DE GRUYTER
OPEN

Acta Geophysica
vol. 63, no. 2, Apr. 2015, pp. 533-546
DOI: 10.2478/s11600-014-0229-2

Discriminating Between Different Streamflow Regimes by Using the Fisher–Shannon Method: An Application to the Colombia Rivers

Jorge O. PIERINI[1], Juan C. RESTREPO[2],
Michele LOVALLO[3], and Luciano TELESCA[4]

[1]Centro Científico y Tecnológico de Bahía Blanca (CCT-BB, CIC, UNS),
CRIBABB, Bahía Blanca, Argentina; e-mail: jpierini@criba.edu.ar
[2] Grupo de Física Aplicada: Océano y Atmósfera, Departamento de Fisica,
Universidad del Norte, Barranquilla, Colombia
[3]Agenzia Regionale per la Protezione dell'Ambiente di Basilicata (ARPAB),
Potenza, Italy
[4]National Research Council of Italy,
Institute of Methodologies for Environmental Analysis (CNR-IMAA), Tito, Italy

Abstract

The Fisher–Shannon (FS) information plane, defined by the Fisher information measure (FIM) and the Shannon entropy power (N_X), was robustly used to investigate the complex dynamics of eight monthly streamflow time series in Colombia. In the FS plane the streamflow series seem to aggregate into two different clusters corresponding to two different climatological regimes in Colombia. Our findings suggest the use of the statistical quantity defined by the FS information plane as a tool to discriminate among different hydrological regimes.

Key words: Caribbean streamflow, Fisher information measure, Shannon entropy.

1. INTRODUCTION

The analysis of the time dynamics of a river streamflow has gained a great importance because it can be considered as the integral of the annual or

---

Ownership: Institute of Geophysics, Polish Academy of Sciences;

interannual climatic fluctuations, which characterize its basin. In fact, the literature dealing with river streamflows is mainly focused on the role of a river streamflow as a sort of climatic indicator to identify and characterize different climatic periods (Pekarova *et al.* 2003, Milliman *et al.* 2008). Indeed, the use of streamflow changes for detecting significant trends, identifying major oscillations periods, and determining relationships between hydrological responses and climatic forcings was performed by several authors. Probst and Tardy (1987), analyzing the annual streamflow data of fifty major rivers distributed all around the world deduced that during the first half of the last century Europe and Asia were affected by a significant humid regime that affected Africa, North and South America in the last half of that century. Pekarova *et al.* (2003) analyzing many streamflows worldwide identified alternating phases of wet and dry periods and extreme cycles of high-low discharge with periods from years to tens of years. But Milliman *et al.* (2008) noted significant changes just in individual rivers and at regional levels. Some authors have also analyzed the relationship between streamflow of Colombian rivers and El Niño-Southern Oscillation (ENSO) phenomenon (Mesa *et al.* 1997, Restrepo and Kjerfve 2000, Gutiérrez and Dracup 2001, Restrepo *et al.* 2014).

From all these studies it is possible to argue that investigating the time structure of streamflows was principally aimed at identifying and possibly quantifying climate-related long-term trends, cycles, scaling, or also anomalous patterns with respect to a certain background.

Streamflows can be considered as the output of systems whose complexity can be measured by its organization and order. In order to get such a knowledge, we need to use appropriate methodological approaches.

In the present paper, we investigate the time dynamics of eight streamflows measured by gauging stations located along the Caribbean plain of Colombia using the Fisher–Shannon (FS) method. The FS approach is based on the information content of the time series using the statistical measures of the Fisher information measure and the Shannon entropy (described in Section 2.2), and is a powerful statistical method to gain into insight the inner dynamics of a complex system.

2. DATA AND METHODS

2.1 Hydrological data

The Caribbean plain of Colombia is located in the northernmost South America. It extends from the Darien tropical rainforest in the Colombia–Panama border, to the Peninsula de La Guajira in the east, and the slopes of the Cordillera de los Andes in the south (Fig. 1). It comprises extensive lowlands with heights below 100 m, plateaus with heights between 200 and

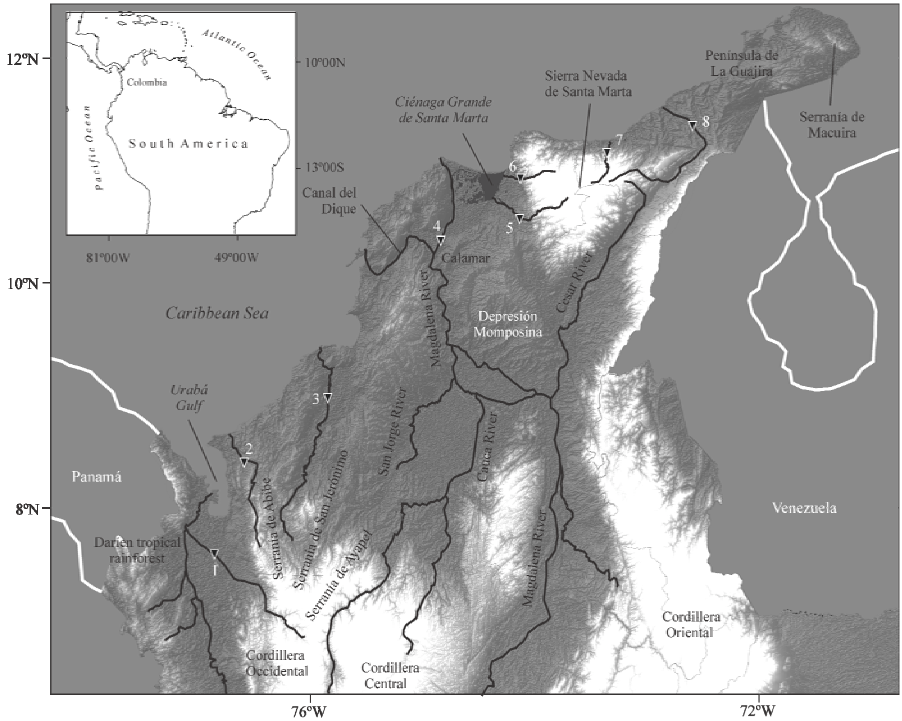

Fig. 1. Caribbean plain of Colombia in northwest South America. Major topographic features, selected rivers and gauge stations: 1 – Sucío River, 2 – Mulatos River, 3 – Sinú River, 4 – Magdalena River (Calamar), 5 – Fundación River, 6 – Frío River, 7 – Palomino River, and 8 – Ranchería River.

Table 1

Drainage basin, river length, headwater, and mean monthly streamflow of selected rivers in the Caribbean plain of Colombia.

River	Drainage basin [× 10^3 km^2]	River length [km]	Headwater [m a.s.l.]	Mean monthly streamflow [m^3 s^{-1}]
Andean Rivers				
Sucío	4.5	169	4080	278.5
Mulatos	0.01	115	2120	4.6
Sinú	14.7	415	3960	398.1
Magdalena	257.4	1540	3650	6497.1
Sierra Nevada Rivers				
Fundación	1.87	161	2586	28.2
Frío	0.32		3975	13.8
Palomino	0.68		4325	25.7
Rancheria	4.23	151	3725	12.8

Table 2

Name, location, and historic record period of studied discharge stations

River	Gage station	Location			Historic record
		Elevation [m a.s.l.]	Longitude	Latitude	
Andean Rivers					
Sucio	Mutata	132	76°26W	7°13N	1976-2010
Mulatos	Pueblo Bello	84	76°31W	8°12N	1977-2010
Sinu	Cotoca Abajo	5	75°51W	9°13N	1970-2010
Magdalena	Calamar	8	74°55W	10°15N	1941-2010 ($Q_{\min}$) 1969-2010 ($Q_{\max}$)
Sierra Nevada Rivers					
Fundacion	Fundacion	55	74°11W	10°31N	1958-2010
Frio	Rio Frio	30	74°09W	10°34N	1967-2009
Palomino	Puente Carretera	30	73°34W	11°14N	1973-2010
Rancheria	Hacienda Guamito	76	72°37W	11°10N	1979-2007

1000 m in the southwest (Serranías de Abibe, San Jerónimo, and Ayapel) and northeast (Serranía de Macuira), and one of the highest coastal mountains of the world, named Sierra Nevada de Santa Marta, with heights up to 5000 m.

The Instituto de Hidrología, Meteorología y Estudios Ambientales – Colombia (IDEAM) provided monthly maximum and minimum streamflow data from the main rivers of the Caribbean plain of Colombia (Fig. 1). The selection of rivers and corresponding gauging stations was based on two important conditions: (1) the station is located close to the downstream part of the basin; and (2) its hydrological data record is longer than thirty years (see Tables 1 and 2 for the characteristics of the selected rivers).

2.2 The Fisher–Shannon information plane

The Fisher–Shannon (FS) information plane represents an efficient tool to investigate the complex temporal fluctuations of nonstationary signals. It is constructed with coordinate axes given by the Fisher information measure (FIM) and the Shannon entropy power (N_X) that are both well known in the context of information theory. The FIM quantifies the amount of organization or order in a system, while N_X measures its degree of uncertainty or disorder. The FIM was developed by Fisher (1925) in the context of statistical estimation. Then, it was utilized for different aims. Frieden (1990) used the FIM to describe the evolution laws of physical systems. Martin *et al.* (1999, 2001) applied it to characterize the temporal fluctuations of electroencephalograms (EEG) and to detect significant dynamical changes. Complex geophysical and environmental phenomena, like volcano-related signals, earthquake-related electromagnetic signals, and atmospheric particulate matter, benefited of the application of the FIM methodology to gain insight into their inner time dynamics and the mechanisms underlying their temporal fluctuations and to reveal precursory signatures of critical phenomena (Lovallo and Telesca 2011, Telesca and Lovallo 2011, Telesca *et al.* 2009, 2010, 2011).

Shannon entropy is used to quantify the uncertainty of the prediction of the outcome of a probabilistic event (Shannon 1948a, b); in fact, it is zero for deterministic events. For continuous distributions the Shannon entropy can take any real positive and negative value. In order to avoid the difficulty arising with negative information measures, the so-called Shannon power entropy N_X (defined below) can be used instead of the Shannon entropy.

Let $f(x)$ be the probability density of a signal x, then its FIM I is given by

$$I = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial x} f(x) \right)^2 \frac{dx}{f(x)} , \tag{1}$$

and its Shannon entropy is defined as (Shannon 1948a, b):

$$H_X = - \int_{-\infty}^{+\infty} f_X(x) \log f_X\left(x\right) dx . \tag{2}$$

As specified above, the notion of Shannon entropy power will be used

$$N_X = \frac{1}{2\pi e} e^{2H_X} . \tag{3}$$

The two measures satisfy the so-called "isoperimetric inequality" $IN_X \geq D$, where D is the dimension of the space (Esquivel *et al.* 2010). The isoperimetric inequality indicates that FIM and Shannon entropy power are linked to each other, and the so-called Fisher–Shannon (FS) information plane represents a tool for the characterization of signals. The product IN_X can also be employed as a statistical measure of complexity (Romera and Dehesa 2004), and for 1-dimensional space, the line $IN_X = 1$ separates the FS plane in two parts, one allowed ($IN_X > 1$), and the other not allowed ($IN_X < 1$).

The calculation of the FIM and the Shannon entropy depends on the calculation of the probability density function $f(x)$ (*pdf*). The *pdf* can be estimated by means of the kernel density estimator technique (Devroye 1987, Janicki and Weron 1994) that approximates the density function as:

$$\hat{f}_M(x) = \frac{1}{Mb} \sum_{i=1}^{M} K\left(\frac{x - x_i}{b} \right) , \tag{4}$$

with b the bandwidth, M the number of data, and $K(u)$ the kernel function, a continuous non-negative and symmetric function satisfying the two following conditions:

$$K(u) \geq 0 \text{ and } \int_{-\infty}^{+\infty} K(u) du = 1 . \tag{5}$$

In our study, we estimated the *pdf* $f(x)$ by means of the algorithm developed by Troudi *et al.* (2008) combined with that developed in Raykar and Duraiswami (2006), that uses a Gaussian kernel with zero mean and unit variance:

$$\hat{f}_M(x) = \frac{1}{M\sqrt{2\pi b^2}} \sum_{i=1}^{M} e^{-\frac{(x-x_i)^2}{2b^2}} . \tag{6}$$

3. RESULTS AND DISCUSSION

We analyzed eight monthly streamflow time series recorded in Caribbean plains of Colombia. Figure 1 shows the geographical location of the gauging stations. Figure 2 shows the raw data. Selection was based on record length and location of the stations in order to cover a long range of years and a wide Colombian coast. Data cover the instrumental period, are proved to contain good quality measurements and correspond to different basins of Caribbean Colombia. We firstly applied the standard well-known power spectral density analysis to detect periodicities. We can identify annual and seasonal oscillations (as it was shown by the peaks in the power spectra) (Fig. 3 shows the power spectrum of Rancheria $Q_{\min}$ data as an example), which we removed before applying the Fisher–Shannon method. Having filtered out the annual and seasonal cycles and normalized the monthly maximum ($Q_{\max,N}$) and minimum ($Q_{\min,N}$) streamflows (in order to avoid any dependence on the real amount of water volume per second), for each site we constructed a residual time series as $(Q_{\max,N} + Q_{\min,N})/2$ (Fig. 4).

Figure 5 shows the Fisher–Shannon (FS) information plane: the y-axis represents the FIM and the x-axis represents the Shannon entropy power N_X; each symbol represents a residual streamflow series. The analysis of the FS information plane allows to discriminate two clusters, the one comprising the north-eastern sites (Rancheria, Palomino, Frio, and Fundación) and the other comprising the south-western ones (Sucio, Sinu, and Magdalena); Mulatos, even though it belongs to the second group, in the FS plane is closer to the first group. The two clusters reflect the different hydrological conditions that characterize the Andean rivers and the Sierra Nevada ones: the Andean rivers drain rather large basins, with Magdalena being the largest fluvial system

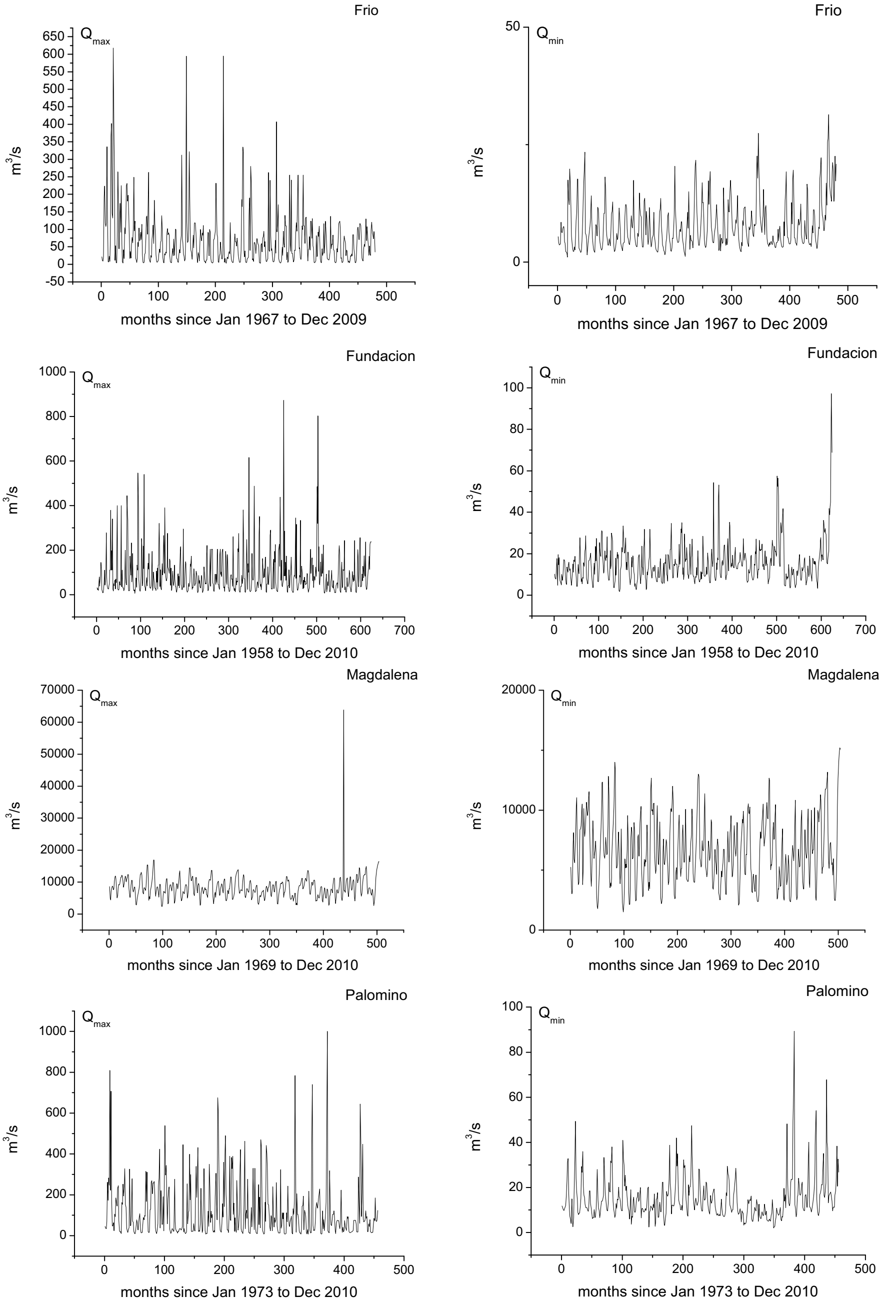

Fig. 2. Continued on the next page.

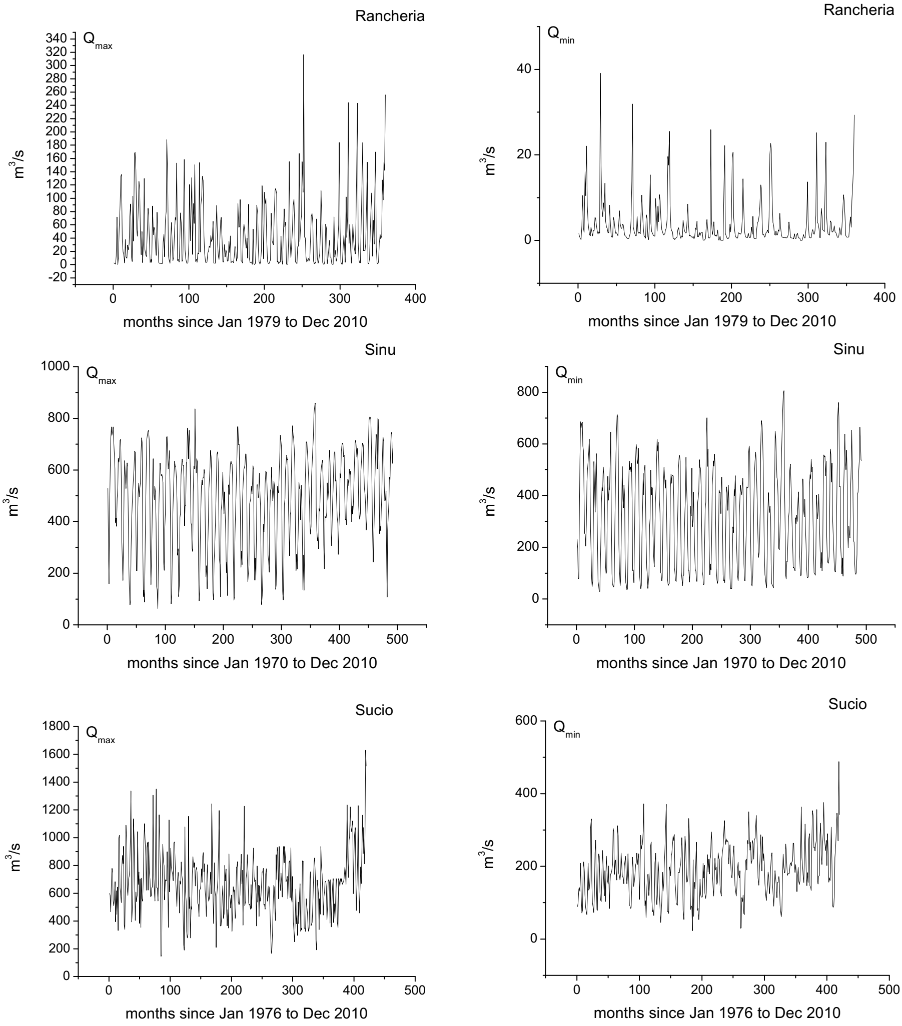

Fig. 2. Mean monthly maximum and minimum streamflows at the selected sites shown in Fig. 1.

in the Caribbean plain; while the rivers of the Sierra Nevada de Santa Marta drain small mountainous basins (< 5000 km^2) with steep gradients and limited alluvial floodplains. The Andean Mulatos river has a local pattern very likely influenced by the local orographic effects, with a small drainage basin that makes it closer to the Sierra Nevada cluster.

The Fisher–Shannon analysis shows that the Andean rivers are characterized by lower FIM (higher N_X) and the Sierra Nevada ones by higher FIM

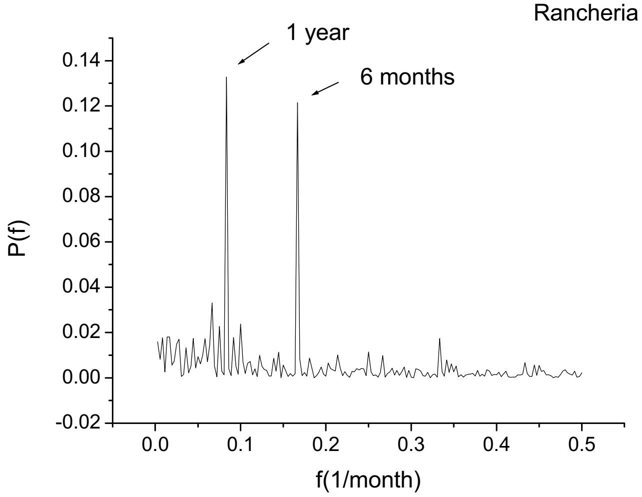

Fig. 3. Power spectrum of Rancheria $Q_{\min}$ data.

(lower N_X). This indicates that the Andean rivers are featured by lower system organization and order, while the Sierra Nevada ones by lower system disorder and higher organization. Hence, the Andean rivers undergo a disordered regime, with no preference for any particular set of states, while the Sierra Nevada ones undergo an ordered regime, in which a particular set of states is preferred. In fact, since the fluvial systems of the Sierra Nevada have drainage areas smaller than 5000 km^2 in mountainous zones, the topographical setting is a primary factor controlling streamflow variability. Instead, the Andean rivers, since they drain extensive plateaus or low-lying alluvial valleys, are subject to the coexistence of several controlling factors that enhance their disorder degree and lower their organization structure. For instance, the Magdalena river has a basin formed by 151 subcatchments, 42 of which are second-order watersheds, several main tributaries, like Cauca river (second largest river in Colombia), and, furthermore, characterized by high tectonic activity, hillslopes commonly exceeding 45°, landslides, steep gradients, and high relief tributary basins (Restrepo and Restrepo 2005, Restrepo *et al.* 2006). Fath *et al.* (2003) asserted that the Shannon entropy can be considered as a measure of the degrees of freedom of a system; and, hence, a system with few degrees of freedom has a high information content. In our case, the Sierra Nevada rivers, being less extensive and with a basin smaller and more concentrated than those of Andean rivers, are characterized by streamflow time variability with a lower number of degrees of freedom and high information content, as clearly indicated by

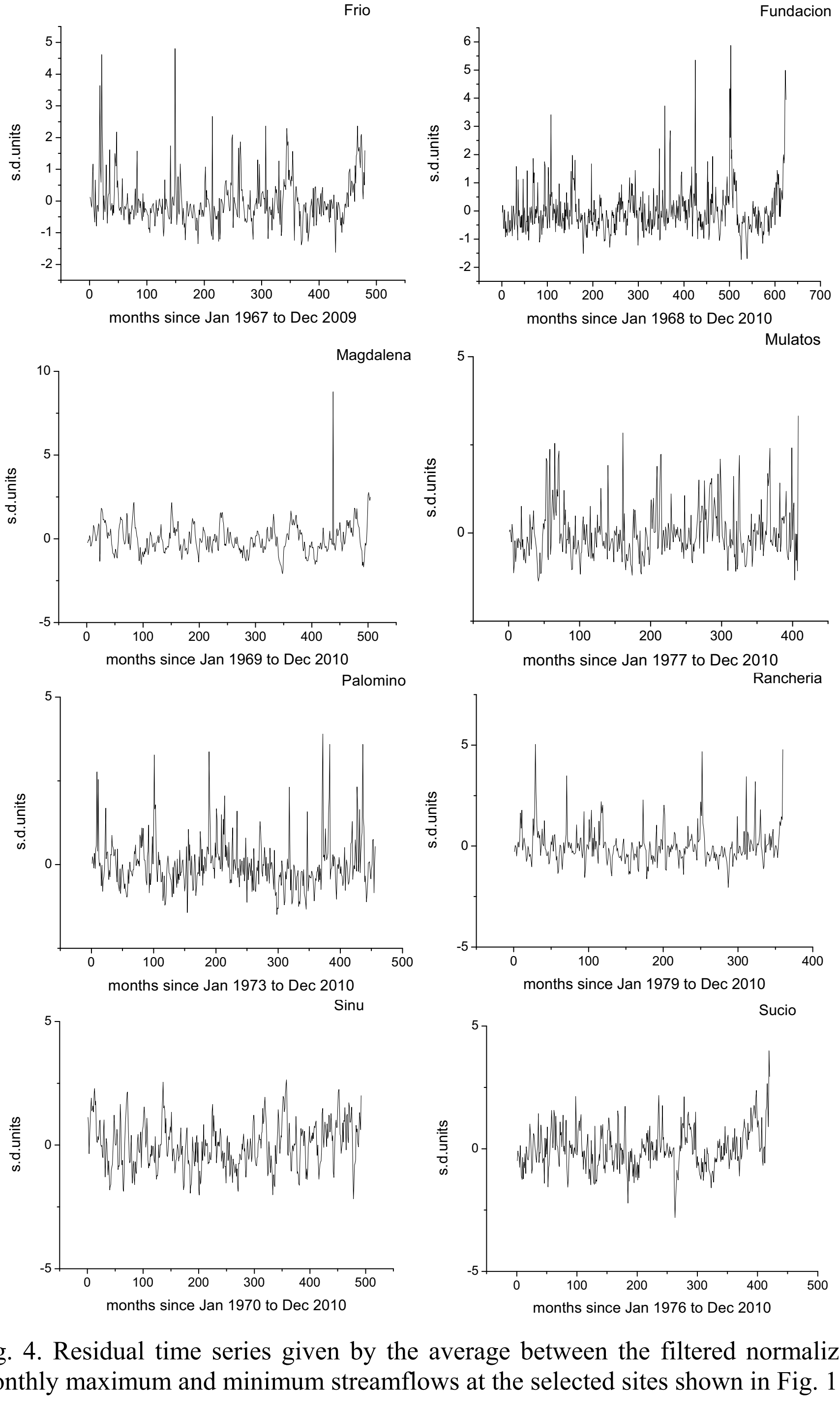

Fig. 4. Residual time series given by the average between the filtered normalized monthly maximum and minimum streamflows at the selected sites shown in Fig. 1.

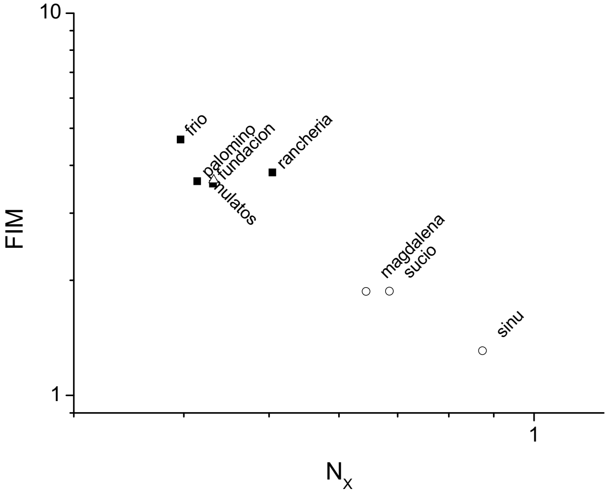

Fig. 5. Fisher–Shannon (FS) information plane: the *y*-axis represents the FIM and the *x*-axis represents the Shannon entropy power N_X; each symbol represents a residual streamflow series.

the FS information plane. As it can be argued, such a number of streamflow controlling factors, larger for the Andean rivers than Sierra Nevada ones, may organize and order the streamflow data, as indicated by the significantly different FIM and Shannon entropy power values.

4. CONCLUSIONS

We analyzed eight monthly streamflow time series recorded in Caribbean plain of Colombia by using the Fisher–Shannon statistical method. The streamflow time series aggregate into two different clusters: one corresponds to Andean Rivers Group (except of Mulatos) and the other corresponding to the Sierra Nevada Group. The Andean Group is characterized by higher Shannon entropy power and lower FIM with respect to the Sierra Nevada Group. This indicates that the Sierra Nevada rivers are featured by higher organization and order than the Andean rivers that, on the contrary, are characterized by greater disorder and uncertainty. A connection between these dynamical properties and the particular characteristics of the topography and drainage area of the two groups was suggested.

Acknowledgement. This research was funded by the Universidad del Norte (Dirección de Investigación, Desarrollo e Innovación – DIDI Grant) and the Universidad Pontificia Bolivariana – Sede Montería (Centro Integral para el Desarrollo de la Investigación – CIDI Grant). The support of Universidad del Norte Fellowship and CEMarin Fellowship (Colombian Center of Excellence for Marine Science Research; sponsored by the German Academic Exchange Service – DAAD), awarded to one of the authors (J.C. Restrepo), is gratefully acknowledged.

References

Devroye, L.A. (1987), *Course on Density Estimation*, Birkhauser, Boston.

Esquivel, R.O., J.C. Angulo, J. Antolin, J.S. Dehesa, S. López-Rosa, and N. Flores-Gallegos (2010), Analysis of complexity measures and information planes of selected molecules in position and momentum spaces, *Phys. Chem. Chem. Phys.* **12**, 7108-7116, DOI: 10.1039/B927055H.

Fath, B.D., H. Cabezas, and C.W. Pawlowski (2003), Regime changes in ecological systems: an information theory approach, *J. Theor. Biol.* **222**, 4, 517-530, DOI: 10.1016/S0022-5193(03)00067-5.

Fisher, R.A. (1925), Theory of statistical estimation, *Math. Proc. Cambridge Philos. Soc.* **22**, 5, 700-725, DOI: 10.1017/S0305004100009580.

Frieden, B.R. (1990), Fisher information, disorder, and the equilibrium distributions of physics, *Phys. Rev. A* **41**, 4265-4276, DOI: 10.1103/PhysRevA.41.4265.

Gutiérrez, F., and J.A. Dracup (2001), An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño–Southern Oscillation indicators, *J. Hydrol.* **246**, 1-4, 181-196, DOI: 10.1016/S0022-1694(01) 00373-0.

Janicki, A., and A. Weron (1994), *Simulation and Chaotic Behavior of alpha-Stable Stochastic Processes*, Marcel Dekker Inc., New York.

Lovallo, M., and L. Telesca (2011), Complexity measures and information planes of x-ray astrophysical sources, *J. Stat. Mech.* **2011**, 3, P03029, DOI: 10.1088/1742-5468/2011/03/P03029.

Martin, M.T., F. Pennini, and A. Plastino (1999), Fisher's information and the analysis of complex signals, *Phys. Lett. A* **256**, 2-3, 173-180, DOI: 10.1016/S0375-9601(99)00211-X.

Martin, M.T., J. Perez, and A. Plastino (2001), Fisher information and nonlinear dynamics, *Physica A* **291**, 1-4, 523-532, DOI: 10.1016/S0378-4371(00)00531-8.

Mesa, O., G. Poveda, and L. Carvajal (1997), *Introducción al Clima de Colombia,* Universidad Nacional de Colombia, Bogotá, Colombia, 390 pp. (in Spanish).

Milliman, J.D., K.L. Farnsworth, P.D. Jones, K.H. Xu, and L.C. Smith (2008), Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951-2000, *Global Planet. Change* **62**, 3-4, 187-194, DOI: 10.1016/j.gloplacha.2008.03.001.

Pekarova, P., P. Miklanek, and J. Pekar (2003), Spatial and temporal runoff oscillation analysis of the main rivers of the world during the 19th-20th centuries, *J. Hydrol.* **274**, 1-4, 62-79, DOI: 10.1016/S0022-1694(02)00397-9.

Probst, J.L., and Y. Tardy (1987), Long range streamflow and world continental runoff fluctuations since the beginning of this century, *J. Hydrol.* **94**, 3-4, 289-311, DOI: 10.1016/0022-1694(87)90057-6.

Raykar, V.C., and R. Duraiswami (2006), Fast optimal bandwidth selection for kernel density estimation. **In**: J. Ghosh, D. Lambert, D. Skillicorn, and J. Srivastava (eds.), *Proc. Sixth SIAM Int. Conf. Data Mining, 20-22 April 2006, Bethesda, USA*, 524-528.

Restrepo, J.C., and J.D. Restrepo (2005), Efectos naturales y antrópicos en la producción de sedimentos de la Cuenca del Río Magdalena, *Rev. Acad. Col. Cien. Exact. Fís. Nat.* **29**, 111, 239-254 (in Spanish).

Restrepo, J.C., J.C. Ortíz, J. Pierini, K. Schrottke, M. Maza, L. Otero, and J. Aguirre (2014), Freshwater discharge into the Caribbean Sea from the rivers of Northwestern South America (Colombia): Magnitude, variability and recent changes, *J. Hydrol.* **509**, 266-281, DOI: 10.1016/j.jhydrol.2013.11.045.

Restrepo, J.D., and B. Kjerfve (2000), Magdalena river: interannual variability (1975-1995) and revised water discharge and sediment load estimates, *J. Hydrol.* **235**, 1-2, 137-149, DOI: 10.1016/S0022-1694(00)00269-9.

Restrepo, J.D., B. Kjerfve, M. Hermelin, and J.C. Restrepo (2006), Factors controlling sediment yield in a major South American drainage basin: the Magdalena River, Colombia, *J. Hydrol.* **316**, 1-4, 213-232, DOI: 10.1016/j.jhydrol.2005.05.002.

Romera, E., and J.S. Dehesa (2004), The Fisher–Shannon information plane, an electron correlation tool, *J. Chem. Phys.* **120**, 19, 8906-8912, DOI: 10.1063/1.1697374.

Shannon, C.E. (1948a), A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 3, 379-423, DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shannon, C.E. (1948b), A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 4, 623-656, DOI: 10.1002/j.1538-7305.1948.tb00917.x.

Telesca, L., and M. Lovallo (2011), Analysis of the time dynamics in wind records by means of multifractal detrended fluctuation analysis and the Fisher–Shannon information plane, *J. Stat. Mech.* **2011**, 7, P07001, DOI: 10.1088/1742-5468/2011/07/P07001.

Telesca, L., M. Lovallo, A. Ramirez-Rojas, and F. Angulo-Brown (2009), A nonlinear strategy to reveal seismic precursory signatures in earthquake-related self-potential signals, *Physica A* **388**, 10, 2036-2040, DOI: 10.1016/j.physa.2009.01.035.

Telesca, L., M. Lovallo, and R. Carniel (2010), Time-dependent Fisher Information Measure of volcanic tremor before the 5 April 2003 paroxysm at Stromboli volcano, Italy, *J. Volcano. Geoth. Res.* **195**, 1, 78-82, DOI: 10.1016/j.jvolgeores.2010.06.010.

Telesca, L., M. Lovallo, H.-L. Hsu, and C.-C. Chen (2011), Analysis of dynamics in magnetotelluric data by using the Fisher-Shannon method, *Physica A* **390**, 7, 1350-1355, DOI:. 10.1016/j.physa.2010.12.005.

Troudi, M., A.M. Alimi, and S. Saoudi (2008), Analytical plug-in method for kernel density estimator applied to genetic neutrality study, *EURASIP J. Adv. Sig. Proc.* **2008**, 739082, DOI: 10.1155/2008/739082.

Received 28 January 2014
Received in revised form 23 June 2014
Accepted 23 June 2014