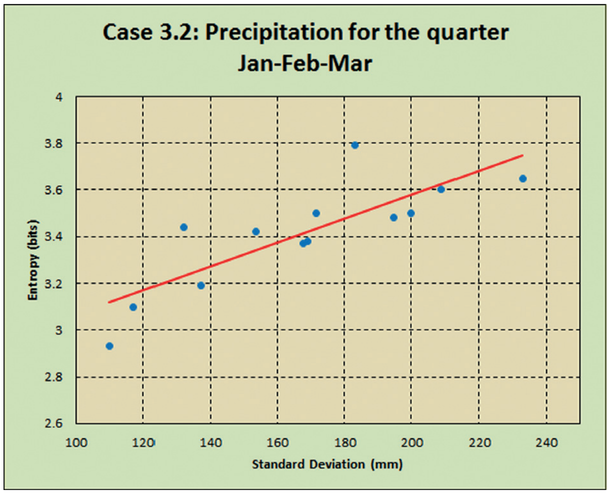

Fig.8 : *Case 3.2*

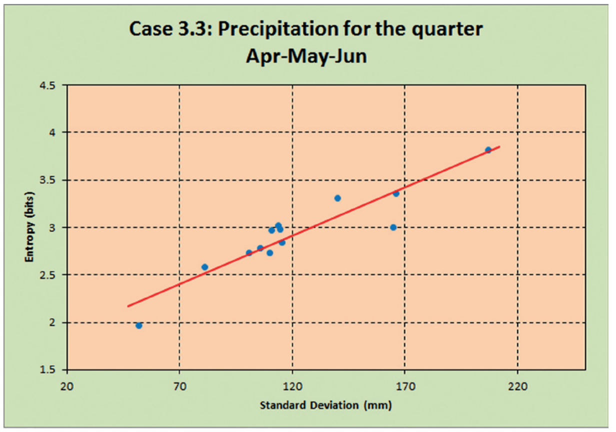

Fig.9 : *Case 3.3*

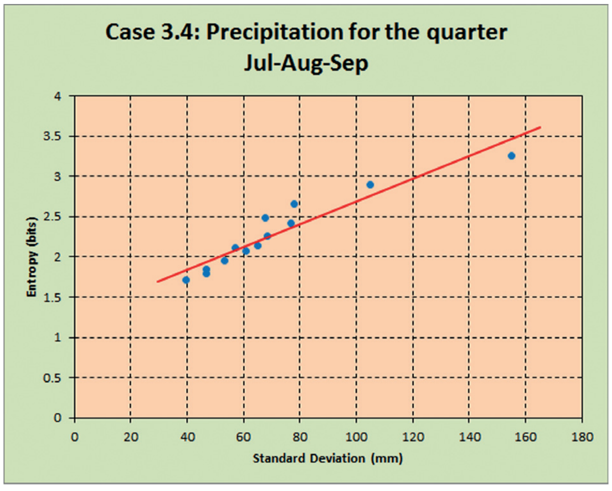

Fig.10 : *Case 3.4*

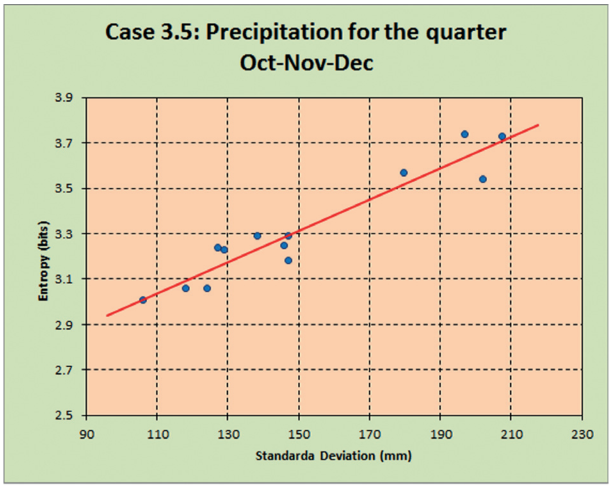

Fig.11: *Case 3.5*

Case 4: Series drawn from Gamma distributed populations, with different coefficients of skewness.

In the previous paper, in which the theoretical bases of the sample entropy were analyzed, many samples with different skewness coefficients were generated, as it is a function of the shape parameter of the Gamma distribution.

In the used daily precipitation series used the skewness coefficient was estimated and the samples was described using uniform class width (Figure 12).

Again what is postulated in the case of synthetically generated samples was checked: the entropy decreases when the skewness coefficient increases.

The rest of the cases raised in the reference paper will be the subject of analysis in a later publication because it characteristics deserve some special treatment due to the related variable: the flow.

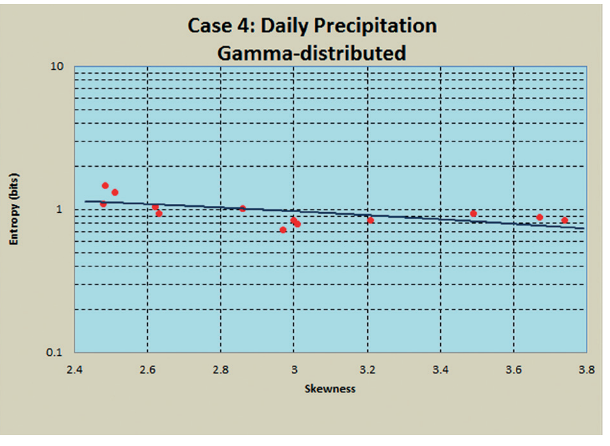

Fig.12 : *Case 4*

Conclusions

The detailed analysis of the entropy of synthetic time series showed some relationships between it and certain descriptive characteristics of samples generated.

Through the study of actual series of precipitation at different levels of aggregation in the study zone described, the above assumptions and conclusions was verified. This means that the sample entropy, evaluated at a point, can be taken as a characteristic of the sample independent of its probability distribution, which is not true when speaking about differential entropy.

The significant positive variation of entropy related to the variability of the sample is remarkable: the greater the variability, the larger the entropy. This is very important when relating the variability with water availability (excesses and deficits). Finding the relationship between the availability and entropy (which is nothing but a measure of the disorder of information) allow these results to define spatially the water balance without further calculations because information, at last, represents the variable and its characteristics.

REFERENCES

Ben-Naim Arieh (2007) Entropy demystified: the second law reduced to plain common sense, Ed. Hackensack: World Scientific.

Boltzmann Ludwig (1872). Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsberichte, der Akademie der Wissenschaften 66 (1872), 275–370. Translation: Further studies on the thermal equilibrium of gas molecules, in Kinetic Theory 2, 88–174, Ed. S.G.Brush, Pergamon, Oxford (1966).

Brillinger D. L. (2002) Second-order moments and mutual information in the analysis of time series, Recent Advances in Statistical Methods, Imperial College, London

Castañeda E. y Barros, V. (1994). Las tendencias de la precipitación en el Cono sur de América al este de los Andes, Meteorológica 19, 23–32.

Chapeau-Bolndeau F. (2007) Autocorrelation versus entropy-based autoinformation for measuring dependence in random signal, Physica A, 380.

Clausius R. (1850). *Annalen der Physik und Chemie* **79**: pp. 368-397, 500-524

Fazlollah M. Reza (1994) An Introduction to Information Theory, Dover Pub. Inc., New York.

Gray R. M. (2007). Entropy and Information Theory, Stanford Uni Press.

Kawachi, T., Maruyama, T., Singh, V. P. (2001) Rainfall entropy for delineation of water resources zones in Japan, J. Hydrol. 246, 36–44.

Kraskov A.; Stöghauser H. & Grassberger Estimating mutual information, Physical Review E 69: 066138

Lazo A. & Rathie, P.(1978) On the entropy of continuous probability distributions - Information Theory IEEE Transactions, 24(1)

Machta J. (1989) Entropy, information and computation, American Journal of Physics, 67.

Park Sung Y.; Bera, Anil K. (2011) Maximum entropy autoregressive conditional heteroskedasticity model, Journal of Econometrics (Elsevier)

Penalba O. y Vargas, W. (2001). Propiedades de precipitaciones extremas en zonas agropecuarias argentinas, Meteorol., 26, 39-55.

Shannon C. E. (1948) A Mathematical Theory of Communication Bell System, Technical Journal, 27.

Silber M. & Vanlesberg S. (2012) - Entropy Analysis of Synthetic Time Series. AQUAmundi Am05038: 001 – 008, DOI 10.4409/Am-038-12-0038

Thomas M. & Thomas, U. J. (1978) Elements of Information Theory, Wiley, New York.

Zurek W. H. (1989) Algorithmic randomness and physical entropy, Phys. Rev. A40.

DOI 10.4409/Am-049-12-0050