Phytopathogenic Status Induced by Xylella Fastidiosa in Olive Groves in Southern Italy Revealed by Visibility Graph Analysis of MODIS Satellite Evapotranspiration Time Series

Luciano Telesca and Rosa Lasaponara
Institute of Methodologies for Environmental Analysis
National Research Council
Tito, Italy
email: luciano.telesca@cnr.it; rosa.lasaponara@cnr.it

***Abstract*—The Visibility Graph (VG) has emerged as a widely used statistical method for characterizing the dynamical properties of time series. This method transforms time series into networks, wherein nodes represent the series values connected by their reciprocal "visibility." Recently, the VG has found applications in the statistical investigation of time series across various research fields. In this study, we leverage the VG to analyze the topological properties of Moderate Resolution Imaging Spectroradiometer (MODIS) satellite evapotranspiration time series in areas covered by olive groves in southern Italy, aiming to detect the presence of *Xylella Fastidiosa*—a highly destructive phytobacterium known for inducing olive quick decline syndrome in olive trees. Our findings suggest that employing the VG enables a very efficient discrimination between infected and healthy sites. This suggests the potential utility of this network analysis method as an operational tool for early diagnosis of plant deterioration caused by *Xylella fastidiosa*.**

Index Terms*—*Visibility graph; network analysis; MODIS; evapotranspiration; Xylella Fastidiosa.

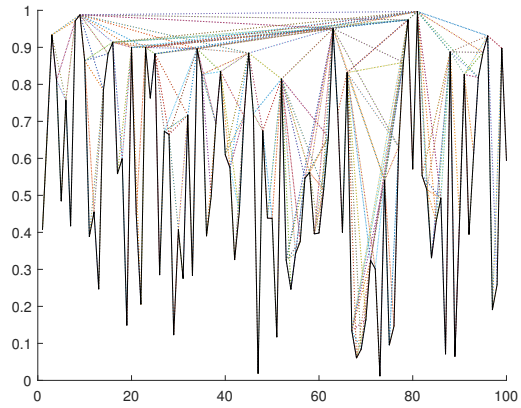

Fig. 1: Representations of the VG for a random signal.

I. INTRODUCTION

Invasive pests and alien plant bacteria emerge as significant global threats, capable of triggering severe plant diseases with profound consequences for both natural ecosystems and agricultural productivity. One such example is *Xylella Fastidiosa*, recognized as one of the most dangerous plant bacteria worldwide. This pathogen induces devastating diseases, particularly affecting grapes, citrus fruits, and olive trees. In the European Union, focusing solely on the impact on olive trees, estimates suggest that *Xylella Fastidiosa* can potentially cause an annual production loss of 5.5 billion euros, impacting 70% of the EU production value of mature olive trees. Hence, the timely detection of plant diseases induced by this bacterium become imperative for implementing measures to mitigate their impact [1].

In 2013, the presence of *Xylella Fastidiosa* was first identified in southeastern Italy [2]. Subsequently, it rapidly spred out to various other European countries, giving rise to a critical phytosanitary emergency [3].

Visual inspection stands as the predominant method of contrasting *Xylella Fastidiosa* due to its simplicity and cost-effectiveness; but its precision strongly depends on the subjective evaluation of disease severity.

Recently, Remote Sensing (RS) approaches have been given significant attention, particularly in the monitoring of vegetation dynamics, including those induced by plant diseases [4]. Various RS applications in phytopathology have concentrated on devising methodologies based on multi-temporal and multi-spectral satellite data for monitoring changes in land cover. Statistical approaches, such as principal component analysis [5] and curve-fitting methods [6], are well-established for detecting alterations in vegetation across the land surface.

The potential use of MODIS evapotranspiration (ET) satellite data for monitoring pest and parasite attacks at both landscape and field scales has been investigated in [7], [8], where the effectiveness of MODIS ET and other vegetation

indices in assessing the deterioration of pine tree vegetation caused by the parasite *Toumeyella Parvicornis* was evidenced.

Furthermore, the application of multifractal detrended fluctuation analysis and the informational method of Fisher-Shannon on MODIS ET time series indicated the suitability of this satellite-derived data as a reliable indicator for assessing the pathogenic impact caused by *Xylella Fastidiosa* in olive trees [9]–[11].

In this paper, we apply the VG method to discern the presence of diseases induced by *Xylella Fastidiosa*. The VG method is used to characterize the topological properties of a signal and, in our context, is employed to analyze the temporal dynamics of MODIS ET satellite data of olive groves affected by *Xylella Fastidiosa*.

In Section II the applied methods of time series analysis will be described and in Section III the investigatred dataset will be presented. Section IV will show the obtained results that will be discussed in Section V.

II. METHODS

A. The visibility graph

The VG [12] converts a time series into a graph or network. It was applied in several research fields, like economics [13]–[15], weather forecasting [16], [17], medicine [18], oceanography [19], etc. The nodes of the graph are the values of the series and the links between them satisfies the following geometrical visbility rule:

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a}, \tag{1}$$

where $t_a < t_c < t_b$. In practice, two values $y_a(t_a)$ and $y_b(t_b)$ are visible to each other if any other value $y_c(t_c)$ fulfils the condition from 1 (Figure 1 shows the application of the VG to a random signal). The adiancency matrix associated to VG is given by

$$A_{pq} = \begin{cases} 1 & \text{if nodes p and q are linked} \\ 0 & \text{if nodes p and q are not linked} \end{cases} \tag{2}$$

The visibility graph associated with a time series is characterized by the following properties: 1) connectivity: every node is connected to at least its closest neighbors on both the left and right sides; 2) undirectionality: the links between nodes do not have a defined direction; 3) invariance under affine transformations: the visibility graph remains unchanged when the time series data undergoes rescaling along both horizontal and vertical axes, as well as horizontal and vertical translations.

The connectivity degree $k_i = \sum_j A_{ij}$ is the number of links departing from a node i. Generally, the higher values of the signal are characterized by the larger connectivity degree, because they behave as hubs of the graph "attracting" more links than those converging to the lower ones.

The degree distribution $P(k)$ is an important quantity that explains some of the structure of the time series. For instance, it is well known that $P(k)$ behaves as an exponential or power-law function of the degree k for random or fractal time series, respectively [12]; while for a periodic series it is formed by a finite number of non-null values, with this number depending on the period [21].

B. The ROC analysis

Receiver Operating Characteristics (ROC) analysis is utilized to evaluate the performance of classifiers. In binary classification scenarios, instances are classified as either "positive" or "negative," and a classifier assigns these instances to predicted classes [20].

When assessing a classifier with respect to an instance, four potential outcomes can occur. The categorization of the instance is as follows: True Positive (TP) if it is positive and correctly classified as positive, False Negative (FN) if it is positive but incorrectly classified as negative, True Negative (TN) if it is negative and correctly classified as negative, False Positive (FP) if it is negative but erroneously classified as positive [20]. We can define the following ratios, the *True Positive rate* (TPr) and the *False Positive rate* (FPr)

$$TPr = \frac{Number \quad of \quad TP}{Total \quad positives}, \tag{3}$$

$$FPr = \frac{Number \quad of \quad FP}{Total \quad negatives}. \tag{4}$$

A ROC curve is a graphical representation with TPr plotted on the y-axis and FPr on the x-axis. The construction of the ROC curve proceeds as follows. Defining the two classes of instances as C_1 and C_2, all the values of the chosen parameter are arranged in ascending order. A threshold T is selected within the range of values, from the minimum to the maximum. Supposing that the mean of C_1 is larger than that of C_2, then, a TP is a value of C_1 that is above T, while a FP is a value that is below T. Thus, for each value of the threshold T within the range of minimum to maximum, the TPr and FPr can be calculated, yielding one point on the ROC curve. By varying the threshold value across the entire range and repeating this process, a curve can be traced through ROC space, known as the ROC curve. In ROC space, the point (0, 1) signifies perfect classification, and one point is considered superior to another if it lies to the northwest of the first point. The diagonal line, represented by the equation y = x, corresponds to random classification, while an effective classifier is depicted by a point located in the upper region of the ROC space.

Each point on the ROC curve corresponds to a tradeoff between TPr and FPr associated with a specific threshold. Typically, to optimize this tradeoff, the point on the ROC curve closest to (0, 1) is chosen, and the corresponding threshold is utilized for classification. Also, the Area Under the ROC Curve (AUC) is frequently employed to quantify the classifier's performance.

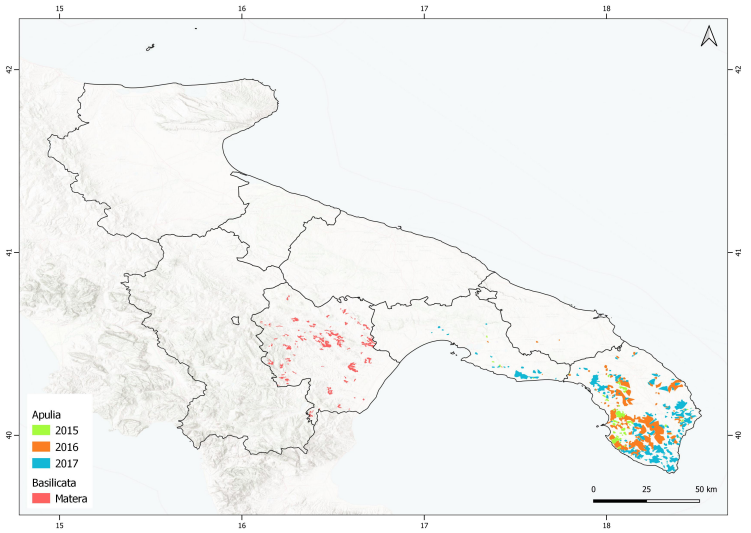

Fig. 2: Investigated area with Matera (unnifected) and X2015, X2016 and X2017 (infected) sites.

III. Data and Study Area

The primary visible indication of an infection caused by *Xylella Fastidiosa* in olive trees manifests as the desiccation of branches [22]. We analysed the MODIS ET data that possess the capability to monitor the water status of plants, enabling the detection of symptoms associated with this disease. The data exhibit a spatial resolution of 500m and an 8-day sampling rate. They are readily accessible online [23] and are also present in the Google Earth Engine (GEE) cloud database.

The ET is calculated summing up soil evaporation (E_s), canopy evaporation (E_c), and canopy transpiration (T_c):

$$ET = Es + Ec + Tc \tag{5}$$

$$E_s = f_w \frac{\Delta A_S + \frac{(1-f_c)\rho_a C_P (e_S - e_a)}{r_a^s}}{\Delta + \gamma \frac{r_S^S}{r_a^S}} + RH^{\frac{(e_S-e_a)}{\beta_{sm}}}(1-f_W)\frac{\Delta A_S + \frac{(1-f_c)\rho_a C_P (e_S - e_a)}{r_a^s}}{\Delta + \gamma \frac{r_S^S}{r_a^S}} \tag{6}$$

$$E_c = f_w \frac{\Delta A_S + \frac{f_c \rho_a C_P (e_S - e_a)}{r_a^{wc}}}{\Delta + \gamma \frac{r_S^{wc}}{r_a^S}} \tag{7}$$

$$T_c = (1-f_w)\frac{\Delta A_C + \frac{f_c \rho_a C_P (e_S - e_a)}{r_a^t}}{\Delta + \gamma(1 + \frac{r_S^t}{r_a^S})}. \tag{8}$$

where f_c is the canopy cover, f_w is the pixel wet surface fraction, *RH* is the relative humidity, Δ is the gradient of the saturation vapor pressure–temperature, A_s and A_c are the available energy to the soil and canopy, respectively, γ is the psychrometric constant, β_{sm} is a parameter related to the soil moisture constraint, r_s^s and r_a^s are the surface and aerodynamic resistance for the soil surface, r_s^{wc} and r_a^{wc} are the surface and aerodynamic resistance for the wet canopy evaporation and r_s^t and r_a^t are the surface and aerodynamic resistance for the canopy transpiration [24].

TABLE I: NUMBER OF PIXELS BEFORE AND AFTER THE GAP FILLING (in bold)

Matera	X2015	X2016	X2017
417	312	446	451
331	**219**	**334**	**371**

In this study, we investigated three datasets consisting of pixels extracted from olive groves affected by *Xylella fastidiosa*: one dataset corresponding to locations where the infection was identified in 2015 (X2015), another in 2016 (X2016), and the third in 2017 (X2017). As reference, we also analyzed a dataset of pixels from unaffected olive groves situated in the Matera area. This reference area shares comparable climatic and topographic conditions with the infected one. The investigated areas are shown in Figure 2. The ET datasets spans from 2010 to 2022, and each pixel's time series has a length of 575 samples, with missing data percentage less than 25%.

IV. Results

The VG can operate only with complete datasets, and therefore, we firstly addressed data gaps in each pixel time series. This was accomplished by replacing the gap with the average of values measured on the corresponding day from other years. In cases where the gap occurred on the same day across all years, the entire pixel was excluded from the analysis. Therefore, after excluding pixels where at least one gap persisted despite the gap-filling procedure, the number of analysed pixels for each dataset is shown in Table I. Figure 3 illustrates an example of gap-filling for a pixel time series.

We applied the VG (code available at [25]) and calculated the degree distribution for the time series of each pixel, as

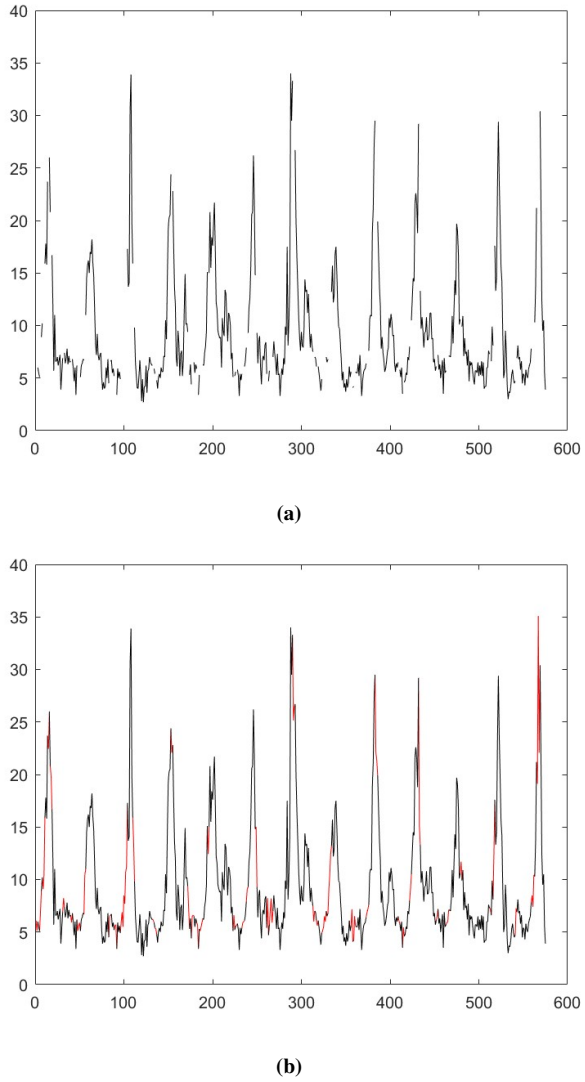

Fig. 3: A pixel time series before (a) and after (b) the gap-filling.

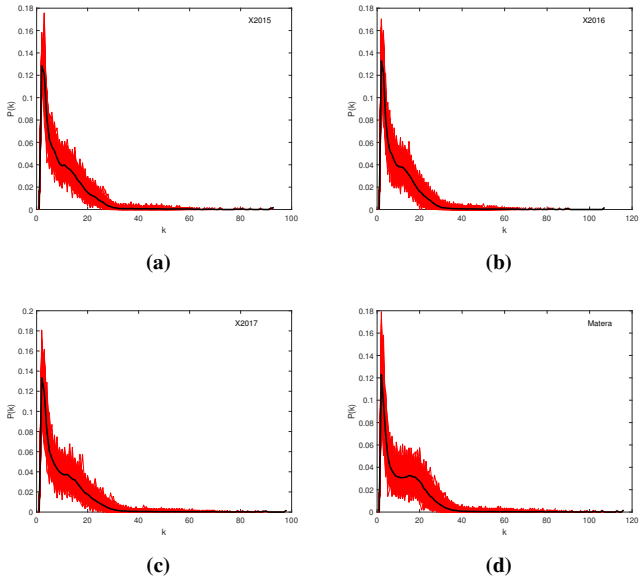

Fig. 4: Degree distributions (red) and their average (black) of the X2015 (a), X2016 (b), X2017 (c) and Matera (d) ET datasets.

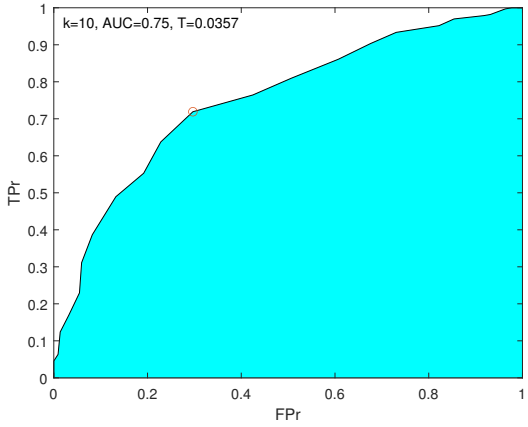

Fig. 5: ROC curve analysis for the degree distribution of the X2015 and Matera datasets at k=10. The AUC is 0.75, with a corresponding threshold of 0.0357, where the TPR is 0.72 and the FPR is 0.30.

depicted in Figure 4, for every dataset, and also shown their respective averages.

To evaluate the classification performance of degree distribution $P(k)$, we employed the ROC curve that represents the relationship between the TPr (the fraction of infected pixels correctly classified as infected) and the FPr (the fraction of uninfected pixels incorrectly classified as infected). By fixing a specific value for the degree distribution k, we performed the ROC analysis for each k value, comparing each infected pixel dataset (X2015, X2016, and X2017) with each uninfected pixel dataset (Matera); this involved calculating the AUC_k, TPr_k, FPr_k, and the threshold T_k for each specific k value.

As an example, Figure 5 presents the ROC curve for the degree distribution value at $k = 10$ for X2015 and Matera datasets. The AUC of 0.75 indicates rather good classification performance between infected pixels in the X2015 dataset and uninfected pixels in the Matera dataset. The optimal threshold, identified at 0.0357, corresponds to TPR of 0.72 and a FPR of 0.30.

TABLE II: RESULTS OF THE ROC ANALYSIS CORRESPONDING TO THE MAXIMUM OF AUC

datasets	ROC parameters				
	k	AUC	TPr	FPr	T
X2015-Matera	21	0.90	0.84	0.20	0.02
X2016-Matera	20	0.88	0.87	0.25	0.02
X2017-Matera	21	0.86	0.84	0.25	0.02

The results of the ROC analysis across various values of the degree k for the datasets X2015 and Matera are displayed in Figure 6. The AUC is generally used to quantify the classification performance. Thus we can select the optimal value of k as corresponding to the maximum AUC. In the

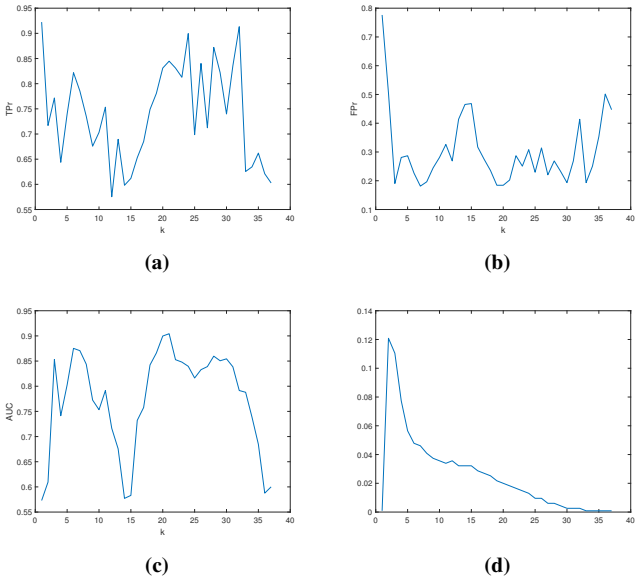

Fig. 6: ROC analysis for the degree distribution of X2015 and Matera datasets: a) TPr, b9 FPr, c) AUC and d) threshold.

specific case, the optimal value of k is determined to be 20, resulting in the following ROC parameters: AUC=0.90, TPr=0.83, FPr=0.18 and threshold T=0.02. A summary of the results obtained applying the ROC analysis to all the datsets are provided inTable II. The optimal value of k for which the AUC is maximized fluctuates between 20 and 21, resulting in AUC values ranging from 0.86 to 0.90, indicative of a highly effective classification performance. The TPr spans from 0.84 to 0.87, while the FPr varies between 0.20 and 0.25. The threshold T remains relatively constant at approximately 0.2.

Since the VG converts the time series into a network, we evaluated the classification performance of some network parameters that are used as network topology metrics. In particular, we analysed the average (μ_k) [26] and the Shannon Entropy (SE) of the connectivity degree [27] defined as:

$$\mathrm{SE} = -\sum_{j=1}^{N_k} P_j(k) \ln P_j(k). \quad (9)$$

Tables III and IV present the outcomes of the ROC analysis conducted on these two parameters. Both parameters exhibit excellent classification performance, with the AUC ranging between 0.93 and 0.97. Additionally, the TPr varies from 0.86 to 0.94, and the FPr ranges from 0.06 to 0.15.

Moreover, focusing on the datasets of infected sites, we observe a slight increase in the averages of both μ_k and SE from 2015 to 2017 (Figure 7).

V. Conclusions

The main factors controlling the evapotranspiration process are influenced by the climate, soil, and vegetation characteristics. In environments where plants inhabit the same ecosystems and encounter similar environmental circumstances, as the cases examined here, differences mainly arise from the varied

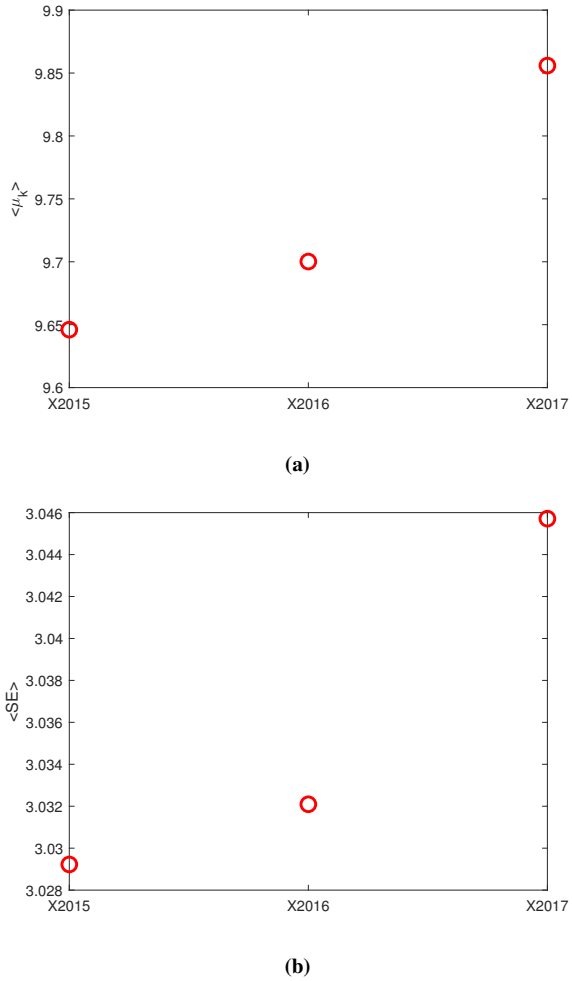

Fig. 7: Averages of a) μ_k and b) Shannon Entropy of the connectivity degree for the infected datasets.

TABLE III: RESULTS OF THE ROC ANALYSIS OF μ_k

datasets	ROC parameters			
	AUC	TPr	FPr	T
X2015-Matera	0.97	0.94	0.07	10.32
X2016-Matera	0.97	0.93	0.08	10.35
X2017-Matera	0.94	0.90	0.15	10.47

TABLE IV: RESULTS OF THE ROC ANALYSIS OF SHANNON ENTROPY

datasets	ROC parameters			
	AUC	TPr	FPr	T
X2015-Matera	0.96	0.92	0.11	3.08
X2016-Matera	0.96	0.88	0.06	3.11
X2017-Matera	0.93	0.86	0.11	3.12

vegetation conditions caused by the presence of *Xylella Fastidiosa* that, since its first detection in Apulia (southern Italy), has affected around 4 million olive trees in the affected region [28]. The presence of this bacterium has led to significant economic losses, particularly in terms of olive trees and oil production, with ongoing repercussions on the economy of the area. Additionally, it has resulted in changing the landscape of the affected area, where olive trees hold deep cultural significance and are a key component of the thriving tourism sector.

The characterization of the impact of *Xylella Fastidiosa* on olive groves is crucial in combating its spread. MODIS ET data appears to offer a valuable tool for achieving this goal, as they enable a differentiation of infected olive orchard areas from those that are uninfected.

Given that one of the main effects of this infection is the fast desiccation of the plant, the utilization of ET seems very suited in effectively characterizing and assessing the impact of this bacterium on plants, since it functions as an indirect indicator of the reduction in water content within plants.

The application of the Visibility Graph to pixels covering both unifected and infected olive groves in southern Italy has suggested to employ classifiers that showed their potential in discriminating between the two classes of pixels, namely the degree distribution, the mean and the Shannon Entropy of the connectivity degree.

It is intriguing to note the rising trend in the average mean connectivity degree and Shannon entropy over time. The datasets for X2015, X2016, and X2017 correspond to sites where the infection was initially identified in 2015, 2016, and 2017, respectively. Therefore, the observed trend in Figure 7 indicates a temporal pattern, implying that the later the plants were infected by the bacterium, the greater the average values of these two network parameters tend to be.

While the results obtained from applying the VG method to MODIS ET time series are still in a preliminary stage, the findings of this study hold significant promise in contributing to the development of operational tools for monitoring vegetation status.

ACKNOWLEDGMENT

This study was supported by the project "Coelum Spies of Climate change and tools for mitigating the effects: EO and AI based methodological approach for Urban Park Management", funded by the National Research Council of Italy.

REFERENCES

[1] https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/monitoring-impacts-xylella-fastidiosa-2019-11-12_en, [retrieved June, 2024]

[2] M. Saponari, D. Boscia, F. Nigro, and G. P. Martelli, "Identification of DNA sequences related to Xylella fastidiosa in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy)," Journal of Plant Pathology, vol. 95, pp. 659–668, 2013.

[3] M. Jeger et al., "Scientific opinion on the updated pest categorisation of Xylella fastidiosa," EFSA Journal, vol. 16, 5357, 2018.

[4] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," Applied and computational harmonic analysis, vol. 27, pp. 265–274, 2009.

[5] J. M. Bioucas-Dias, and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Transactions on Image processing, vol. 16, pp. 2992–3004, 2007.

[6] G. Chen, and D. Needell, "Compressed sensing and dictionary learning," Finite Frame Theory: A Complete Introduction to Overcompleteness, vol. 73, pp. 201-241, 2016.

[7] L. Telesca et al., " Exploring Long-Term Anomalies in the Vegetation Cover of Peri-Urban Parks Using the Fisher-Shannon Method," Entropy, vol. 24, 1784, 2022.

[8] L. Telesca et al., "Urban and Peri-Urban Vegetation Monitoring Using Satellite MODIS NDVI Time Series, Singular Spectrum Analysis, and Fisher–Shannon Statistical Method," Sustainability, vol. 15, 11039, 2023.

[9] L. Telesca, N. Abate, F. Faridani, M. Lovallo, and R. Lasaponara, "Revealing Traits of Phytopathogenic Status Induced by Xylella Fastidiosa in Olive Trees by Analyzing Multifractal and Informational Patterns of MODIS Satellite Evapotranspiration Data," Physica A, vol. 629, 129163, 2023.

[10] L. Telesca, N. Abate, F. Faridani, M. Lovallo, and R. Lasaponara, "Discerning Xylella fastidiosa-Infected Olive Orchards in the Time Series of MODIS Terra Satellite Evapotranspiration Data by Using the Fisher–Shannon Analysis and the Multifractal Detrended Fluctuation Analysis," Fractal and Fractional, vol. 7, 466, 2023.

[11] L. Telesca, N. Abate, M. Lovallo, and R. Lasaponara, "Investigating the Impact of Xylella Fastidiosa on Olive Trees by the Analysis of MODIS Terra Satellite Evapotranspiration Time Series by Using the Fisher Information Measure and the Shannon Entropy: A Case Study in Southern Italy," Remote Sensing, vol. 16, 1242, 2024

[12] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: The visibility graph," Proceedings of the National Academy of Sciences, vol. 105, pp. 4972–4975, 2008.

[13] R. Zhang, B. Ashuri, Y. Shyr, and Y. Deng, "Forecasting construction cost index based on visibility graph: A network approach," Physica A, vol. 493, pp. 239-252, 2018.

[14] Y. Long, "Visibility graph network analysis of gold price time series," Physica A, vol. 392, pp. 3374–3384, 2013.

[15] N. Wang, D. Li, and Q. Wang, "Visibility graph analysis on quarterly macroeconomic series of China based on complex network theory," Physica A, vol. 391, pp. 6543–6555, 2012.

[16] W. Jiang, B. Wei, J. Zhan, C. Xie, and D. Zhou, "A visibility graph power averaging aggregation operator: A methodology based on network analysis," Computers & Industrial Engineering, vol. 101, pp. 260–268, 2016.

[17] S. Chen, Y. Hu, S. Mahadevan, and Y. Deng, "A visibility graph averaging aggregation operator," Physica A, vol. 403, pp. 1–12, 2014.

[18] M. Yu, A. Hillebrand, A. A. Gouw, and C. J. Stam, "Horizontal visibility graph transfer entropy (HVG-TE): A novel metric to characterize directed connectivity in large-scale brain networks," NeuroImage, vol. 156, pp. 249-264, 2017.

[19] L. Telesca, M. Lovallo, and J. O. Pierini, "Visibility graph approach to the analysis of ocean tidal records," Chaos, Solitons & Fractals, vol. 45, pp. 1087–1091, 2012.

[20] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, pp. 861-874, 2006.

[21] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, J., "Horizontal visibility graphs: Exact results for random time series," Physical Review E, vol. 80, 046103, 2009.

[22] M. Saponari, A. Giampetruzzi, G. Loconsole, D. Boscia, and P. Saldarelli, "Xylella fastidiosa in Olive in Apulia: Where We Stand," Phytopathology, vol. 109, pp. 175–186, 2019.

[23] https://lpdaac.usgs.gov, [retrieved June, 2024].

[24] L. Laipelt, R. H. B. Kayser, A. S. Fleischmann, A. Ruhoff, W. Bastiaanssen, T. A. Erickson, and F. Melton, "Long-term monitoring of evapotranspiration using the SEBAL algorithm and Google Earth Engine cloud computing ISPRS," Journal of Photogrammetry and Remote Sensing, vol. 178, pp. 81–96, 2021.

[25] https://www.mathworks.com/matlabcentral/fileexchange/70432-fast-natural-visibility-graph-nvg-for-matlab, [retrieved June, 2024]

[26] G. Bounova, and O. de Weck, "Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles," Physical Review E, vol. 85, 016117, 2012.

[27] B. A. Gonçalves, L. Carpi, O. A. Rosso, and M. G. Ravetti, "Time series characterization via horizontal visibility graph and Information Theory", Physica A, vol. 464, pp. 93–102, 2016.

[28] K. Schneider, W. van der Werf, M. Cendoya, M. Mourits, J. A. Navas-Cortés, A. Vicent, and A. Oude Lansink, "Impact of *Xylella fastidiosa* subspecies *pauca* in European olives,", Proceedings of the National Academy of Sciences of the United States of America, vol. 117, pp. 9250–9259, 2020.