HAL
open science

Spatial pattern analysis of flavescence dorée repartition in vineyards from the Bordeaux region

Florian Rançon, Lionel Bombrun, Barna Keresztes, Christian Germain

To cite this version:

Florian Rançon, Lionel Bombrun, Barna Keresztes, Christian Germain. Spatial pattern analysis of flavescence dorée repartition in vineyards from the Bordeaux region. Journal of Agricultural Informatics, 2017, 8 (2), pp.1-11. hal-01647929

HAL Id: hal-01647929

https://hal.science/hal-01647929v1

Submitted on 24 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. x, No. x
journal.magisz.org

Spatial pattern analysis of flavescence dorée repartition in vineyards from the Bordeaux region

Florian Rançon [1], Lionel Bombrun [2], Barna Keresztes [3], Christian Germain [4]

I N F O
Received
Accepted
Available on-line
Responsible Editor:

Keywords:
Flavescence dorée, Vine-Growing, Epidemics, Geostastistics

A B S T R A C T
The devastating incidence of flavesence dorée on vineyards throughout the world motivates for a better comprehension of this epidemic disease. In this study, we highlight the characteristic spatial non-random distribution of flavescence dorée diseased plants on a set of 7 vineyards from the Bordeaux region. First, we propose a simple statistical framework using Monte-Carlo simulations in order to assess the randomness of the disease repartition. Several statistics are considered such as the mean distance to the nearest diseased neighbor or divergence (using Kullback-Leibler dissimilarity symmetric variant) of the distance histogram to the average distance histogram of random simulations. The performance of these statistics is first evaluated on a set of generated repartitions at different randomness levels using ROC curves as a visual representation of the risks associated with the test. Histogram comparison was found to be more effective and robust for the detection of non-random configurations. The proposed algorithm is then used on real data, showing significant aggregations patterns and edge effect on some of the plots for flavesence dorée diseased plants but also uprooted plants.

1. Introduction

First reported in the french vineyards in the second half of the twentieth century (Caudwell 1957), Flavescence dorée (FD) is a quarantine disease with huge consequences on the vine-growing economy, including yield loss and degraded grape quality. It is caused by a phytoplasma vectored by the leafhopper species Scaphoideus titanus, making it an epidemiologic disease whose threat stems from fast propagation in the vineyard. Thus, insecticide treatment is mandatory in grapegrowing areas where the disease is widespread. Vineyards contaminated at a rate superior than 20% are rogued to stop the disease spreads. Affected vines show summer foliar symptoms depending on the cultivar, including yellowing and leaf curving (in a similar fashion as Bois noir disease) but also delayed or absent august hardening. Depending on the case, a short growth can also be observed during spring. Susceptibility to the disease seems to be highly cultivar-dependent and symptoms may not be shown with equal intensity (Caudwell 1990). FD was reported in most of the vinegrowing countries, including Italia (Vidano 1964), Switzerland (Schaerer et al. 2007) and Spain (Rahola et al. 1997). In France, FD is present in most vinegrowing areas, including the Bordeaux region (Bonfils & Schvester 1960), menacing vineyard durability. Knowledge and forecasting of the disease spread are closely related to its spatio-temporal repartition. Thus, a regional surveillance system may include the analysis of repartition data (collected from ground survey or from remotely-sensed data) to check for potential small aggregated structures that might be problematic for the vineyard over the next years.

[1] Florian Rançon
Laboratoire IMS Université de Bordeaux, Bordeaux Sciences Agro, *florian.rancon@u-bordeaux.fr*
[1] Lionel Bombrun
Laboratoire IMS Université de Bordeaux, Bordeaux Sciences Agro, *lionel.bombrun@agro-bordeaux.fr*
[1] Barna Keresztes
Laboratoire IMS Université de Bordeaux, *barna.keresztes@u-bordeaux.fr*
[1] Christian Germain
Laboratoire IMS Université de Bordeaux, Bordeaux Sciences Agro, *christian.germain@agro-bordeaux.fr*

In this work, we study the patterns associated with the FD presence in vineyards located in La Brède, France, presenting various infection levels and various cultivar combinations, focusing here on the spatial analysis of FD repartition in the end of summer 2016. For this purpose, expert ground survey indicating the precise locations of symptomatic plants is considered as a basis for repartition analysis. Another approach may be the use of Unmanned Aerial Vehicle (UAV) or satellite based technologies coupled with image processing techniques, such as in (Albetis et al. 2017). However, an automatic detection of symptomatic plants based on reddening or yellowing spots may lead to a large amount of unreliable detections (false positives), due to the complex diversity of grapevine diseases. Thus, in the following, we do not focus on this detection step and decide to use the expert map of symptomatic plants to study the spatial pattern analysis of FD repartition in vineyards.

Usually, the goal of spatial and temporal analysis is to exhibit the presence of spatial structures in the occurrences and propagation of symptomatic plants in a 2D space, which means the presence of non-random patterns. These popular methods in ecology fields (Velazquez et al. 2016) can also be applied to epidemics studies (Gatrell et al. 2009), due to the availability of field labeled data (position of symptomatic plants) spatialized on a lattice grid or in a GIS system. Examples include the study of disease patterns in tomato (Kawaguchi & Suenaga-Kanetani 2014) or mint (Johnson et al. 2006) fields. More specifically, pattern analysis applied on FD datasets was conducted in previous studies.

A first description of FD incidence was reported in (Pearson et al. 1985). More recently, (Morone et al. 2007) gathered 5 years of FD incidence in 7 severely affected vineyards in Piemonte, Italia, describing the temporal infectivity and recovery rates of the disease. (Beanland et al. 2006) used spatial analysis over an 8-year period and surveyed the abundance of leafhoppers using traps. Inspired by the works of (Gray et al. 1986) and (Nelson 1995), their proposed methodology consists in a statistical test based on the distance between two symptomatic plants, using random simulations in order to empirically estimate the distribution of the statistic under the null hypothesis. Significant clustering and nonrandom repartitions were reported in that study. Similarly, (Pavan et al. 2012) observed on 4 vineyards monitored over a 5-year period a significant border effect in the occurrences of the symptoms on some of the plots, using the nearest distance to an edge of the vineyard. On one vineyard and 5 years of data, (Maggi et al. 2017) used nearest-neighbor but also quadrat based statistics to show clustering behaviors on a multiscale level. The latter shows some similarity with Ripley's K function describing the spatial occurrences of a phenomenon in a given neighborhood (Besag 1977). A measure of the year-to-year isotropicity was also achieved but symptomatic plants seemed to progress in a non-isotropic way.

The first contribution of this paper is to adapt the nearest neighbor method, in a similar fashion as Ripley's K function, by considering the mean distance (rather than the number of occurrences) between an infected plant and its neighbors at a specific scale. The second contribution is the proposal of an empirical nonparametric method to detect non-random patterns. This latter is based on a hypothesis test whose statistic is computed as the Kullback-Leibler divergence between distance histograms. In the following, this statistic will be compared to two other ones based on distance histograms. The paper is structured as follows. Section 2 presents the methodology and the different statistics used. Section 3 presents the evaluation of the methods on generated non-random datasets. Section 4 and 5 present the results on real FD data acquired during summer 2016. Finally some conclusions and future works are drawn in Section 6.

2. Methodology

2.1. Principle of the randomness test

Similarly to the approach of (Beanland et al. 2006), randomness of the symptoms repartition is constructed as a simple statistical test. The test hypothesis can then be written as:

- H0: the spatial occurrences of the plants can be considered random.
- H1: the spatial occurrences aren't random.

To test this hypothesis, nonparametric statistics are often used in order to highlight significant clustering or dispersion patterns in the vineyard at different scales. In this paper, we follow a similar approach. In the case of FD, clustering is the most probable expected phenomenon. H_0 hypothesis means symptom presence at a given spot is uniformly probable. Thus in order to simulate random repartitions, we model each plant location in the vineyard as a random variable X with a Bernoulli distribution. Given the proportion f of observed symptomatic plants in the real data, the distribution of this random variable is given as:

$$\Pr(X = 1) = f \ , \Pr(X = 0) = 1 - f \ \ (equation\ 1)$$

where 1 corresponds to the values of symptomatic locations and 0 to the values of non-symptomatic ones. It can be noted that, while the mean proportion of observed symptomatic plants per simulation tends towards f for a large number of repetitions, the individual proportion varies for each experiment (binomial distribution). In order to take into account the massive presence of missing plants in some of the plots, random simulations were conducted only on non-missing plants when considering the FD repartition.

Once the test statistic distribution has been estimated under the null hypothesis, the p-value of the test can be computed. If the p-value is lower than the significance level α (e.g. 5%), the H_0 hypothesis is rejected, meaning the spatial repartition is likely to be not random. However, there is a risk of an incorrect rejection of the null hypothesis (type I error or false positive) or an incorrect validation of the null hypothesis (type II error or false negative), meaning every decision comes with a risk.

The proposed test statistics must account for the natural variability of random configurations in a given scale and be sensitive to non-random patterns in the vineyards. Spatial structure is related to the distance between diseased plants at different scales. This means computing the whole distance matrix (full neighborhood) or only considering distances within a fixed radius (local neighborhood). This distance matrix allows the construction of a distance histogram which can then be summarized using two different statistics.

2.2. Statistic based on the mean distance to nearest-neighbor

The nearest-neighbor method (as used in (Maggi et al. 2017) and (Beanland et al. 2006)) reduces the search radius to the nearest observed symptomatic plants. This is an indicator of small scale spatial relationships between plants. Aggregated patterns (or positive spatial autocorrelation) will be characterized by low nearest-neighbor distance while negative spatial autocorrelation means high nearest-neighbor distance. Note that there is no radius for the nearest-neighbor method. To summarize, this approach extracts the shortest pairwise distances between symptomatic plants and uses the mean distance as a statistic, whose simple form can be written as:

$$S = \frac{\sum shortest\ pairwise\ distances}{n} \ (equation\ 2)$$

where n is the number of observed symptomatic plants.

2.3. Statistic based on the mean distance to the neighbors in a radius

In order to obtain a better description, the mean distance to neighbors method uses a fixed radius to threshold the distance histogram and thus is a generalization of the nearest-neighbor statistic by considering a scale parameter. It can also be viewed as a variant of the Ripley's K function as used in (Maggi et al. 2017). Trends at larger scales (larger radius) can be encapsulated using the mean of that thresholded histogram. Similarly, small mean distance in a neighborhood may indicate a clustered structure in the dataset. Depending on the scale (radius), the significance of the test may vary, indicating aggregation patterns at specific scales. In that case, the statistic can be written in a simple form as:

$$S = \frac{\sum(distances < radius)}{m} \ (equation\ 3)$$

where *m* is the number of pairs separated by a distance lower than the radius threshold.

If we do not consider any threshold (i.e. infinite radius), m is equal to $\frac{n(n-1)}{2}$ since all the unique pairwise distances are used for the computation. For all the other cases, *m* depends on the threshold and on the actual symptom repartition. It is worth noting that Ripley's K at a given radius is proportional to *m* (Besag 1977).

2.4. Proposed histogram divergence statistic

However, when considering a mean distance value, information about the underlying distribution is lost. For that reason, we propose a statistic which takes into account the whole distance histogram. A measure of how far a given distance histogram is from an empirically estimated (again, using Monte-Carlo runs) distance histogram is thus needed. This is achieved using the Kullback-Leibler (KL) divergence (Kullback 1951) variant called Jensen-Shannon divergence (JSD) (Endres & Schindelin 2003). Given two continuous probability distribution *p* and *q,* the KL dissimilarity is computed as:

$$D_{KL}(q\|p) = \int q(x)\log\frac{q(x)}{p(x)}dx \quad (equation\ 4)$$

By working with discrete probability distributions *P* and *Q*, the KL divergence can then naturally be defined as:

$$D_{KL}(Q\|P) = \sum_i Q(i)\log\frac{Q(i)}{P(i)} \quad (equation\ 5)$$

Switching from continuous probability law to discrete histogram induces problematic cases where histogram count may be null, meaning potential division by zero. To overcome this issue, we remove zero values in the KL divergence computing. Using average distribution $M = \frac{P+Q}{2}$ allows us to compute the Jensen-Shannon divergence as:

$$JSD(Q,P) = \frac{D_{KL}(Q\|M) + D_{KL}(P\|M)}{2} \quad (equation\ 4)$$

The JSD is similar to the standard KL but bears the advantage of being symmetric. Higher divergence values indicate high departure from the random histogram, meaning a one-tailed test can be set up using divergence histogram as the distribution under the null hypothesis. Similarly to the Ripley's function, results allow to check the CSR hypothesis at different sampling scales as illustrated in Figure 1, in which the radius corresponds to the maximal distances for which two infected plants are considered neighbors and thus are used to compute the statistic. Here, grey zone corresponds to the acceptation zone of the H0 hypothesis (CSR).

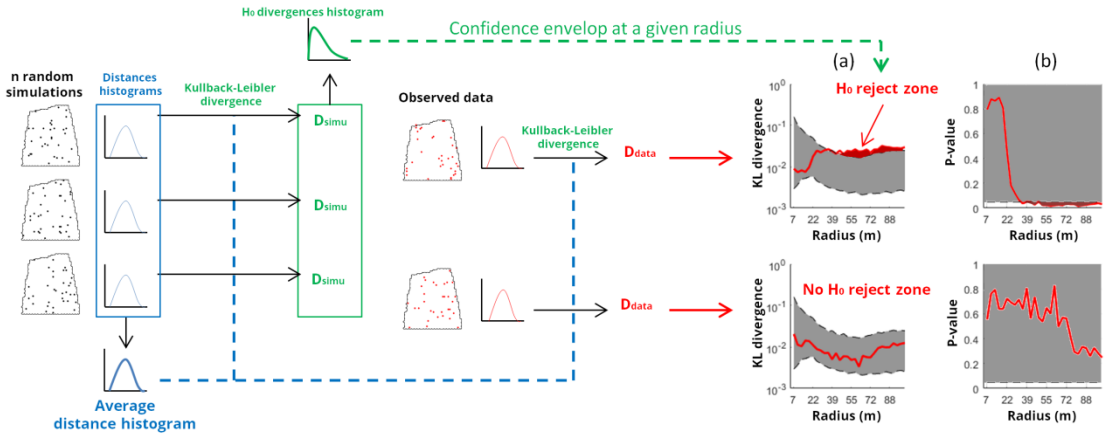

Figure 1 - Schematic explanation of the repartition analysis at several distances using the JSD between histograms as test statistic. (a) - Observed statistic (red curve) at a given radius can be compared to its distribution (grey confidence envelop, 5% and 95% quantiles) under the null hypothesis in order to confirm or reject it. (b) – Evolution of the related p-value, the dashed line represents the significance level (here α = 5%).

In this paper, we use the euclidean distance as a measure of the distance between two plants. In the following sections, the above mentioned methods will be compared to simulated non-random datasets, then the one with the best performance will be applied on real datasets. All the experiments are conducted using the programming language MATLAB (R2015b, The MathWorks Inc.).

3. Results

3.1. Results on generated datasets

In this section, we wish to evaluate how well each method performs on datasets with different properties. An accurate method must be able to discriminate random repartitions from nonrandom ones with minimal risk. The proposed methods are first applied on simple datasets with for example distinctive aggregation (maximal positive autocorrelation on set 1 of Figure 2.a) and dispersion patterns (maximum negative autocorrelation on set 2). The aim of this experiment is to evaluate the potential of the proposed models to identify 10 non-random repartitions. Figure 2 presents the evolution of the three test statistics on 10 user-created sets (Figure 2.a) for different radiuses. Using the mean-distance (Figure 2.c) and the divergence statistic (Figure 2.d), all the datasets lead to the reject of the H0 hypothesis for most radiuses using an alpha risk of 1%. The only exception here is set number 2 with its uniform repartition. In that case, while the distribution can't be considered random, the mean-distance appears nearly identical as the ones computed for random datasets. Departure from random sets can be clearly seen using the nearest-neighbor statistic (Figure 2.b, very high mean distance to nearest neighbor) but also with the proposed histogram divergence statistic.

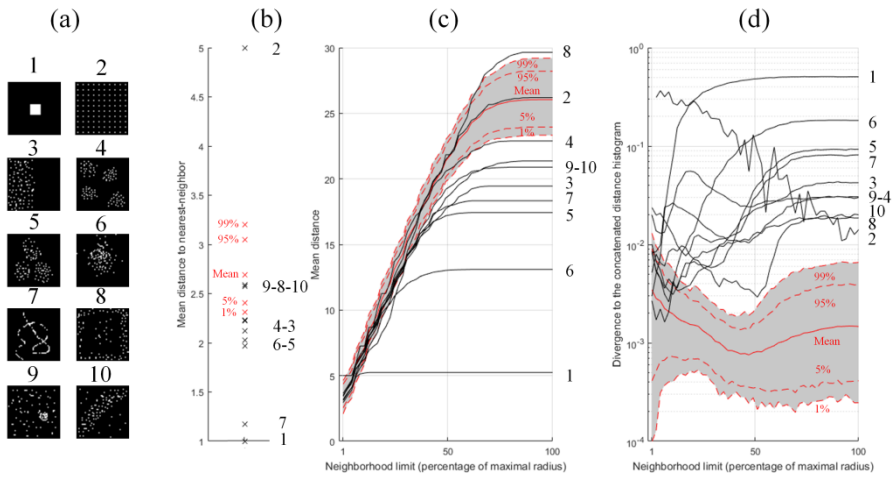

Figure 2 – (a) 10 simulated dataset with non CSR property (*f*=0.04, 50x50 grid) (b) - Mean nearest-neighbor distance (unitless), (c) – Mean neighbors distance (unitless) and (d) - mean divergence (unitless) observed on the 10 datasets. Red lines correspond to the empirical quantities observed for random repartitions and different quantiles (potential H0 reject values for a corresponding α risk), black lines correspond to the observed distances on the 10 datasets. For clarity, divergence axis is presented on a log-scale.

Methods are then applied to generated datasets with different properties. The symptom frequency *f* is considered here as the first critical parameter to take into account, the other one being the 'randomness level' (called *r*) of the repartition. Randomness is however hard to quantify. To obtain a quasi-continuous randomness level, we alter the uniformly equal random probability field under the H0 hypothesis to generate intermediate states between total randomness and a fully determined state. This is done by adding another random field using 2D Gaussian negative and positive spots of random sizes and orientations. Examples of generated repartitions using this method are shown in Figure 3. Symptom frequency *f* was chosen to fit the naturally observed disease and missing plants frequencies (e.g. from 1% to 32% as illustrated in Figure 5). The *r* values were chosen using linearly sampled values ranging from 0 (deterministic field) to 1 (unaltered uniform random field).

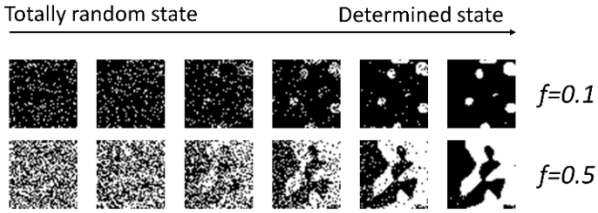

Figure 3 – Examples of generated symptom repartitions for 2 frequencies (f) and 6 randomness levels (r)

This approach can be used for different symptom frequencies and repeated (here, we use 500 repetitions per (*f,r*) couple to generate a wide array of repartitions) in order to obtain a set of simulated nonrandom repartitions. We can now compare totally random sets with non-random sets in order to assess the efficiency of the previously described methods.

In order to evaluate the discriminative ability of each method, we use ROC curves (Fawcett 2006), a popular machine-learning visualization tool, applied on the p-value of the statistical test. In that case, ROC curves serve as a visual illustration of the first and second order risks. For a given false alarm rate (proportion of false decisions induced by the test reject), ROC curves allow to determine the expected true alarm rate (proportion of true positives). The area under the curve (AUC) is also an indicator (ranging from 0 to 1) of how well random repartitions are separated from non-random ones. Figure 4 gives an illustration on the process behind the creation of ROC curves. Here values in red indicate the p-values obtained from random datasets while p-values in blue correspond to the ones obtained on non-random datasets. The best case means that it is possible to set a threshold to perfectly separate random repartitions from non-random ones based on the computed p-values. In that case, the AUC is equal to 1.

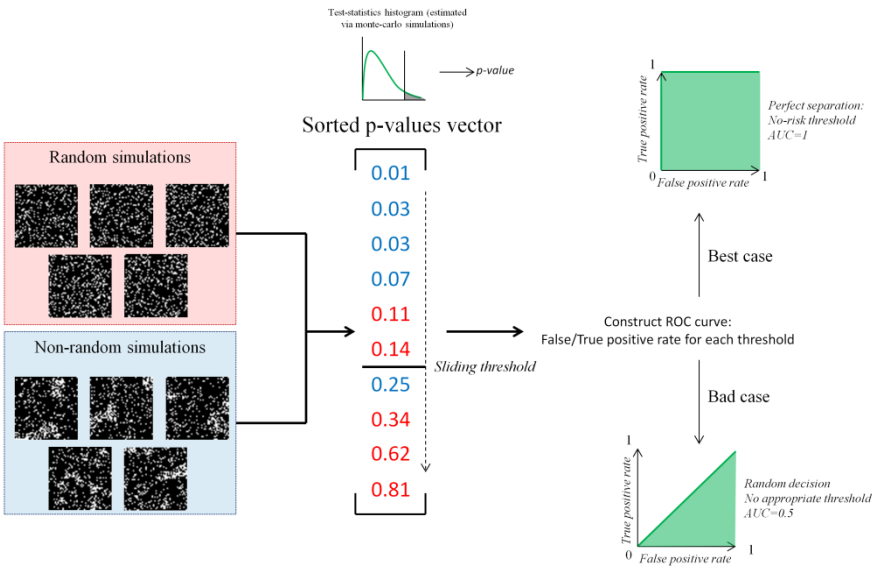

Figure 4 - ROC curves generation methodology. AUC indicates how well non-random disease repartitions can be separated from random repartitions when fixing a critical p-value threshold to the test

Results on the generated datasets (*f* ranging from 0.01 to 0.32, *r*=10 randomness levels) can be consulted in Figure 5. Total randomness column compares random repartitions with random repartitions, thus all the ROC curves follow the *y*=*x* line. For almost every set of parameters, the divergence statistic yields to the best AUC, meaning better ability to set apart random repartitions from non-random ones and more sets with close-to-zero risk (AUC=1) when rejecting the H_0 hypothesis. It also shows less sensitivity to the *f* parameter than the other two methods. As expected, the nearest-neighbor method tends to perform poorly on repartitions with high disease occurrences. In such case, the mean distance to the nearest diseased neighbor may be very low, and may lead in some cases to random decision (AUC=0.5). The nearest-neighbor method seems to be nearly as effective as the other methods if the local density of infected plants is not too high.

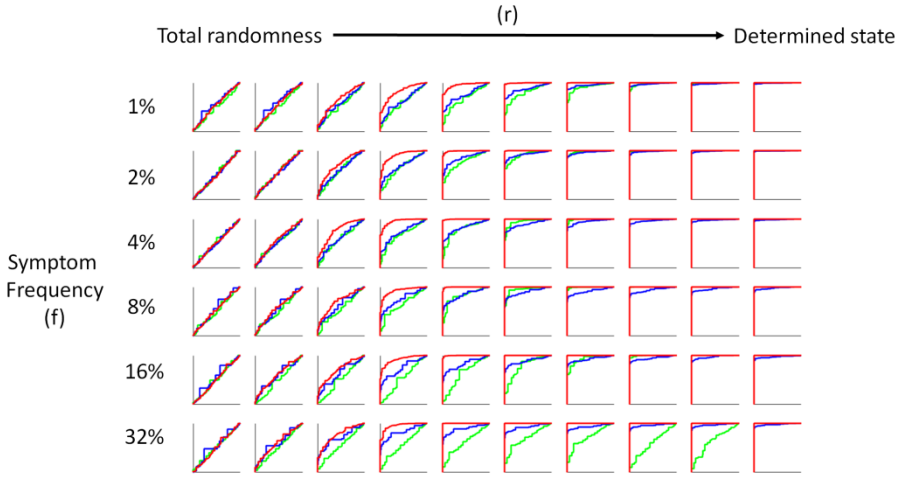

Figure 5 - ROC curves computed for different (frequency, randomness) couples using nearest-neighbor distance statistics (green curves), local mean-distance statistics (blue curves) and local divergence statistics (red curves). Maximal radius was considered for this experiment

Since the proposed statistic based on the histogram divergence has shown the best performance on these datasets, this method will be used in the following on real FD dataset acquired on La Brède test site.

4. Results on real flavescence dorée data

4.1. Studied vineyards and field campaign

Experiments were conducted in 7 vineyard plots near La Brède in the Bordeaux region (Figure 6). In this region, FD occurrence was diminishing from 2012 to 2015 but it seemed on the rise since then. Average plant density in the vineyards is about 2000ha^{-1}. The vineyards exhibit various sizes, orientations and cultivar combinations combined with FD outbreaks.

Figure 6 - Geographic location of the 7 study (L1 to L7) vineyards in the La Brède municipality (WGS 84 coordinates: -0° 31' 58.58",44° 41' 29.87"), Nouvelle Aquitaine region, France. Satellite image: Google Earth.

The monitoring of FD symptoms was conducted during summer 2016, on September 8. The presence of FD was evaluated using a protocolar grid indicating FD severity on a 4 levels scale (from faint foliar symptoms to full foliar symptoms) and possible confusion risks with other diseases. Photographs of the symptomatic plants were also taken in order to check later on for these confusions. Figure 7 presents the result of the field campaign in a grid representation of the symptomatic and missing plants in the 7 vineyards. Using aerial imagery, gridded symptomatic plants positions were also georeferenced (relatively to the position of rows extrema observed on the aerial image) and included in a GIS software (QGIS 2.16), allowing a more precise distance measurements between symptomatic plants.

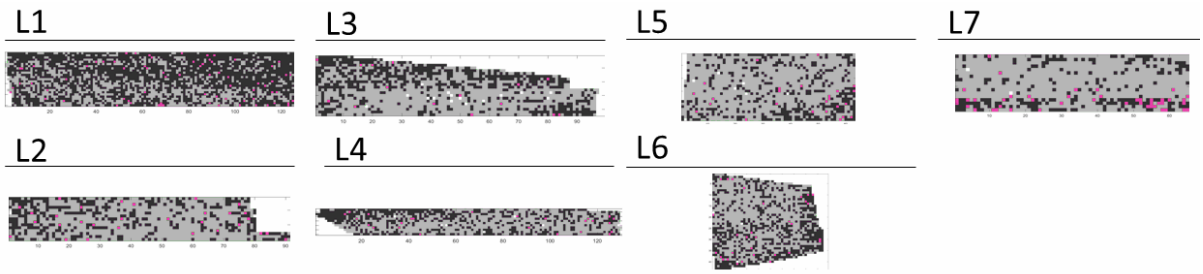

Figure 7 - Grid view of the field notation campaign data. Gray pixels: Healthy Plants. Black pixels: Missing plants, Pink pixels: FD symptomatic plants.

Most missing plants in the vineyards likely come from former uprooted FD symptomatic plants. Vineyard L1 particularly bears the trace of previous epidemics, with a missing plant proportion superior to 50%. On the other vineyards, it ranges from 20% to 40%. FD symptom occurrences vary from one vineyard to another, ranging from 1% to 4% of the total sampled plants. Observed proportions are summarized in Table 1, as well as a list of the cultivars present in the vineyard.

Table 1 - Description of the 7 studied vineyards

Vineyard	L1	L2	L3	L4	L5	L6	L7
Cultivars	*Cabernet-Sauvignon*	*Merlot, Sémillon*	*Merlot, Sauvignon Blanc*	*Cabernet-Sauvignon, Merlot*	*Merlot, Sémillon*	*Cabernet-Sauvignon*	*Cabernet-Sauvignon*
Number of plants	*3084*	*1222*	*1688*	*1448*	*1103*	*1652*	*1603*
Number of rows	*25*	*15*	*22*	*12*	*17*	*43*	*26*
Frequencies (f)							
Healthy plants	*25.3%*	*65.7%*	*57.7%*	*52.7%*	*73.7%*	*57.8%*	*61.9%*
Missing plants	*62.5%*	*30.8%*	*37.5%*	*39.6%*	*19.9%*	*38.3%*	*25.4%*
Flavescence dorée	*2.4%*	*2.8%*	*1%*	*2%*	*4.1%*	*2%*	*1.6%*
Other diseases	*9.8%*	*0.7%*	*3.8%*	*5.7%*	*2.3%*	*1.9%*	*11.1%*

Figure 8 summarizes the main results obtained on this dataset. Several tests are performed on the 7 studied vineyards using the best-performing methods previously described. The first test involves simple random hypothesis. It aims to determine if according to the spatial occurrences, the repartition can be considered as random. It is applied to FD plants (Figure 8.a), but also on missing plants (Figure 8.b). The second test aims at determining if a border effect is present in the vineyard (Figure 8.d), i.e. FD/missing plants are closer to the edges than in random repartitions. This is achieved using the nearest-neighbor statistic method, which is preferred here since only the distance to the nearest border is of interest. Eventually, we wish to know if a spatial relationship between FD symptoms locations and missing-plants location exists. For that purpose, we construct a test based on the divergence method, using the distance from FD affecting plants to missing plants, simulating both random repartitions (using respective *f* values) for each Monte-Carlo run (Figure 8.c).

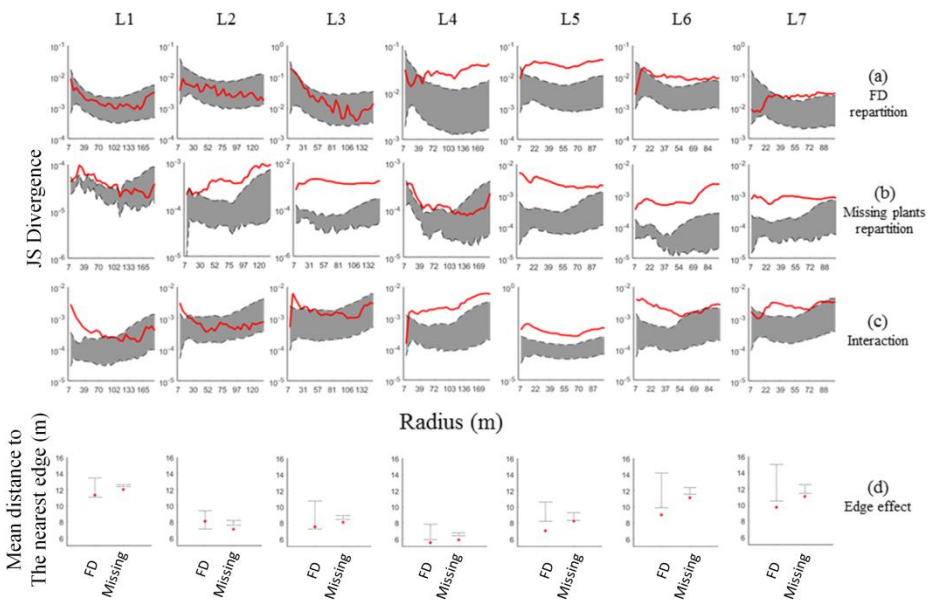

Figure 8 – Test statistics (red curves and points) obtained on the La Brède dataset. Values falling outside the grey confidence envelope (5% confidence) indicate H_0 reject and non-random behaviors. (a) to (c) test based on JSD (for varying threshold distances) while (d) considers the nearest-neighbor distance and thus is conducted without any distance threshold.

When setting the significance risk α to 5%, all the vineyards reject the null hypothesis for the missing plants repartition for small maximum distance (radius) indicating a nonrandom repartition, the reject was however not significant at large scale for plots 1 and 4. FD repartition was found to be nonrandom on plots 4 to 7, while plots 1 to 3 never reject the H_0 hypothesis. Similar interpretations

can be made for the edge effect test, featuring reject on 4 plots for FD repartition and all the 7 plots for the missing plants repartition, which supports the results of (Pavan et al. 2012) and (Maggi et al. 2017) on the observed edge effect in vineyards affected by FD. Tests considering the proximity of diseased plants with missing plants (Figure 8.c) yield reject on most vineyards at small scale except for plots 2 and 3, leading to the confirmation that missing plants and FD diseased plants are spatially linked, showing significant attraction patterns. The proximity with missing plants seems especially significant on vineyard 4 and 5 in which nonrandom patterns can be observed at small and large scales, which corroborates the visual observation of FD affected plants located nearby clusters of rogued out plants.

5. Discussion

While statistical non-parametric methods using Monte-Carlo simulations provide a flexible framework for the repartition analysis, several potential pitfalls remain. Important missing plants proportions are still an issue because they limit the potential number of possible locations for simulated symptoms, meaning fewer combinations and blurrier frontier between random and non-random patterns. This is the case for vineyards with more than half of missing plants such as vineyard number 1. Similarly, the joint presence of several cultivars in the vineyards calls for prudent interpretations. This doesn't apply to pure *Cabernet Sauvignon* plots such as 1, 6 and 7 but the other exhibit mixed cultivars, whose susceptibility to the disease may be different. The presence of a different cultivar on the border of the plot, such as in vineyard 5, may lead to the hypothesis reject because symptomatic plants are more frequent near this border and thus the repartition may not appear random. This study provides a repartition analysis tool which can be embedded into an epidemics surveillance system. Prior steps to the analysis may include ground field notation or more innovative tools such as satellite or UAV imagery, both used to obtain spatialized information about the disease repartition. Symptomatic plants positions can then be analyzed in order to check for potential threatening aggregations patterns, even for low FD prevalence.

6. Perspectives and conclusion

In this study, we have proposed a simple nonparametric statistical framework using a divergence histogram statistic taking into account the distance histogram in a given radius. Results on generated datasets indicate better separation between random repartitions and non-random ones using the divergence method. Better generation of simulated repartitions may include the control over specific aggregation patterns such as outbreak size or edge-effect. Applications on real FD data indicate significant non-random aggregation patterns on some of the studied plots, as well as an edge effect and a possible spatial correlation between missing plants and FD symptoms. This method could be applied to the temporal study of FD propagation, using yearly datasets. Spatial pattern studies could also be coupled with other exogenous variables such as spatial leafhopper abundance in the vineyard as in (Pavan et al. 2012).

7. Acknowledgement

This study has been carried out with financial support from the Région Nouvelle Aquitaine and the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02). First author is financed by Fonds Unique Interministériel (FUI) ADVANTAGE driven by Vegepolys/ Agri Sud-Ouest Innovation and with support from Région Nouvelle Aquitaine/Bpifrance.

8. References

Albetis, J. et al., 2017. Detection of Flavescence dorée Grapevine Disease Using Unmanned Aerial Vehicle (UAV) Multispectral Imagery. *Remote Sensing*, 9(4), p.308.

Beanland, L., Noble, R. & Wolf, T.K., 2006. Spatial and Temporal Distribution of North American Grapevine Yellows Disease and of Potential Vectors of the Causal Phytoplasmas in Virginia.

Environmental Entomology, 35(2), pp.332–344.

Besag, J., 1977. Contribution to the discussion of Dr Ripley's paper. *Journal of the Royal Statistical Society*, 39(2), pp.193–195.

Bonfils, J. & Schvester, D., 1960. Les cicadelles (Homoptera Auchenorrhyncha) dans leurs rapports avec la vigne dans le sud-ouest de la France. *Ann Epiphyt*, 3, pp.325–336.

Caudwell, A., 1957. *Deux années d'études sur la flavescence dorée nouvelle maladie grave de la vigne*, Paris: Institut National de la Recherche agronomique.

Caudwell, A., 1990. Epidemiology and characterization of flavescence dorée (FD) and other grapevine yellows. *Agronomie*, 10(8), pp.655–663.

Endres, D.M. & Schindelin, J.E., 2003. A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7), pp.1858–1860.

Fawcett, T., 2006. An introduction to ROC analysis. , 27, pp.861–874.

Gatrell, A.C. et al., 2009. Spatial point analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographer*, 21(1), pp.256–274.

Gray, S.M., Moyer, J.W. & Bloomfield, P., 1986. Two-dimensional distance class model for quantitative description of virus-infected plant distribution lattices. *Phytopathology*, 76(2), pp.243–248.

Johnson, D., Zhang, H. & Alldredge, J.R., 2006. Spatial pattern of verticillium wilt in commercial mint fields. *Plant disease*, 789-797(June).

Kawaguchi, A. & Suenaga-Kanetani, H., 2014. Spatiotemporal distribution of tomato plants naturally infected with leaf mold in commercial greenhouses. *Journal of General Plant Pathology*, 80(5), pp.430–434.

Kullback, S., 1951. on Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), pp.79–86.

Maggi, F. et al., 2017. Space-Time Point Pattern Analysis of Flavescence Dorée Epidemic in a Grapevine Field: Disease Progression and Recovery. *Frontiers in Plant Science*, 7(January).

Morone, C. et al., 2007. Epidemiology of flaveseence doree in vineyards in northwestern Italy. *Phytopathology*, 97, pp.1422–1427.

Nelson, S.C., 1995. Spatiotemporal distance class analysis of plant disease epidemics. *Phytopathology*, 85(1), pp.37–43.

Pavan, F. et al., 2012. Border effect in spatial distribution of flavescence dorée affected grapevines and outside source of Scaphoideus titanus vectors. *Bulletin of Insectology*, 65(2), pp.281–290.

Pearson, R.C. et al., 1985. Occurrence of flavescence dorée-like symptoms on "White Riesling" grapevines in New York, U.S.A. *Phytopathologia Mediterranea*, 24(1/2), pp.82–87.

Rahola, J. et al., 1997. La flavescencia dorada en los viñedos del Alt Empordà (Girona). *Bol Sanid Veg Plagas*, 23(403), p.416.

Schaerer, S. et al., 2007. Flavescence dorée : la maladie et son extension. *Revue suisse Vitic. Arboric. Hortic.*, 39, pp.107–110.

Velazquez, E. et al., 2016. An evaluation of the state of spatial point pattern analysis in ecology. *Ecography*, 39(11), pp.1042–1055.

Vidano, C., 1964. Scoperta in Italia dello Scaphoideus littoralis Ball cicalina americana collegata alla Flavescence dorée della vite. *Ital Agric*, 101, pp.1031–1049.