Informational Analysis of MODIS Satellite Evapotranspiration Data of Vegetation Cover: a Method to Reveal the Presence of Plant Diseases

Luciano Telesca, Rosa Lasaponara
Institute of Methodologies for Environmental Analysis, National Research Council
C.da S.Loja, 85050 Tito (PZ), Italy
email:luciano.telesca@imaa.cnr.it; email:rosa.lasaponara@imaa.cnr.it

Farid Faridani
Department of European and Mediterranean Cultures, Environment, and Cultural Heritage, University of Basilicata
85100 Potenza, Italy, email:farid.faridani@unibas.it

Nicodemo Abate
Institute of Heritage Science, National Research Council, C.da S. Loja, 85050 Tito Scalo, Italy
email:abate.nicodemo@gmail.com

Michele Lovallo
ARPAB, 85100 Potenza, Italy, email:michele.lovallo@arpab.it

***Abstract*—The main goal of this paper is the evaluation of the potential of Fisher-Shannon statistical method applied to MODIS evapotranspiration satellite time series to explore the inner time dynamics of vegetation cover. In particular, we focused on two types of vegetation areas, peri-urban parks and olive orchards. For the first, we selected five sites in Italy, one of which (Castel Volturno) is affected by *Toumeyella Parvicornis*, a parasite that has been adversely impacting the Pinus trees of that area in the recent years. For the second, we selected seven sites in Southern Italy, four of which are affected by *Xylella Fastidiosa*, considered one of the most dangerous phytopathogenic bacteria in the world. For all the investigated sites, to remove the trend and seasonal variability, we firstly applied the Singular Spectrum Analysis (SSA); then, we analysed the de-trended series by means of the Fisher-Shannon statistical method, which combines the Shannon Entropy Power (SEP) and the Fisher Information Measure (FIM). In the Fisher-Shannon Information Plane (FSIP), the infected vegetated areas appear well characterized by the lowest FIM and the highest SEP. These preliminary results seem to envisage the usefulness of the Fisher-Shannon method as a reliable statistical tool to be included in an operational system for early diagnosis of status of deterioration of vegetation.**

Keywords-Fisher-Shannon method; singular spectrum analysis; MODIS; vegetation.

I. INTRODUCTION

With the worsening of climate change and the increasing of global trade, plant diseases have been accelerating in outbreaking and spreading out. Invasive pests and alien plant bacteria are considered one of the major threats worldwide, because they can induce serious plant diseases with devastating impacts on both natural ecosystems and agriculture production with huge environmental (loss of biodiversity) and economic damage. For instance, *Xylella Fastidiosa*, considered one of the most dangerous plant bacteria in the world, causes a number of devasting diseases of significant economic importance in many crops as, but not only, grapevine, Citrus, Olive trees etc. As an example, in the EU only considering the impact on olive trees, it has been estimated that this bacterium has the potential of causing an annual production loss of 5.5 billion euros, affecting 70% of the EU production value of older olive trees. Thus, detecting, quantifying and identifying plant diseases is extremely crucial for assessing tempestive measures to contrast them [1].

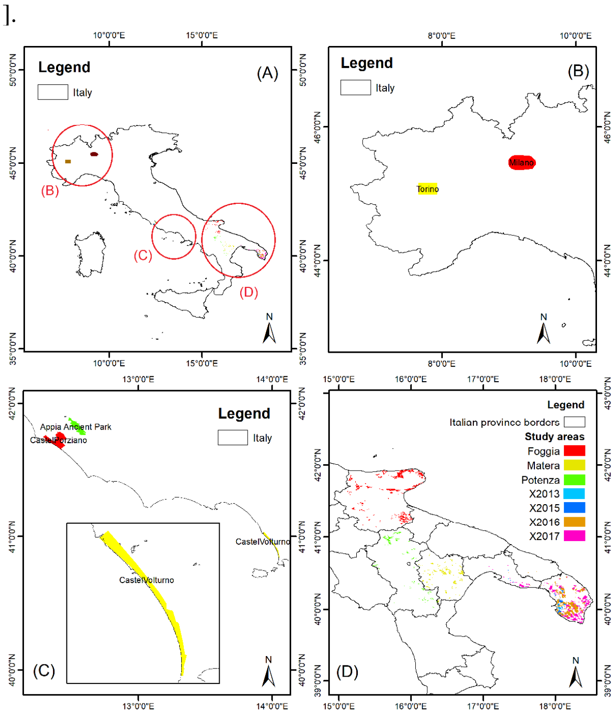


Figure 1. Study areas (A); peri-urban parls (B, C); olive orchards (D).

In the recent years, Remote Sensing (RS) approaches have been gaining special attention in monitoring vegetation dynamics resulting, among the others, from plant diseases

 ISBN: 978-1-68558-057-5

[2]. Several RS applications in phytopathology have been focused on the development of methodologies based on multi-temporal and multi-spectral satellite data for monitoring land-cover changes. Statistical approaches, such as principal component analysis [3] and curve fitting methods [4], are well known for detecting vegetation changes of land surface.

In this paper, we present a statistical approach, namely the Fisher-Shannon (FS) method, to capture evidence of the presence of plant diseases. The FS method relies on the informational content of a time series and, in our case, is used to analyse the time dynamics of MODIS Evapotranspiration (ET) satellite data of different vegetation covers, affected by *Toumeyella Parvicornis* and by *Xylella Fastidiosa*.

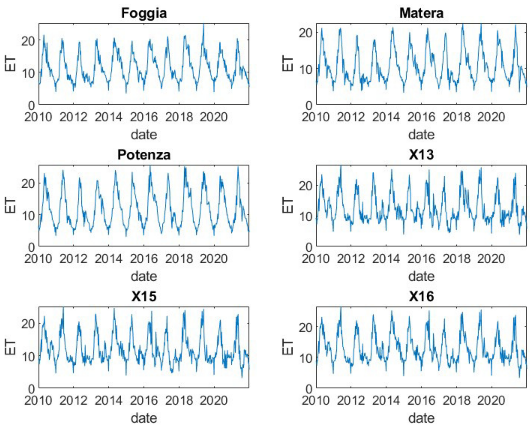
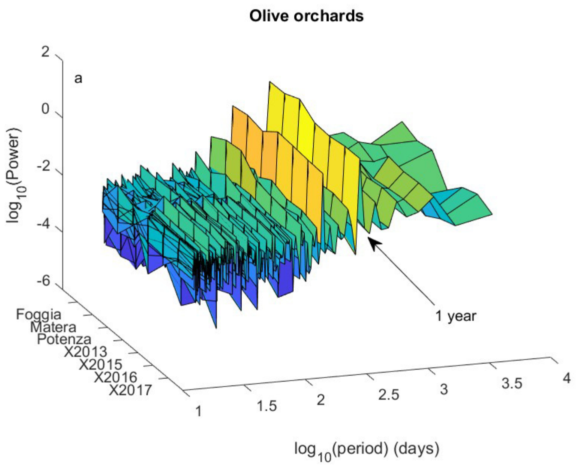


Figure 2. An example of evapotranspiration data.

The MODIS ET product is based on the logic of the Penman-Monteith equation, which includes inputs of daily meteorological reanalysis data along with satellite information. It is expected that ET will suitably characterize and capture the impact of plant infected by *Toumeyella Parvicornis* and by *Xylella Fastidiosa*, since one of the recognizable effects is that the plant dries up and dies.



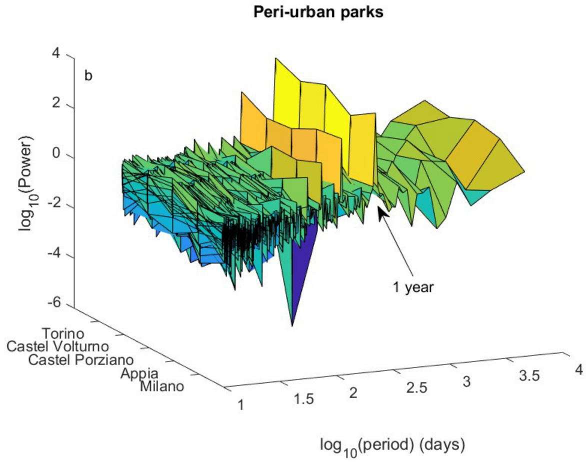


Figure 3. Power spectra of ET time *series* of olive orchards (a) and peri-urban parks (b).

II. DATA AND METHODS

For the purpose of this study, five peri-urban parks in Italy were selected, Milano, Torino, Appia, Castel Porziano and Castel Volturno, the last one attacked by the *Toumeyella Parvicornis* since 2015. Furthermore, seven olive orchard areas were selected, Foggia, Potenza, Matera, X2013, X2015, X2016 and X2017, the last four located in Southern Apulia and infected by *Xylella Fastidiosa* in different periods from 2013 to 2017 (Figure 1). For each site, one MODIS-based ET time series was obtained by averaging the ET values of all the 500m resolution pixels covering each investigated site. The sampling time of the MODIS ET satellite data is 8 days. Some examples of the analysed MODIS ET time series are shown in Figure 2.

The statistical approach used in investigating the data is composed by two steps: the singular spectrum analysis and the Fisher-Shannon method, described in the following subsections.

A. *Singular Spectrum Analysis*

The decomposition of a time series into independent components can be performed by using several techniques, among which the Singular Spectrum Analysis (SSA) [5] represents an efficient and well-known tool based on phase-lagged copies of the series.

The independent components obtained by means of the SSA can be easily recognizable as slowly changing trend, oscillatory components and structureless noise [6].

Let us consider a time series y_i (i =1, .., N) and a lag M, then the Toeplitz lagged correlation matrix can be constructed:

$$c_{ij} = \frac{1}{N-|i-j|}\sum_{k=1}^{N-|i-j|} y_k y_{k+|i-j|}\,, 1 \leq i,j \leq M \qquad (1)$$

Sorting its eigenvalues λ_k in decreasing order, the corresponding eigenvectors $E_{k,j}$ where j and k vary from 1 to M, are used to calculate the k-th principal component i

 ISBN: 978-1-68558-057-5

$$a_{ik} = \sum_{j=1}^{M} y_{i+j} E_{jk} \qquad (2)$$

for $0 \leq i \leq N\text{-}M$, and the k-th reconstructed component of the time series:

$$R_k = \frac{1}{M} \sum_{j=1}^{M} a_{i-j,k} E_{jk} \qquad (3)$$

for $M \leq i \leq N\text{-}M\text{+}1$. Since the eigenvalue λ_k represents the fraction of the total variance of the original series explained in k-th reconstructed component R_k, the decreasing order of the eigenvalues also reflects the decreasing order of the reconstructed components by the fraction of the total variance of the series [7]. SSA requires that the lag M is properly selected, on the base of a trade-off between the quantity of information extracted (large M) and the degree of statistical confidence in that information (large ratio N/M). Khan and Poskitt [8] calculated the maximum $M = (\log N)^c$, $1.5 \leq c \leq 2.5$.

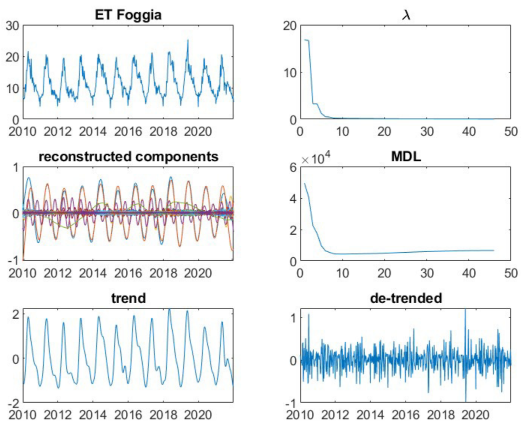


Figure 4. Application of SSA to Foggia MODIS ET data.

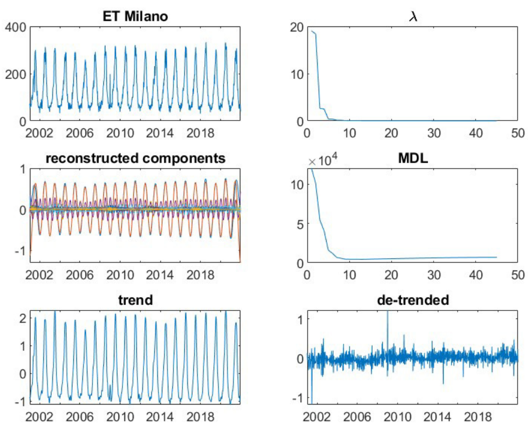


Figure 5. Application of SSA to Milano MODIS ET data.

The minimum description length (MDL) criterion [9]

$$MDL(k) = -\log\left(\frac{\prod_{i=k+1}^{p} \lambda_i^{\frac{1}{p-k}}}{\frac{1}{p-k}\sum_{i=k+1}^{p} \lambda_i}\right)^{(p-k)N} + \frac{1}{2} k(2p-k)\log N \qquad (4)$$

is used to separate the series into two parts that we can define as trend and detrended series; λ_k are the eigenvalues, p is the number of eigenvalues, identical to M, and N is the length of the original series. The separation occurs at the value of $k \in \{0, 1, 2, \ldots, p\text{-}1\}$ for which the MDL is minimized.

B. Fisher-Shannon method

The informational properties of a time series can be analysed by the Fisher Information Measure (FIM) and the Shannon entropy (SE) that quantify respectively the local and global smoothness of the distribution of a series. The FIM and SE can be utilized for characterizing the complexity of non-stationary time series described in terms of order and organization [10]. The FIM measures the order and organization of the series, and the SE its uncertainty or disorder [11]. The FIM and SE are defined by the following formulae:

$$\mathrm{FIM} = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial x} f(x)\right)^2 \frac{dx}{f(x)}, \qquad (5)$$

$$\mathrm{SE} = -\int_{-\infty}^{+\infty} f_X(x) \log f_X(x)\, dx, \qquad (6)$$

where $f(x)$ is the distribution of the series x. Instead of SE, it is generally used the Shannon entropy power (SEP) N_X

$$N_X = \frac{1}{2\pi e} e^{2SE}, \qquad (7)$$

that is defined positive. FIM and N_X are not independent of each other due to the isoperimetric inequality $\mathrm{FIM} \cdot N_X \geq D$ [12], where D is the dimension of the space (1 for time series).

FIM and N_X depend on $f(x)$, whose accurate estimation is crucial to obtain reliable values of informational quantities. For calculating FIM and N_X we applied the kernel-based approach that is better than discrete-based approach in estimating the probability density function [13]. Thus applying the kernel density estimator method for $f(x)$ [14], [15] as shown in the following formula:

$$\hat{f}_M(x) = \frac{1}{Mb} \sum_{i=1}^{M} K\left(\frac{x - x_i}{b}\right), \qquad (8)$$

where M and b denote the length of the series and the bandwidth respectively, while $K(u)$ is the kernel that is a continuous, symmetric and non-negative function satisfying the two following constrains:

$$K(u) \geq 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} K(u)\,du = 1, \qquad (9)$$

 ISBN: 978-1-68558-057-5

$f(x)$ is estimated by means of an optimized integrated procedure using the algorithms of Troudi et al. [16] and Raykar and Duraiswami [17] with the Gaussian kernel:

$$\hat{f}_M(x) = \frac{1}{M\sqrt{2\pi b^2}} \sum_{i=1}^{M} e^{-\frac{(x-x_i)^2}{2b^2}} \quad (10)$$

Due to the isoperimetric inequality, the Fisher-Shannon information plane (FSIP), which has the N_X as x-axis and FIM as y-axis, represents a very useful tool to investigate the complexity of time dynamics of signals [18]. For scalar signals, the curve FIM·N_X=1 separates the FSIP into two parts, and each signal can be represented by a point located only in the space FIM·N_X>1.

III. RESULTS

The SSA requires that the phase lag M is selected to capture the main periodicities of the series. Thus, we firstly calculated the power spectrum of each ET time series (Figure 2) and identified the annual cycle as the main periodicity.

Thus, to detect at least the annual cycle, M was set as 46, consistently with the sampling time of the data, which is 8 days. As an example, Figure 4 and Figure 5 show the application of the SSA to the ET time series of two sites. After normalized the series, the SSA eigenvalue spectrum λ was obtained along with the reconstructed components; each eigenvalue represents the contribution of the corresponding component to the total variance of the original series. The behaviour of the reconstructed components varies from oscillatory trend with amplitude modulation to seemingly noisy. Applying the MDL criterion the signal is separated into a trend and a de-trended series; the value of k_{min} corresponding to the minimum MDL represents the number of the first reconstructed components to sum up for obtaining the trend. For the series Foggia, for instance, k_{min}=9; thus, the trend is obtained summing up the first 9 reconstructed components and the de-trended series by subtracting the trend from the original normalized series. Table I and Table II show the SSA parameters (phase lag M and k_{min}) used for each time series.

The trend is featured by an oscillatory behaviour and represents the seasonal cycles of meteo-climatic origin. The de-trended series, although apparently noisy, represents the inner time dynamics of the series that might not be influenced by external driving mechanisms. Thus, since our aim is the characterization of the time dynamics of inner vegetation by using the Fisher-Shannon method, for each site we analysed the de-trended series. Figure 6 and Figure 7 show the FSIP of de-trended MODIS ET time series of the peri-urban parks and the olive orchard areas, respectively. The FSIP indicates that among the peri-urban parks Castel Volturno that is effectively attacked by the *Toumeyella Parvicornis* is characterized by the lowest FIM and the highest SEP. Among the olive orchards, the four sites X2013, X2015, X2016 and X2017 that were infected by *Xylella Fastidiosa* occupy the bottom-right part of the FSIP, indicating a higher level of disorder and lower level of organization of the vegetation index, similarly to Castel Volturno.

TABLE I. SSA PARAMETERS USED FOR PERI-URBAN PARKS

Peri-urban parks		
Site	***M***	***k_{min}***
Torino	46	7
Castel Volturno	46	5
Castel Porziano	46	7
Appia	46	11
Milano	46	12

TABLE II. SSA PARAMETERS USED FOR OLIVE ORCAHRDS

Olive orchards		
Site	***M***	***k_{min}***
Foggia	46	9
Matera	46	8
Potenza	46	8
X2013	46	7
X2015	46	7
X2016	46	6
X2017	46	6

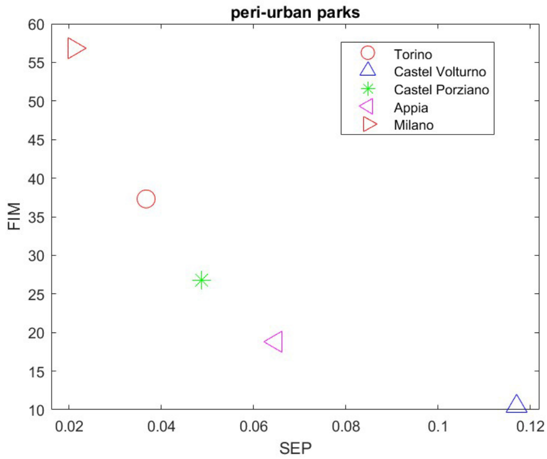


Figure 6. Fisher-Shannon Information Plane for the de-trended MODIS ET data of peri-urban parks.

IV. DISCUSSION

The fatality rate for both *Toumeyella Parvicornis* (affecting pines) and *Xylella* Fastidiosa (affecting olive orchads) is as high as 100%, and their early detection is the critical issue to eradicate the disease and stop tree mortality. Therefore, the main question is how to quickly find the infected trees?

From the operational point of view, the existing solutions are only based on in situ analysis and visual inspection, and, therefore, are not suitable to identify early signals of plant diseases not visible at a naked eye.

Earth Observation (EO) technologies provide, instead, imaging beyond the visible and therefore much more information than those obtained solely from the ground. Moreover, EO undoubtly offer cost-effective tools for monitoring wide areas at both local and global scale.

Nevertheless, previous studies based on satellite EO or drone surveys did not utilize analyses of long term time series that enabled the identification of early signals of degradation. Moreover, the use of evapotranspiration time series as proxy indicator of plant conditions also improved the early detection capability (and facilitate the forecasting of pest outbreak) that is the critical issue to eradicate destructive disease and pest infestations, as those from both for *Xylella* and *Toumeyella Parvicornis*.

There is a strong requirement for reliable operational tools for multiscale, multi-sensor, multi-temporal monitoring of biophysical parameters relating to the state of vegetation to assess and monitor land degradation and capture early signs of both degradation productivity declines and related temporal dynamics that often precede tree mortality years to decades before death. In more details, the following tasks would be useful to implement: i) setting up of EO-based metrics/indicators suitable for an early diagnosis of vegetation deterioration trends to improve ability to forecast tree disease and mortality from local up global scale; ii) effective satellite-based near real time monitoring of forest disease and pest damage for the development of prevention and control strategies.

V. CONCLUSIONS

The vegetation of several study areas from the North to the Southern part of Italy was analysed. The study areas were peri-urban parks and olive orchards. The peri-urban parks were selected as key in improving environmental quality, being rich in biodiversity and allowing urban areas to be more sustainable, helping to combat climate change and make cities more comfortable. The olive orchards were selected as extremely important for the economy of Southern Italy; in fact, Apulia (were some of the investigated sites are located) accounts for about 40% of Italy's olive oil production.

Thus, for each site we focused on the SSA de-trended series since this represents the inner time dynamics of the vegetation.

Our findings point out to the following results: (i) the trend of each series is characterized by an oscillatory behaviour that might be linked with the meteo-climatic cycles, (ii) the de-trended series, although apparently noisy, might be not influenced by external driving mechanisms; (iii) among the investigated peri-urban parks, Castel Volturno, and among the olive orchards, X2013, X2015, X2016 and X2017 are characterized by the lowest FIM and the highest SEP; (iv) Castel Volturno, X2013, X2015, X2016 and X2017 share similar phytopathogenic conditions, which is induced by *Toumeyella Parvicornis* for Castel Volturno and *Xylella Fastidiosa* for the remaining four sites; (v) a plant disease seems to be well revealed by analysing the informational properties of MODIS ET time series.

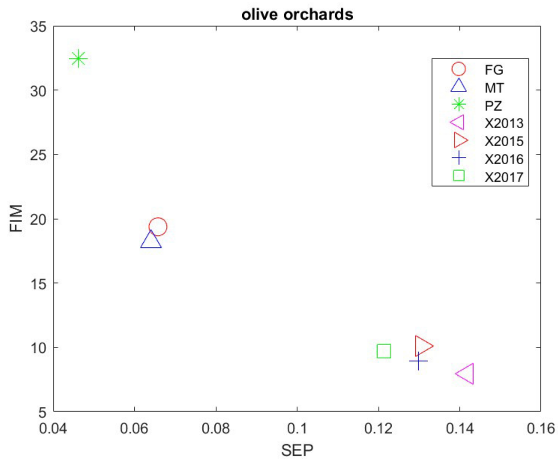


Figure 7. Fisher-Shannon Information Plane for the de-trended MODIS ET data of olive orchards.

Our results could contribute to the definition of methodologies able to diagnose the deterioration and operational tools for the monitoring of biophysical parameters of the status of vegetation.

ACKNOWLEDGMENT

This work was supported by the project COELUM funded by CNR.

REFERENCES

[1] N. Zhang, et al. "A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades," Remote Sens., vol. 12, 3188, 2020.

[2] D. S. Reddy, and P. R. C. Prasad, "Prediction of vegetation dynamics using NDVI time series data and LSTM. Model," Earth Syst. Environ., vol. 4, 409–419, 2018.

[3] M. Hall-Beyer, "Comparison of single-year and multiyear ndvi time series principal components in cold temperate biomes," IEEE Trans. Geosci. Remote Sens., vol. 41, 2568–2574, 2003.

[4] G. L. Galford, J. F. Mustard, J. Melillo, A. Gendrin, and C. E. P. Cerri, "Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil.," Remote Sens. Environ., vol. 112, 576–587, 2008.

[5] R.Vautard, and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," Phys. D, vol. 35, 395–424, 1989.

[6] H. Hassani, "Singular Spectrum Analysis: Methodology and Comparison," J. Data Sci., vol. 5, 239–257, 2007.

[7] D. Schoellhamer, "Singular spectrum analysis for time series with missing data, " Geophys. Res. Lett., vol. 28, 3187–3190, 2001.

[8] M. Khan, and D. S. Poskitt, D.S. "Description Length Based Signal Detection in singular Spectrum Analysis," Monash Econometrics and Business Statistics Working Papers 13/10;

Monash University, Department of Econometrics and Business Statistics: Melbourne, Australia, 2010.

[9] M. Wax, and T. Kailath, “Detection of signals by information theoretic criteria,” IEEE Trans. Acoust. Speech Signal Process., vol. 33, 387–392, 1985.

[10] R. A. Fisher, “Theory of Statistical Estimation,” Math. Proc. Camb. Philos. Soc., vol. 22, 700–725, 1925.

[11] C. E. Shannon, “A Mathematical Theory of Communication,” Bell Syst. Tech. J., vol. 27, 379–423, 1948.

[12] K. D. Sen, J. Antolín, and J. C. Angulo, “Fisher-Shannon Analysis of Ionization Processes and Isoelectronic Series,” Phys. Rev. A, vol. 76, 032502, 2007.

[13] L. Telesca, and M. Lovallo, “On the performance of Fisher Information Measure and Shannon entropy estimators,” Physica A, vol. 484, 569–576, 2017.

[14] A. Janicki, and A. Weron, Simulation and Chaotic Behavior of Alpha-Stable Stochastic Processes,” Chapman & Hall/CRC Pure and Applied Mathematics; CRC Press: Boca Raton, FL, USA, 1993.

[15] L. Devroye, “A Course in Density Estimation; Progress in Probability,” Birkhäuser Boston Inc.: Cambridge, MA, USA, 1987.

[16] M. Troudi, A. M. Alimi, and S. Saoudi, “Analytical Plug-In Method for Kernel Density Estimator Applied to Genetic Neutrality Study,” EURASIP J. Adv. Signal Process., vol. 2008, 739082, 2008.

[17] V. C. Raykar, and R. Duraiswami, “Fast optimal bandwidth selection for kernel density estimation,” in Proceedings of the 2006 SIAM International Conference on Data Mining; Proceedings; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2006.

[18] C. Vignat, and J.-F. Bercher, “Analysis of Signals in the Fisher–Shannon Information Plane,” Phys. Lett. A, vol. 312, 27–33, 2003.