Adaptive Bandwidth Selection in Nonparametric Kernel Density Estimation for Time Signal Anomaly Detection using Jensen-Shannon Divergence Controlling

Antonio Squicciarini[a], Elio Valero Toranzo[b] and Alejandro Zarzo[a]

a) Universidad Politécnica de Madrid (Spain), *b*) Universidad Rey Juan Carlos (Spain)

1) Introduction

- Entropic, Information and Complexity measures showed capability describing non-stationary signals, but in **case-specific application** with little or no focus on main parameter tuning.
- **Almost exclusively discrete inference**. Limited attention has been given to continuous non-parametric inference solutions.
  - Underlying phenomena are continuous, and the signal is the result of a discretization of it.
  - Some information measures (such as non-parametric Fisher) are uniquely defined in the discrete case.

With respect to continuous inference, the biggest limitation is the absence of a solution for Kernel Density Estimation (KDE) [3] bandwidth optimization specifically designed for time series data that:

- Quantifies the **overfitting-underfitting tradeoff**
- Does not require any assumptions about the data distributions
- Is optimizable offline

2) Methodology Steps

Given a discrete signal composed of N equispaced samples,

$$\{x_i\}_{i=1}^{N} = \{x(t_i) = x_i \in \mathbb{R},\ i = 1, \dots, N\} \tag{1}$$

1. Apply overlapping window division is defined through sliding temporal window as:

$$W_j(\delta, \Delta) = \{x_i,\ i = 1 + j\delta - \Delta, \dots, j\delta\}\,. \tag{2}$$

2. Compute the KDE non-parametric inference

$$\hat{p}(x) = \mathcal{K}_h[W_j(\delta, \Delta)] = \frac{1}{h\Delta} \sum_{i=1+j\delta-\Delta}^{1+j\delta} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

3. Apply the entropic, information and complexity functionals (for example, in case of differential Shannon Entropy: $a_j = \mathbb{H}[\hat{p}_j(x)]$)

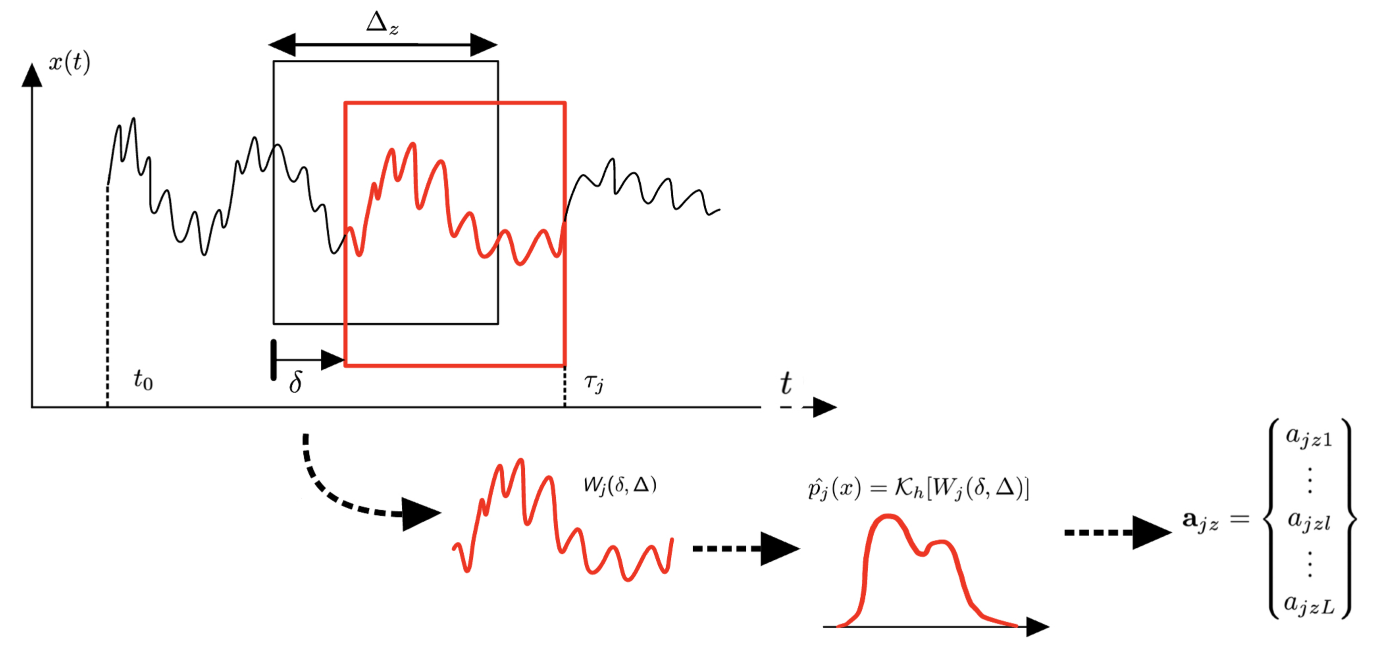

Figure: This figure illustrates the methodology employed.

Repeat the methodology considering:

- A set of Z window scales
- L desired continuous entropic, information and complexity measures

Consequently, we obtain an ordered sequence of matrices $\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_M$, where $\boldsymbol{A}_j \in \mathbb{R}^{Z\times L}$. The simultaneous usage of various measures reveals unique characteristics of the underlying time signal dynamic [5, 2].

3) JSD-h algorithm

Jensen-Shannon Divergence (JSD) with more than two Probability Density Functions (PDFs) ($M \geq 2$),

$$JS^{\pi}(\{p_j\}_{j=1}^{M}) = \sum_{j=1}^{M} \pi_j D_{KL}(p_j||m) = \mathbb{H}\left[\sum_{j=1}^{M} \pi_j p_j\right] - \sum_{j=1}^{M} \pi_j \mathbb{H}[p_j]\,. \tag{4}$$

Where $m = \sum_{j=1}^{M} \pi_j p_j$. The upper bound of the JSD is given in terms of the entropy of the PDF weight distribution as,

$$JS^{\pi}(\{p_j\}_{j=1}^{M}) \leq \mathbb{H}\left[\{\pi_j\}_{j=1}^{M}\right]\,. \tag{5}$$

The JSD score utilised to pick the bandwidth, h, is defined as,

$$S^{(JS)}(h, \Delta, \delta) = JS^{\pi}\left[\{\mathcal{K}_h[W_j(\delta, \Delta)]\}_{j=1}^{M^*}\right] = JS^{\pi}\left[\left\{\frac{1}{h\Delta}\sum_{i=\delta\cdot j}^{\delta\cdot j+\Delta} K\left(\frac{x - x_i}{h}\right)\right\}_{j=1}^{M^*}\right], \tag{6}$$

with M^* the total number of healthy PDFs.

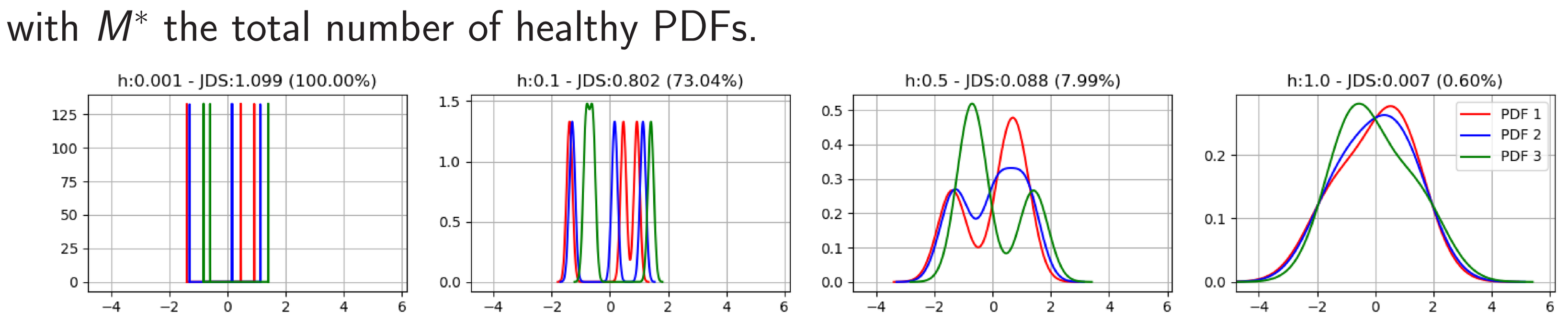

Figure: Jensen-Shannon Divergence (JSD) computed between three simple PDFs, with varying values of the smoothing parameter h.

The $S^{(JS)}$ is monotonically decreasing with respect h, thus, fixed a threshold th^{JS}, the h that achieved this exact balance can be found with bisection methods.

$$S^{(JS)}(h, \Delta) = th^{JS} \log M^* \to h^*(\Delta)\,. \tag{7}$$

This approach allows us to control the bandwidth at each scale, Δ, with only one hyperparameter, th^{JS}.

4) Numerical Simulations

Entropic and Information functionals utilised: Shannon entropy (8) [1], Tsallis entropy (9) [6], and Rényi entropy (10) [4].
Here, $p(x)$ represents a PDF ($\int_\Omega p(x) = 1$ and $p(x) \geq 0\ \ \forall x \in \Omega \subseteq \mathbb{R}$).

$$\mathbb{H}[p] = -\int_\Omega p(x) \ln p(x)\, dx, \tag{8}$$

$$\mathbb{H}_q[p] = \frac{1}{1-q}\left(1 - \int_\Omega p(x)^q\, dx\right) \quad q \in \mathbb{R} \tag{9}$$

$$\mathbb{H}_\alpha[p] = \frac{1}{1-\alpha} \ln\left(\int_\Omega p(x)^\alpha\, dx\right) \quad \alpha \in \mathbb{R}. \tag{10}$$

$$\mathbb{I}[p] = \int_\Omega \frac{\left(\frac{d}{dx}p(x)\right)^2}{p(x)}\, dx = \mathbb{E}\left[\left(\frac{\partial}{\partial x}\log p(x)\right)^2\right]. \tag{11}$$

The equation describing our synthetic signal is,

$$x(t) = g(t) \sum_{k=1}^{K_n} \mathbf{Re}\left(A_k\, e^{-i(2\pi f_k t + \phi_k)}\right) + (1 - g(t)) \sum_{k=1}^{K_a} \mathbf{Re}\left(A_k^{(a)}\, e^{-i\left(2\pi f_k^{(a)} t + \phi_k^{(a)}\right)}\right) + \epsilon(t), \tag{12}$$

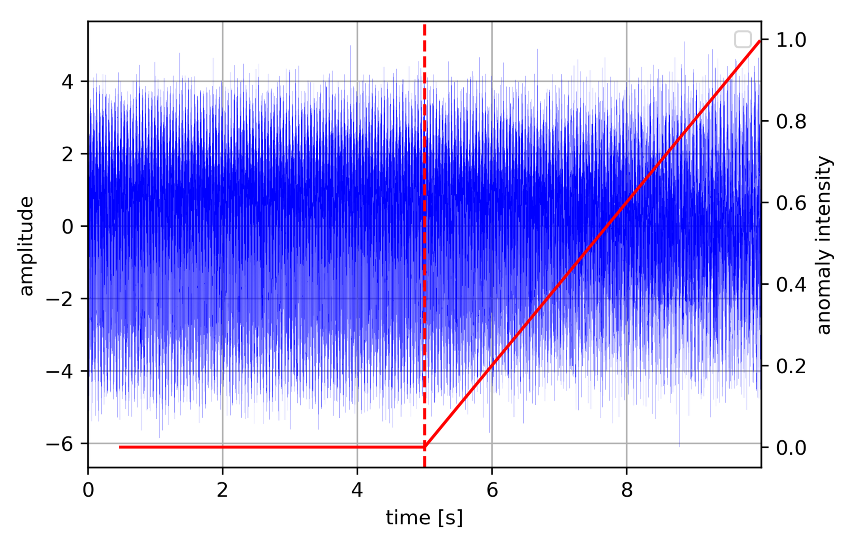

Figure: Linear increasing anomaly.

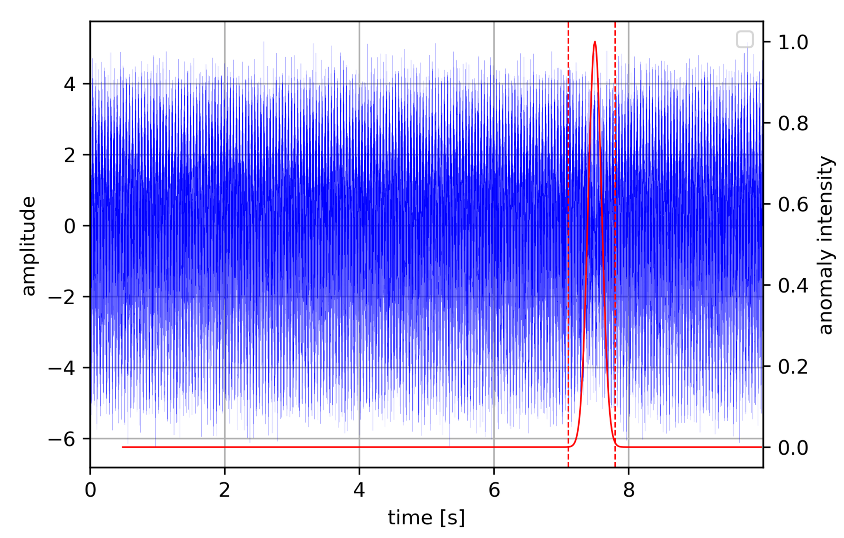

Figure: Gaussian localised anomaly.

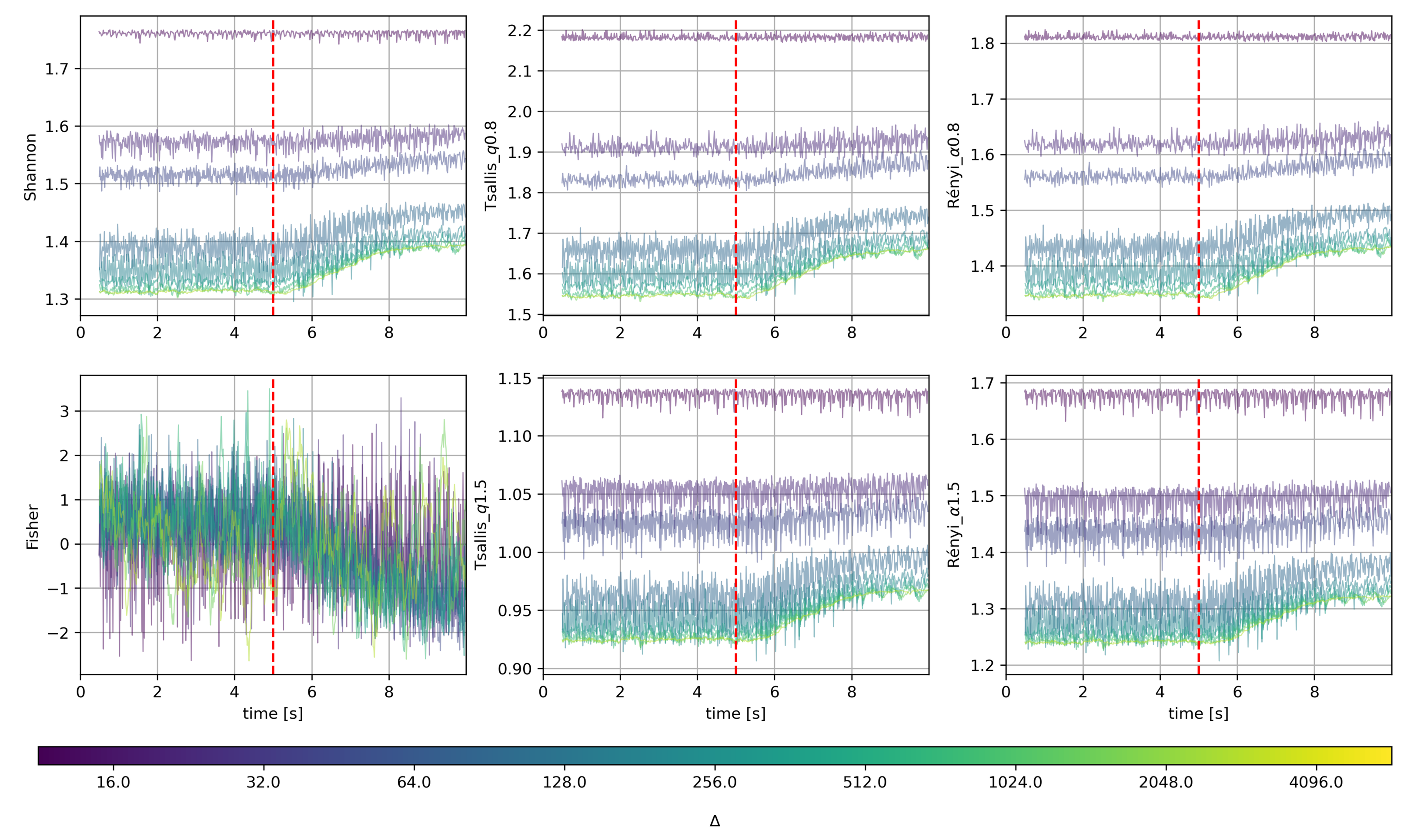

Figure: Entropic and information time-dependent plots related to linear increasing anomaly.

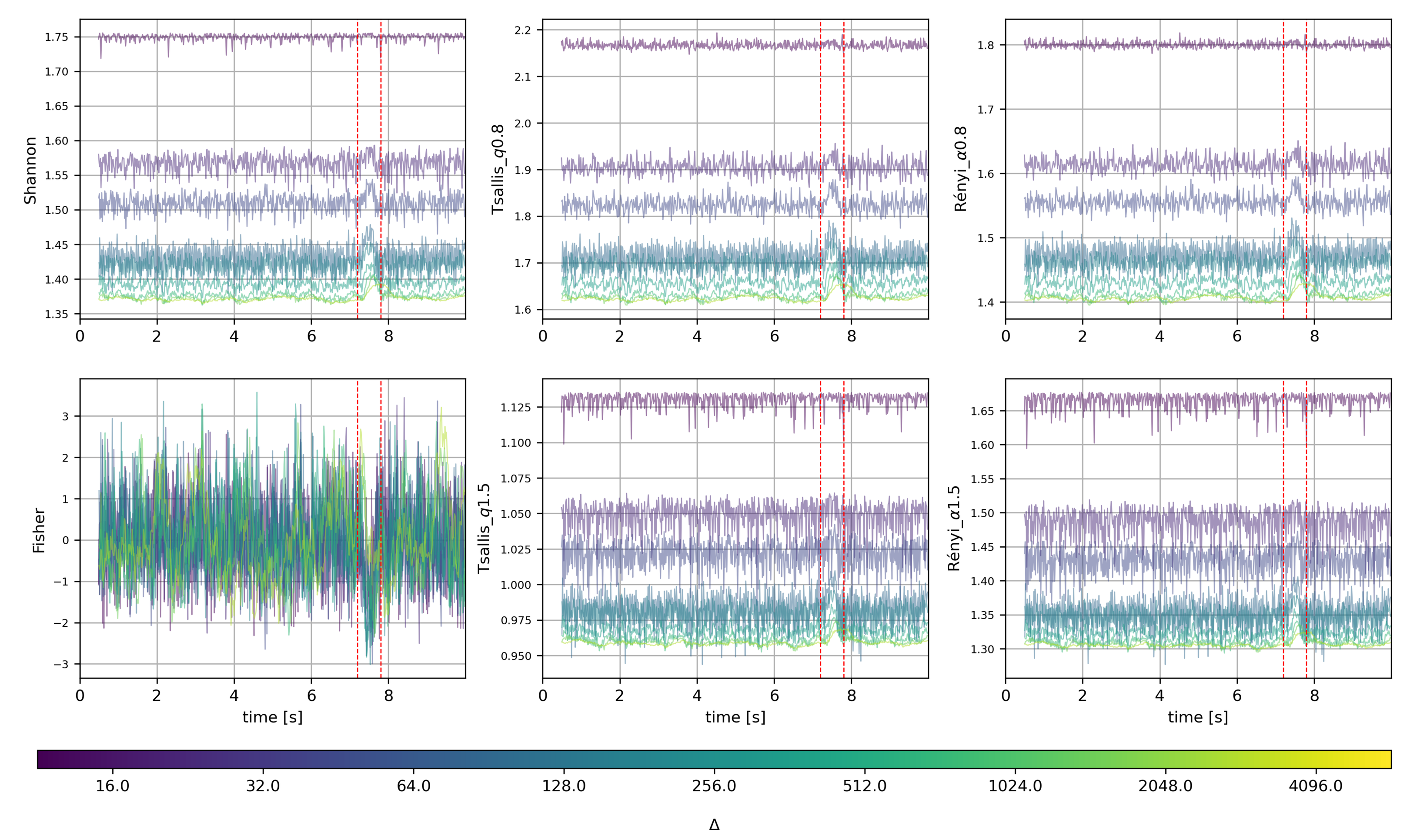

Figure: Entropic and information time-dependent plots related to Gaussian localised anomaly.

5) Conclusions

- The proposed methodology effectively describes different types of anomalies.
- Analyzing at different scales is crucial to increasing the sensitivity of the transformation.
- A new bandwidth optimization algorithm is proposed that:
  - Balances the overfitting-underfitting tradeoff given a sequence of distributions, with quantitative metrics, crucial for consistent results across different scales.
  - Does not make any assumptions about the data distribution (no i.i.d. is required).
  - Can be optimized offline within the anomaly detection scope, given reference time signals.

REFERENCES

- Thomas M Cover and Joy A Thomas. ELEMENTS OF INFORMATION THEORY. page 774, 2006.
- Fabian Guignard, Mohamed Laib, Federico Amato, and Mikhail Kanevski. Advanced Analysis of Temporal Data Using Fisher-Shannon Information: Theoretical Development and Application in Geosciences. *Frontiers in Earth Science*, 8:255, July 2020.
- Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962.
- Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pages 547–562. University of California Press, 1961.
- O.A. Rosso, M.T. Martin, A. Figliola, K. Keller, and A. Plastino. EEG analysis using wavelet-based information tools. *Journal of Neuroscience Methods*, 153(2):163–182, June 2006.
- Constantino Tsallis. Entropic nonextensivity: A possible measure of complexity. *Chaos, Solitons & Fractals*, 13(3):371–391, March 2002.