

# OBSERVED CHANGES IN THE PERSISTENCE AND CONDITIONAL ENTROPY ASSOCIATED WITH THE DAILY THERMAL STRUCTURE IN SOUTHERN SOUTH AMERICA

Gustavo Naumann\* and Walter M. Vargas

National Scientific and Technological Research Council (CONICET)  
Department of Atmospheric and Oceanic Sciences. Faculty of Sciences. University of Buenos Aires  
Buenos Aires - ARGENTINA

## Abstract

In this work we discussed about the spatial and temporal behavior of specific properties of the climate system. This was analyzed as changes in the persistence and system memory using the information theory and represented by the conditional entropy.

This study is performed by coupling a cluster analysis algorithm with conditional entropy with long-term reference series of daily maximum and minimum temperature distributed in southern South America between S23° and S55° latitude.

In a spatial analysis of the entropy it was found that exists an meridional gradient in the entire region, whit its maximum in the southern region (westerlies region). Taking into account the lower predictability of this system, small changes in this property imply significant variability. In this study the gradient imply that the predictability in the northern regions are two times more than in the southern part of South America.

An evidence of the temporal changes in the conditional entropy was observed with quasi-cyclical variations. If we analyse the low frequency variability on the conditional entropy trough the spectral estimate can be inferred that the dominant wave around 18 years. A nearly inverse course reveals that the conditional entropy of cold/warm days must decrease in the case of increasing persistence and the conditional entropy increase with the more presence of transitions of warm to cold and cold to warm events.

Taking into account the changes observed in the persistence and conditional entropy especially in groups that representing the warm and cold days. These results suggest changes in objective forecasting.

## 1. INTRODUCTION

The character of the dynamics, linear or non linear, and the precision of the measurement of the initial states decide about the horizon of predictability. For most complex systems, like meteorological or financial processes, we have at best a few general ideas about their predictability (Feistel, Ebeling (1989) and Ebeling, (2002)).

The problem we discuss here is about the spatial and temporal behavior about our chances to predict future states considering the persistence are good and in which cases they are rather bad in the southern South America region. A basic tool to analyze these questions are the information theory trough the conditional entropies introduced by Shannon (See Shannon 1948 and 1950) and used by many researchers as Leung and North (1990), Schürmann and Grassberger, 1996, Ebeling (1997), Werner et al (1999), Molgedey, Ebeling: (2000), Roulston and Smith (2002) and Tang et al (2008) among others.

In order to discriminate processes involved in these properties, introducing a non-hierarchical cluster analysis on the reference series of daily maximum and minimum temperature anomalies. With this algorithm each day is represented by each group, which is characterized by the values of their center of mass, the algorithm assigns each day to one of four groups: warm, wet, cold and dry (Vargas and Naumann, 2008). Also In this paper is found that the thermal properties of each group are associated with atmospheric processes of synoptic-scale, with circulation patterns defined for each group. The wet days are the primarily responsible for the changes observed in the series of maximum and minimum temperatures. These findings are consistent with trends in the series of precipitation in the region. In addition, it was detected a significant increase in the occurrence of warm days and a decrease in the frequency of cold days.

In this paper we analyze the conditional entropy of spell length of n-states clustered temperature by Shannon's Information-theoretical methods. A spatial and temporal analysis of these parameters allows to infer about the regions or in which periods we have more risks in the predictions accuracy due by the internal variability of the climatic system.

Section 2 presents the daily data used for the study. The conditional entropy and its relations whit the spell length is presented in Section 3. The spatial distribution of the uncertainties and its climatological interpretation are discussed in Section 4. Section 5 shows the temporal behavior of the conditional entropy. Finally, Sect. 4 summarizes the main conclusions.

---

\* *Corresponding author address:* Gustavo Naumann  
Departamento de Ciencias de la Atmósfera y los Océanos,  
Facultad de Ciencias Exactas y Naturales,  
Universidad de Buenos Aires, CONICET,  
Buenos Aires, Argentina  
e-mail: [gnaumann@at.fcen.uba.ar](mailto:gnaumann@at.fcen.uba.ar).

## 2. DATA AND METHOD

Stations or reference series used had to have long records (ideally covering the instrumental period) with good quality measurements and represent different or specific climatic regions. This reference series of daily maximum and minimum temperature was selected according to have a long quantity of data to produce a stable estimation of the clusters (in this case more than 20000 values) and a representative geographic distribution of stations, trying as much as possible to include the climatic regions of southern South America and to cover a wide latitudinal selection (we cover S23°–S55°; see Table 1).

Station	Longitude	Latitude	start	end
S. Rosa	-64.26	-36.54	1937	2004
R. Gallegos	-69.45	-51.99	1896	2004
Pergamino	-60.53	-33.90	1931	2000
Corrientes	-58.74	-27.43	1894	2004
Pilar	-63.85	-31.64	1931	2004
S. M.	-65.20	-26.80	1891	2000
O.C.B.A.	-58.42	-34.57	1906	2000
Campinas	-47.12	-23.00	1890	2003

Table 1. Description, geographic location and observation period of the 8 reference stations analyzed.

In a cluster analysis,  $p$ -dimensional explicative variables  $X$  are used for  $N$  objects; the objective is to group them into  $K$  groups ( $K \setminus N$ ), so that the variables of items that belong to one group resemble each other as much as possible and differ as much as possible from those of other groups.

After making the cluster analysis we obtained four groups characterized by four centroids. These provide a

new series of discrete values in which each day is no longer represented by various temperature anomalies, but by a symbol indicating that it belongs to the group. For more details of the classification see Vargas and Naumann (2008).

To characterize climatological processes, it is important to have a detailed knowledge of the temporal behavior of the parameters, like the memory of the process (Ebeling, 2002). These specific properties of the system can be analyzed as changes in the persistence and system memory represented by the conditional entropy. According with the cluster analysis algorithm each day is described by only one of the symbols wet, dry, cold or hot. We define  $P_i(X_1; \dots; X_i)$  as the probability of each group or symbol sequence of length  $i$ . This probability describes the persistence of each cluster over  $i$  time steps. The sequence of the probabilities  $p_i(x)$  states that  $x$  will be maintained over  $i$  days before it switches over to another sequence. This quantity is called "exit time distribution" (Nicolis et al., 1997).

Figure 1 shows the probability that the weather corresponding to each cluster for Buenos Aires. Here we can see that for the cold and warm days the distribution is similar as a Markov process, but for the wet and dry days the persistence is higher than for a Markov processes.

If this distribution changes with time we may associate this with a climatic change. In general, the exit time distribution can be fit to an analytical function. Several possible expressions were tested (Gabriels and Neumann, Nicolis et al., 1997). In most cases the best fit for each cluster are corresponding to an exponential distribution:

$$p_{(i)} = p_1 e^{(b(1-i^2))} \quad (1)$$

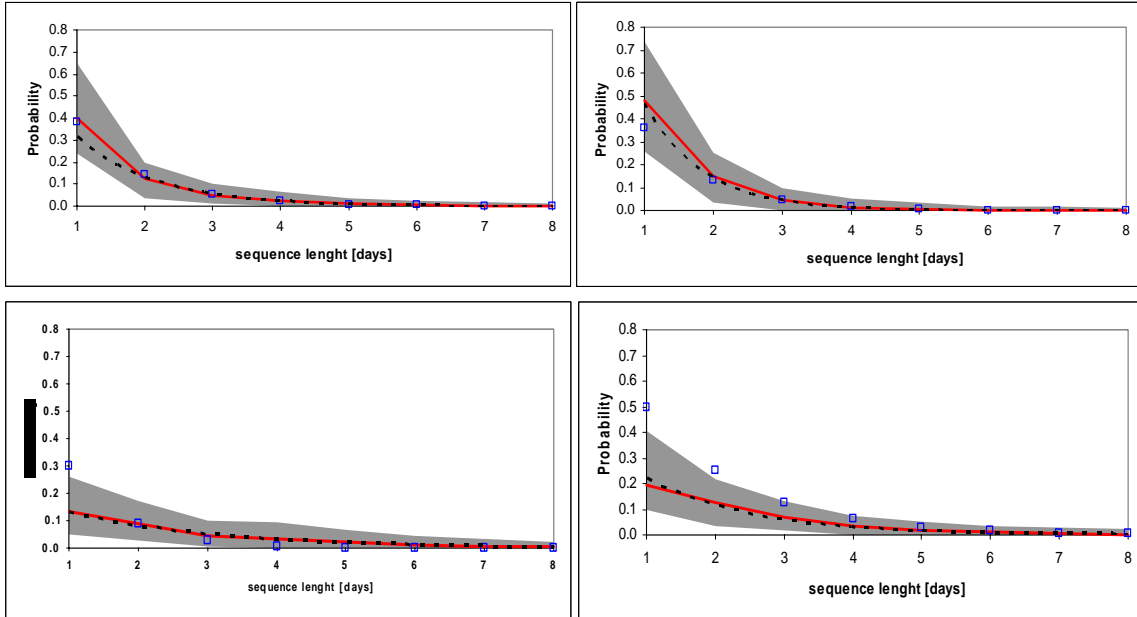


Figure 1. Calculated (solid line) and exponential fit (dashed line) exit time distribution for a) wet, b) dry, c) warm and d) cold days at Buenos Aires, and the attached values of the Markov process (squares).

or

$$p_{(i)} = (1 - p_1) p_1^{(i)} \quad (2)$$

where  $p_1$  is the transition probability for each group and  $p_{(i)}$  is the probability of a sequence of  $i$  days in the same state.

### 3. Conditional Entropy

The Shannon Entropy (Shannon, 1948 and 1950) is computed as follows: Let  $x$  be a discrete variable which may take  $i=1...m$  possible values. In this work  $x$  are the clustered temperature that can take four possible states; wet, dry, cold and hot.

$$H(x) = - \sum_{x=i}^m p_i(x) \log_2(p_i(x)) \quad (3)$$

From equation (2) it can be seen that the highest amount of uncertainty from an information source is realized when the output symbols of the source are equally probable. The entropy varies from 0 to  $\log(m)$ , so the entropy may be standardized so that it would range from 0 to 1, by dividing it by its maximum.

It may be easier to compare the amount of disorder of two systems, knowing that one system encountered more states than the other. Conditional entropy relies on transitional probabilities, for the conditional probability of state  $j$  of variable  $Y$  given state  $i$  of variable  $X$ ,  $[p(Y=j|X=i)]$ . The equation for computing the conditional entropy is [2]

$$H(y|x) = - \sum_{x=i}^{m1} p_i(x) \sum_{y=j}^{m2} p_{ij}(y|x) \log_2 p_{ij}(y|x) \quad (4)$$

This conditional entropy (mutual information) measures the uncertainty of predicting a state one step into the future, given a history consisting of  $n$  states; i.e. the present state and the previous states are known (Ebeling and Nicolis, 1992). Predictability in this work is measured by differences of Shannon entropies, in other words by conditional entropies.

The existence of long correlations is expressed by long decreasing tails of the conditional entropies. In general our expectation is that any long-range memory decreases the conditional entropies and improves the chances for predictions.

Let the  $n$ -tuple  $\{A_1...A_n\}$  be the  $\lambda$  states of a given spell of length  $n$ . Let further  $p(A_1...A_n)$  be the probability to find a spell with the states  $A_1...A_n$ , so the entropy per spell of length  $n$  is defined as

$$H_n = - \sum_{\{A_1...A_n\}}^{\lambda^n} p(A_1...A_n) \log_{\lambda} p(A_1...A_n) \quad (5)$$

From this entropy we derive conditional entropy  $h_n$  as the differences

$$h_n = H_{n+1} - H_n \quad (6)$$

The maximum of the uncertainty (in units of  $\log(\lambda)$ ) is  $h_n=1$ . Hence one can define the average predictability as the difference between the maximal and the actual uncertainty

$$r_n = 1 - h_n \quad (7)$$

It means that the predictability is related to the certainty that we have about the next state in the future in comparison to the available knowledge.

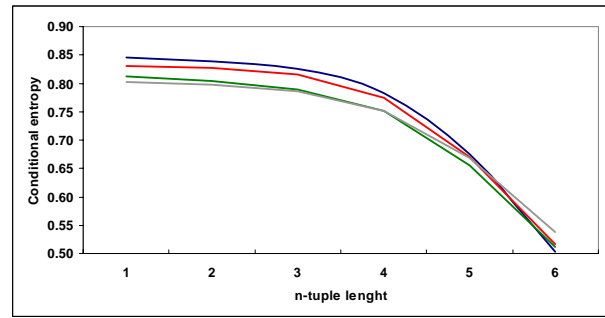


Figure 2. Conditional entropy (uncertainty) as a function of  $n$ -tuple length at Campinas (grey), Tucumán (green), Buenos Aires (red) and Rio Gallegos (Blue). For  $n > 6$  the calculation of the conditional entropy is not reliable due to large statistical errors.

In order to analyse these properties Figure 2 show the conditional entropy as function of spell length at four reference series. Here the differences between the conditional entropies of the six  $n$ -tupel are largest between the third and four spell length than between the other ones. Thus the growth in information is larger between this first and the second day than between the following days. For the forecast of climatological parameters we can say that the consideration of longer durations leads to an improvement in the predictability of the following day.

If analyze the spatial behavior of this parameter we can see a latitudinal behavior with greater uncertainties in Río Gallegos ( $lat=S52^\circ$ ) and lesser uncertainties in the tropical regions (Tucumán and Campinas).

### 3. Spatial variations of the uncertainties

Following this point of view, we analyze the spatial behavior of the entropy for all the stations in the region that ensure information of daily maximum and minimum temperature in at least 20000 days for guarantee an stable estimation of the clusters.

Figure 3 shows the field of entropies  $H$  for the spell length 2 calculated according equations (5). This caption reinforces the idea of a meridional gradient,

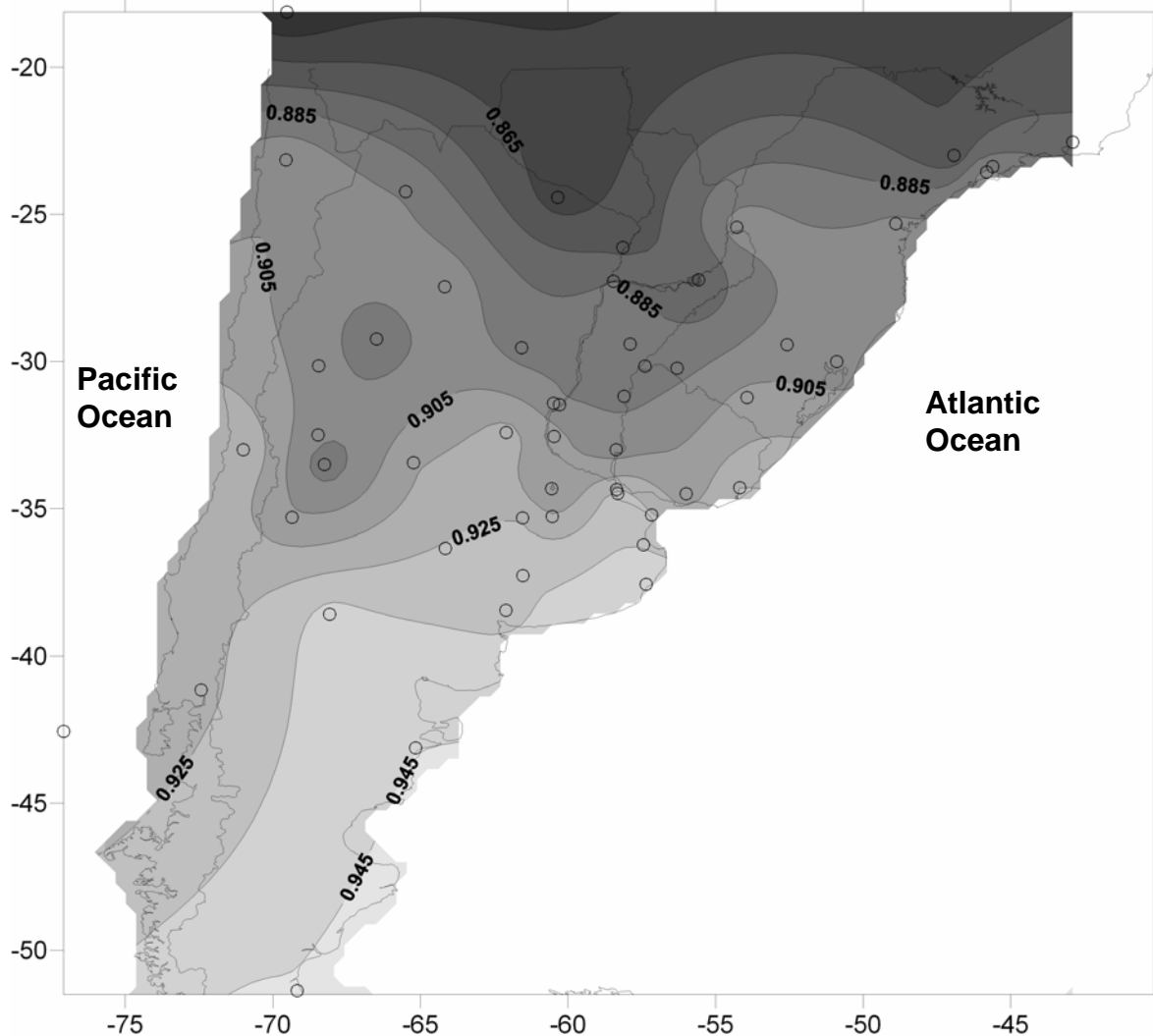


Figure 3. Spatial distribution of Block-entropy H for the spell length 2 at 54 stations of the region.

initially suggested by the results in figure 2 and also some inferences about the regional circulation and its predictability emerge.

In the Patagonia region (southern than S40°) where the higher entropies were observed are pointing out the preferential direction of the synoptic system (cold fronts, cyclonic and anticyclonic systems, etc.). Another aspect that we need to mention is the Andes role over the synoptic system trajectories evidenced by a zonal gradient mainly in the Patagonia where the west winds are dominant. The Andes effect is less important for latitudes northern than S30°.

In central Argentina and the coastal regions of Brazil exists a transition region between the mid-latitudes and tropical regimes. This last one characterized by the higher predictability due to the persistence of the warm advections from the Amazonia.

These results suggest that the risks associated with the same model prediction are different for each region. In other words the entropy measures the quantity of information that is needed for built an efficient model.

### 3. Temporal variations of conditional entropy and persistence

Initially our analysis of secular variations of the annual occurrence of each group reveals a significant increase in humid days, i.e., days with little thermal amplitude, at all stations between 25 and 40S (figure 3). This is also observed at Campinas (S23°), although at a lower significance level (10%). These results agree with the increase in precipitation observed by different authors, as Barros et al (2008), Haylock et al (2006), Liebmann et al (2004), Minetti et al (2003) and including the IPCC reports. This consistently shows that dry day occurrence decreases significantly in Santa Rosa, Pergamino and Buenos Aires (stations in central Argentina).

For practical applications, one is not so much interested in an average uncertainty of predictions but even more in a concrete prediction based on the observation of a concrete string of finite length  $n$ . Figure 4 shows the conditional entropy for warm and cold days and

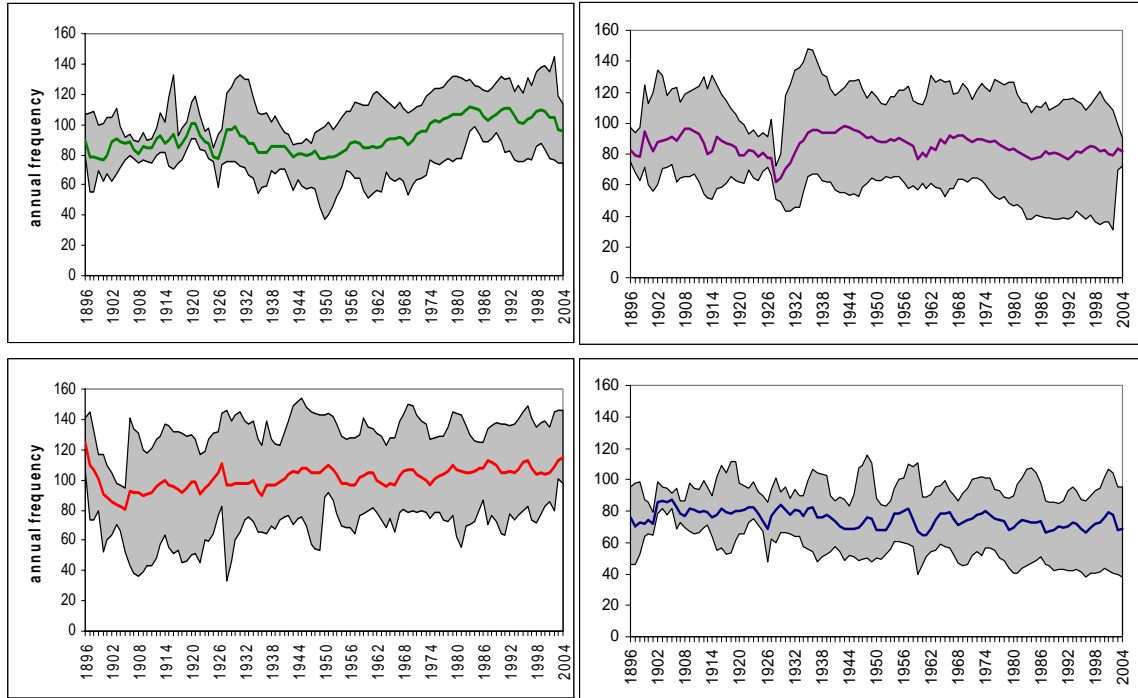


Figure 3. Annual frequency of occurrence of the a) wet days, b) dry days, c) warm days and d) cold days

transition probability between cold to cold and hot to hot days and between cold to hot and hot to cold, it means the persistence and antipersistence respectively. Here we can see that conditional entropy shows significant variations with time. In both cases these variation is  $\pm 0.10$  that represent changes in the predictability values between 15 -25%. That means that exist special years were the predictability increase/decrease up to two times higher than the average.

In the same way the nearly inverse course reveals that the conditional entropy of cold/warm days must decrease in the case of increasing persistence and the conditional entropy increase with the more presence of the transitions of warm to cold and cold to warm events. In this variation we found some periodicities in the entire region. Figure 5 shows the power spectrum of the conditional entropy for the warm group at Buenos Aires. This show a periodicity of about 18 years were the security of prediction based on persistence

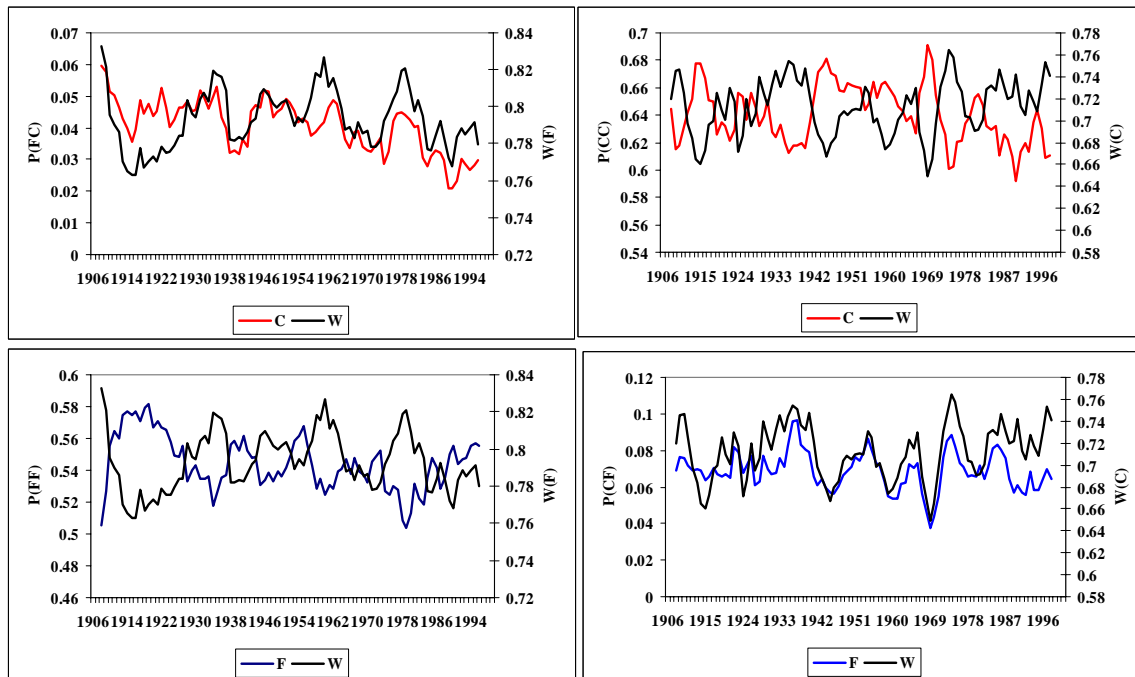


Figure 4. Conditional entropy (black) for warm days [W] left panel and for cold days [C] right panel, and the transition probability between W-C (a), W-W (b), C-C (c) and W-C (d) at Buenos Aires.

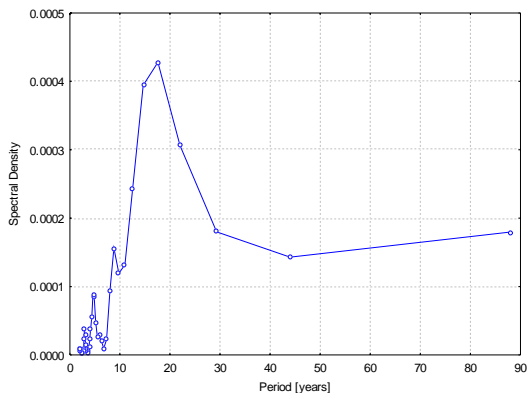


Figure 5. Spectral density of the conditional entropy for the warm group at Buenos Aires.

increased/decreased whit time. An increase of the conditional entropy is observed in the 1930s, 1950s and at the end of the 1970s, the last two maximum are coincident with two climatic jumps observed in many variables at the region (Minetti and Vargas, 1998). This may be related, because an increase in the disorder of the system could imply a change of state. After these events start a decrease of the conditional entropy reaching his minimum approximately ten years after.

#### 4. Conclusions

By coupling a cluster analysis algorithm with the analysis of the system memory trough conditional entropy and the exit time distribution complex climatological structures can be described better (Werner et al, 1999).

Classification of the daily series of temperature with the longest records produced four groups. The names of these groups were assigned according to their thermal properties (which reflect aspects of humidity, cloudiness, orography, etc.) and their association with precipitation on annual and daily scales. The groups are warm, wet, cold and dry. The groups differentiate weather types, which represent circulation patterns.

In a spatial analysis of the entropy it was found that exists an meridional gradient in the entire region, whit its maximum in the southern region of South America (westerlies region). Taking into account the lower predictability of this system, small changes in this property imply significant variability. In this study the gradient imply that the predictability in the northern regions are two times more than in the westerly part of South America.

An evidence of the temporal changes in the conditional entropy was observed with quasi-cyclical variations. If we analyse the low frequency variability on the conditional entropy trough the spectral estimate can be inferred that the dominant wave around 18 years. A nearly inverse course reveals that the conditional entropy of cold/warm days must decrease in the case of increasing persistence and the conditional entropy increase with the more presence of transitions of warm to cold and cold to warm events.

Taking into account the changes observed in the persistence and conditional entropy especially in groups that representing the warm and cold days. These results can suggest changes in objective forecasting models which involved time series.

In other way is suggested changes in the distribution of the circulation patterns could affect directly the objective forecast also.

Finally the study of the long-term variability of the entropy could suggest more evidences of climate changes.

**Acknowledgments.** This research was sponsored by projects UBA X-228 and CONICET PIP 5139.

#### References

Barros VR, Doyle ME, Camilloni IA, 2008: Precipitation trends in southeastern South America: relationship with ENSO phases and with low-level circulation. *Theoretical and Applied Climatology* 93(1-2): 19

Ebeling W., G. Nicolis, 1992: *Chaos, Solitons & Fractals* 2, 635.

Ebeling, W., 1997: Prediction and entropy of sequences with LRO. *Physica D* 109 42-50.

Ebeling W., 2002: Entropies and Predictability of Nonlinear Processes and Time Series. Notes In Computer Science; Proceedings of the International Conference on Computational Science-Part III. Springer-Verlag London, UK. Vol. 2331 Pages: 1209 – 1217.

Feistel, R., Ebeling, W., 1989: *Evolution of Complex Systems*, Kluwer Academic Publ., Dordrecht.

Gabriel K. R, J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv,. *Quarterly Journal of the Royal Meteorological Society.* 88, 375. 90-95.

Haylock, M.R., T.C. Peterson, L.M. Alves, T. Ambrizzi, Y.M.T. Anunciação, J. Baez, V.R. Barros, M.A. Berlato, M. Bidegain, G. Coronel, V. Corradi, V.J. Garcia, A.M. Grimm, D. Karoly, J.A. Marengo, M.B. Marino, D.F. Moncunill, D. Nechet, J. Quintana, E. Rebello, M. Rusticucci, J.L. Santos, I. Trebejo, and L.A. Vincent, 2006: Trends in Total and Extreme South American Rainfall in 1960–2000 and Links with Sea Surface Temperature. *J. Climate*, 19, 1490–1512.

Leung, L.Y., and G.R. North, 1990: Information Theory and Climate Prediction. *J. Climate*, 3, 5–14.

Liebmann, B., C.S. Vera, L.M.V. Carvalho, I.A. Camilloni, M.P. Hoerling, D. Allured, V.R. Barros, J. Báez, and M. Bidegain, 2004: An Observed Trend in Central South American Precipitation. *J. Climate*, 17, 4357–4367.

Minetti J., Vargas W.M., 1998: Trend and Jumps in the annual precipitation in South America, south of the 15°S. *Atmósfera* 11:205–223 México.

Minetti, J.L., W.M.Vargas, A.G.Poble, L.R.Acuña And G.Casagrande, 2004: Non-Linear trends and low frequency oscillation in annual precipitation over Argentina and Chile, 1931-1999. *Atmósfera* 16, 119-135.

Molgedey L., W. Ebeling, 2000: Local order, entropy and predictability of financial time series, *Eur. Phys. J B* 15 733-737.

Nicolis, C., Ebeling, W., Baraldi, C., 1997: Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. *Tellus*, 49A, 10-118.

Roulston, M.S., and L.A. Smith, 2002: Evaluating Probabilistic Forecasts Using Information Theory. *Mon. Wea. Rev.*, **130**, 1653–1660.

Shannon C.E., 1948: A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656.

Shannon, C. E., 1950: Prediction and entropy of printed English, *The Bell System Technical Journal*, 30:50-64.

Schürmann T. and P. Grassberger, 1996: Entropy estimation of symbol sequences, *CHAOS* 6 414.

Tang, Y., R. Kleeman, and A.M. Moore, 2008: Comparison of Information-Based Measures of Forecast Uncertainty in Ensemble ENSO Prediction. *J. Climate*, **21**, 230–247.

Vargas W. M., G. Naumann, 2008: Impacts Of Climatic Change And Low Frequency Variability In Reference Series On Daily Maximum And Minimum Temperature In Southern South America. *Regional Environmental Change*. Vol. 8 N° 2 45-87.

Werner PC, Gerstengarbe FW, Ebeling W, 1999: Changes in probability of sequences, exit time distribution and dynamical entropy in Potsdam temperature record. *Theor Appl Climatol* 62:125–132.