Towards estimating the proportion of dead and missing vines at the field level

Baptiste Oger, Cécile Laurent, Philippe Vismara, Bruno Tisseyre

iVES OENO One vine & wine

DOI: https://doi.org/10.20870/oeno-one.2025.59.1.8061

OPEN ACCESS

ORIGINAL RESEARCH ARTICLES

Towards estimating the proportion of dead and missing vines at the field level

Baptiste Oger[1,*], Cécile Laurent[2], Philippe Vismara[3], Bruno Tisseyre[1]

[1] ITAP, Univ. Montpellier, L'institut Agro Montpellier, INRAE, France.
[2] Fruition Sciences, MIBI, 672 Rue du Mas de Verchant, 34000 Montpellier, France.
[3] MISTEA, Univ. Montpellier, L'institut Agro Montpellier, INRAE, France.

Article number: 8061

*correspondence:
baptiste.oger@hotmail.com

Associate editor:
Franco Meggio

Received:
19 March 2024

Accepted:
19 November 2024

Published:
10 January 2025

ABSTRACT

This work aimed to improve practices for estimating dead and missing vines, particularly when these estimations are made via sampling. Vine mortality is a well-known phenomenon that many studies have attempted to better explain, but only a few have focused on estimating the number of dead and missing vines at the field-level, which remains a challenging task for grapegrowers. The present article aimed to determine to what extend an increased an increased sampling effort improves estimation accuracy and thus help practitioners define a sample size adapted to vineyard properties and the accuracy expected by the grower. The first part of the study investigated whether vineyard properties (year of planting and variety) and available ancillary data (soil resistivity and fraction of vegetation cover [Fcover]) can provide *a priori* information on the proportion of dead and missing vines and guide the sampling strategy. The analysis was based on an exhaustive dataset created by individually mapping 14,199 dead and missing vines across 29 fields in a 20 ha vineyard. In this vineyard, regression models showed an increase in dead and missing vines of around 2.2 % per year from the tenth year onwards. Plantation year, variety and vigour (assessed by Fcover) were identified as providing valuable prior information on the proportion of dead and missing vines within the fields. The second part of the study focused on characterising the estimation error values that can result from these estimations. It focuses on sampling estimation, highlighting to what extent field properties (such as the number of dead and missing vines and their spatial autocorrelation) affect the accuracy of estimates. The results showed that the same sampling protocol can result in different accuracies from one field to another. Thus, they underline the importance of leveraging prior information, such as available field data, to tailor sampling efforts accordingly. Based on these results, this article introduces new guidelines to facilitate the estimation of the proportion of dead and missing vines in vineyards.

KEYWORDS: sampling, vineyard, mortality, vine decline, grape variety, viticulture, vegetation index

INTRODUCTION

In viticulture, the presence of dead and missing vines (DMV), is a pervasive issue that can have profound implications for vineyard productivity and sustainability. The causes of vine mortality are multifaceted and can include a wide range of biotic (Chuche & Thiéry, 2014; Borgo *et al*., 2016; Lade *et al*., 2022) and abiotic (Suarez *et al*., 2019; Gambetta *et al*., 2020) factors, such as pests, diseases, environmental stresses and management practices. Over the past twenty years, vine decline, encompassing vine mortality and yield loss, has been reported as a growing phenomenon in most wine-producing regions (Fontaine *et al*., 2016). At the individual vineyard field level, knowledge about the number of DMV can serve different purposes, which are either direct (monitoring DMV to decide on vine replanting or vineyard renewal) or indirect (for yield estimation, disease spread monitoring etc.). Furthermore, estimating the proportion of DMV can serve as an indicator of underlying issues within vineyards that necessitate prompt attention and intervention to mitigate further losses and ensure the long-term sustainability of grape production. Despite its importance, the presence of DMV remains a relatively understudied aspect in viticulture, and only a few studies have focused on the actual description of DMV as a practical indicator for decision-making. Recent studies, for example, have highlighted the economic consequences of the presence of DMVs (van Zoeren *et al*., 2020), proposed action plans to combat vine decline (Riou *et al*., 2016) and highlighted the perception grapegrowers have of DMV and their effects on yield (Merot *et al*., 2023).

In the last few years, with the development of data collection technologies, many studies have highlighted the potential of automatic detection methods for DMV (Chanussot *et al.,* 2005; Tang *et al.,* 2016; Primicerio *et al.,* 2017; Di Gennaro *et al.,* 2020; Fernandes *et al.,* 2021). However, these technologies are considered costly and applicable only in very specific conditions. Therefore, they are not widely deployed and are rarely used by grapegrowers. As a result, the proportion of DMV remains poorly monitored in most vineyards, with all the technical and economic implications that this entails. When it is monitored, it is often estimated using simpler approaches, such as sampling. However, the uncertainty associated with sampling protocols represents the main obstacle to its use, as for the more general characterisation of the proportion of DMV.

This sampling uncertainty arises from two main phenomena. First, the proportion of DMV can vary greatly from one field to another, depending on internal factors, such as age and variety (Fussler *et al.,* 2008), as well as on external factors, such as climatic conditions (Gambetta *et al.,* 2020) and vine diseases (Lade *et al.,* 2022). Within the same commercial vineyard, it is possible to observe fields with almost no DMV, while other fields contain up to 40% of DMV (van Zoeren *et al.,* 2020). However, from a statistical point of view, the uncertainty associated with the sampling estimation of a qualitative variable such as the presence of DMV is related to the frequency of the observed phenomenon (Brown *et al.,* 2001). As a result, the same sampling protocol can lead to totally different accuracies depending on whether the field has a low or high % of DMV. The relationship between the number of observations and the uncertainty is also elusive. While it is commonly known that increasing the relative sampled area reduces uncertainty, available tools in viticulture do not provide sufficient information to quantify this phenomenon. Therefore, it is difficult for grapegrowers to assess the reliability of an estimate, and it is always challenging to define a sampling protocol that ensures the desired accuracy. Surprisingly, no recent work has addressed DMV sampling in viticulture, even though this is important information that has multiple applications to support decision-making in vineyards. It is, therefore, crucial to provide information regarding the estimation of DMV proportion at the vineyard field level which corresponds to the typical management unit in viticulture.

Through the lens of geostatistics, this article aims at providing new knowledge on how best to estimate the proportion of DMV. To achieve this, the article utilises a high spatial resolution (individual vine scale) dataset of DMV across a large number of fields within a vineyard. The first part investigated whether any auxiliary information could provide *a priori* information on the proportion of DMV to guide the sampling strategy. To this end, the relationship between the proportion of DMV and available ancillary variables was studied through regression analyses. In the second part, this study examines the relationship between the size of the relative sampled area, vineyard field properties and sampling estimation errors. Finally, guidelines were derived from the results of both parts to support grapegrowers when defining sample sizes and protocols that are adapted to their own field and at an accuracy that will allow the grapegrower to make decisions.

MATERIALS AND METHODS

1. Data

1.1. Data acquisition

Dead and missing vines (DMV) were counted and georeferenced in May 2022 across 20 ha and 29 fields of a vineyard located near Montpellier in the South of France (latitude: 43.532300; longitude: 3.864230). Figure 1 shows the location of the study and the spatial organisation of the fields. Each vine in the 29 fields was checked and the position of all DMV was recorded by a single operator. To ensure precise georeferencing, the operator utilised the Mergin Map application (developed by Lutra Consulting, UK) on a smartphone connected to a GNSS RTK (Global Navigation Satellite System Real-Time Kinematic) receiver. The GNSS receiver was linked to the Centipede RTK network (Ancelin *et al*., 2022) to achieve centimeter-level accuracy. The final dataset includes the locations of 14,199 DMV. Healthy vines were not mapped. Figure 1C shows a part of this dataset for some of the surveyed fields.

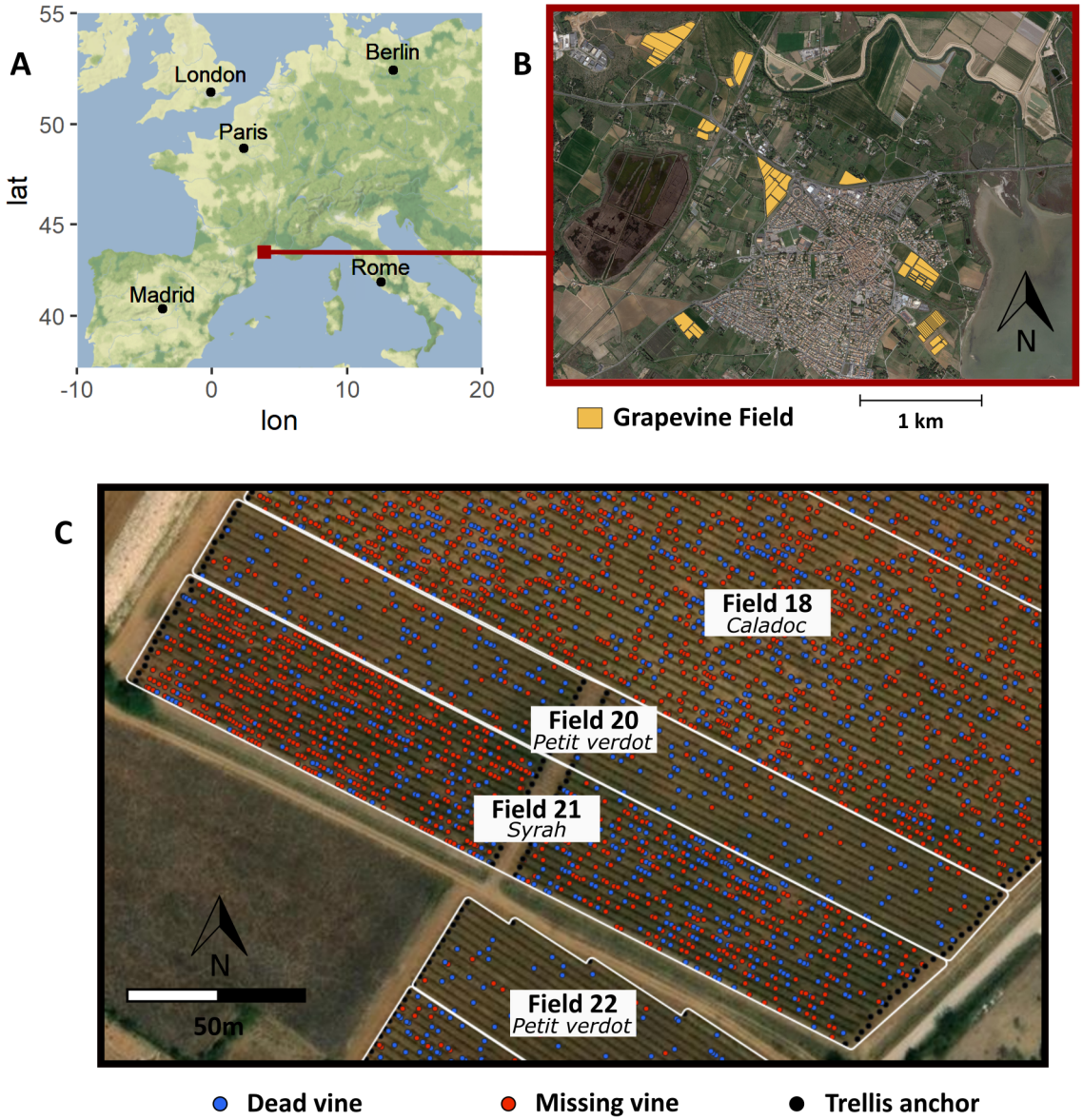

FIGURE 1. Location (A), representation (B) of the grapevine fields, and snapshot of the final dataset (C).

The vineyard produces winegrapes (*Vitis vinifera*) and grown varieties include Grenache, Mourvedre, Merlot, Syrah, Marselan, Viognier, Chasan, Aranel, Caladoc, Petit Verdot and Terret blanc. The field areas vary, being as small as 0.02 ha and as large as 1.89 ha, with plantation years spanning 1998 to 2017. Row spacing is consistent at 2.5 m and vine spacing at 1 m or 1.2 m. Specific information for each field is available in the supplementary data. All vines were trained in the short-spur cordon de Royat system, supported by a three-wire trellis system. Vines are all pruned using a mechanical pre-pruning operation performed as early as possible after leaf fall (November-December). The mechanical pre-pruning was followed by a traditional manual pruning carried out during January, February, and the first 15 days of March. Weeding was primarily mechanical, with the use of an under-row weeding machine in October, March and April, as well as middle row plowing in October, December and April. Only two fields were weeded chemically in March.

For each of the 29 fields, ancillary data providing supplementary information were acquired. To characterise vegetation cover, the commercial service Oenoview (ICV, Lattes, France) was used to provide FCover (Fraction of vegetation Cover) maps at 1.5 m² pixels resolution (Tondriaux *et al.*, 2018) derived from SPOT 6-7 satellite images. Images were acquired in early July 2022. During this period, 24 out of the 29 fields had little to no weeds in the inter-row areas. Measurements of soil electrical resistivity were performed with a commercial service provider (Geocarta, France) in 2018. The measurements were taken with electrode Wiener devices giving a maximal depth of investigation of 1 m (Panissod *et al.*, 1998). Data averaged at the field level can be found in supplementary data.

1.2. Indicators

For each field, the proportion of DMV was computed as the ratio of the number of observed DMV to the number of vines

initially planted in the field. The later value was calculated from the plantation density and the area of the field.

$$DMV_{Field}(\%) = \frac{Number_{dead \wedge missing vines}}{\left(\frac{Area}{Distance_{interow} \times Distance_{interplant}}\right)} \quad (1)$$

A second indicator of spatial autocorrelation was computed from the measurements. This indicator was based on variographic analysis. As variographic analyses are performed on quantitative variables, and not on qualitative variables such as productive vines vs. dead or missing vines, the data were re-rasterised for the sole purpose of computing this indicator. To achieve this, the fields were divided into 5m-by-5m squares using the grid function in the QGIS software (QGIS; QGIS Association, http://www.qgis.org), and the number of DMV was retrieved for each pixel. The 5 x 5 m resolution was chosen as it represents a good compromise, capturing a significant portion of the spatial information without distorting its structure (Tisseyre *et al*., 2018). This method draws inspiration from analyses commonly used in forestry (Bellehumeur *et al*., 1997) when studying planting densities. Variographic analysis was then performed based on the number of observed DMV in each pixel using the "gstat" package (Pebesma, 2004). This indicator represents the percentage of variance explained by a spatial phenomenon (hereafter referred to as "Spatial autocorrelation") and was computed from (*i.e*., the nugget effect that represents the variance share that is not spatialised) and (*i.e*., the partial sill that represents the variance share that is spatialised) as shown in Equation 2:

$$Spatialautocorrelation(\%) = \frac{C1}{C0+C1} \quad (2)$$

1.3. Simulated data

Simulated data were used to better understand the effect of spatial autocorrelation on the DMV estimation error. The distribution of estimation errors is conditioned to the independence of measurements when sampling (Thompson, 2012). In crop production, this independence may be compromised by the spatial organisation of measurements (spatial autocorrelation). This is often driven by spatially structured environmental factors (soil and elevation, etc.); *i.e*., environmental factors that are not randomly distributed over space. This results in geographically close measurements that are more similar to each other than distant ones. In viticulture, soil factors related to vine decline (Desassis *et al*., 2005) and vine mortality (Vaudour *et al*., 2017) have been shown to be spatially structured. Therefore, the experimental dataset was supplemented with simulated data to isolate and identify the effect of a factor by ensuring that all other field parameters remained constant. The simulated data are therefore a good complement to the real data set, which did not allow this kind of comparison to be carried out.

Four 100 m × 100 m (1 ha) fields were simulated. The row spacing was set at 2.5 m and the vine spacing at 1 m, resulting in a plant density of 4,000 vines/ha. The objective was to generate fields with two levels of spatial auto-correlation (10 % and 30 %) and two levels of proportion of DMV (10 % and 30 %). As proposed by Oger *et al*. (2023), a spatially structured quantitative variable was generated for each field using a two-step approach. First, Gaussian fields with no nugget effect were simulated using the "gstat" package (Gräler *et al*., 2016). Their sills were respectively set at 10 % and 30 % of the total variance. These Gaussian fields represented the spatialised part of the simulated fields. Second, to obtain the unstructured variability of the simulated fields, a random nugget effect was then added to each Gaussian field. The nugget effects were added by using a simple centred normal distribution of variance that was 90 % and 70 % of the total variance (σ^2). Finally, the 10 % or 30 % of the vines with the highest spatially structured quantitative values were considered as dead or missing in the simulated fields.

2. Data analysis and explanatory factors for vine mortality

2.1. Linear regression

Linear regressions were performed to relate the observed proportion of DMV in the fields to the age of the vines. Twelve of these models were calibrated: one general model integrating the data from all eleven varieties and one model per variety. All the models were set to cross the same point to i) exclude unrealistic models, which present negative theoretical proportions of DMV or already positive proportions at planting (age = 0), ii) improve the robustness of the approach for varieties with few repetitions, and iii) facilitate the comparison of different models by focusing on the regression coefficient as the sole parameter. Given that mortality is low in the early years and dead vines were most often replaced within the first ten years, the point with coordinates [0,10] was selected. The two youngest fields, planted in 2014 and 2017 and showing very few missing vines, were therefore not included in the models. Additionally, this approach allowed for an easier comparison of varieties using a single parameter (the regression coefficient), instead of two parameters. The models were programmed using the "lm" function from the "stats" package (R Core Team, 2013).

2.2. Varieties classification based on mortality rate

To compare the variety effect with the effect of other variables that could explain mortality and overcome the small sample sizes associated with each grape variety, the grape varieties were divided into 3 classes of mortality: low, intermediate and high. Classification was performed based on regression coefficients obtained with the linear model.

Grape varieties with a regression coefficient that lay within the standard error of the average regression coefficient (with all varieties included) were considered as having "intermediate mortality". Grapes varieties with a regression coefficient that was higher or lower than the interval $[regressioncoefficent(allvarieties) \mp standarderror]$ were considered to have "low mortality" or "high mortality" respectively.

2.3. ANCOVA model

An analysis of covariance (ANCOVA) was conducted to explain the differences in observed mortality among fields based on the available data. This statistical analysis was performed using the "anova_test" function from the "rstatix" package (Kassambara, 2023).

3. Sampling estimation of DMV

3.1. Sampling protocol

To evaluate the sampling estimation of the proportion of DMV in the field, a Row-Based Random Sampling (RBRS) protocol was defined. This protocol randomly selects n rows out of the r rows of a field and the selected rows are then fully observed. The advantage of such sampling procedures is that they account for operational constraints encountered in viticulture and correspond to what is done in practice by grapegrowers (Oger *et al.*, 2021). In some specific cases, fields are intersected by paths that allow a change of rows in the middle of the field. In such situations in this study, the two parts separated by the path were considered as corresponding to two different rows.

3.2. Sample properties

Each sample was described by two attributes: the relative sampled area in the field and the estimation error. The relative sampled area in the field was determined by summing the lengths of the rows sampled and dividing it by the total length of all rows within the field. This metric enables the comparison of sampling efforts, particularly when vineyard fields are non-rectangular and rows vary in length. This is not the case for the simulated fields where all rows have exactly the same length, but the same approach has been applied to all fields, whether they were real or simulated. This metric is expressed as a percentage.

The estimation error associated with a sample is computed by comparing the proportion of DMV within the sample (DMV_{Sample}) to the actual proportion of DMV within the field (DMV_{Field}). A relative error (%) is computed as described in Equation 3:

$$Error(\%) = \frac{DMV_{Sample} - DMV_{Field}}{DMV_{Field}} \quad (3)$$

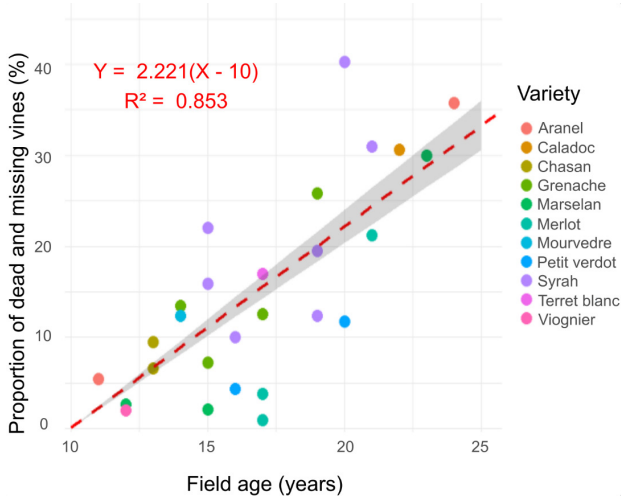

FIGURE 2. Proportion of DMV per field as a function of age. The varieties are represented by different colours. The red dashed line represents a linear model assuming negligible mortality before 10 years, the grey area represents the standard error of the linear model.

RESULTS

1. Explanatory factors for vine mortality

Figure 2 shows a scatter plot of the different vineyard blocks according to plantation year versus proportion of dead and missing vines (DMV). The two fields planted > 10 years ago had a very low proportion of DMV and are not represented.

Out of the 29 available fields, the proportion of DMV varied from 0.2 % to 40.3 %. Figure 2 highlights the direct relationship between vine age and the proportion of DMV in the fields. The regression model, which relates the proportion of DMV with age, has a coefficient of 2.23 ± 0.17, meaning that the fields lost an average of 2.23 % of their vines each year after their tenth year. Beyond the age effect, some varieties, such as Petit Verdot and Merlot, were mostly below the regression model, while others, like Syrah and Aranel, were mostly above it. These observations could indicate the existence of a variety effect, but require further confirmation. This was consistent with the fact that not all varieties have the same mortality rate and may exhibit varying resistance to the different stresses that affect them.

Table 1 shows the classification of the grape varieties according to the three mortality levels low, intermediate and high, as observed in the fields of the survey. Varieties with intermediate mortality corresponded to those whose regression coefficient fell within the standard error of the regression line (Figure 1). Low mortality varieties had a regression coefficient lower than the regression line, while varieties with a high mortality had a regression coefficient higher than the regression line.

Regarding the varieties Syrah, Caladoc, Grenache, Petit Verdot and Merlot, the classification was consistent with the scientific literature on vine mortality in France (Fussler *et al.*, 2008) and the sensitivity of different varieties to trunk diseases (Lade *et al.*, 2022), especially the fungal disease Esca (Borgo *et al.*, 2016). For example, Syrah has been described as being more sensitive to fungal diseases than Merlot in these three publications. To the best of our knowledge, there is no comparative study on the other varieties in the scientific literature; therefore, this classification represents new knowledge in this regard. As Viognier, Chasan, Mourvèdre and Terret blanc are represented less than three times in the dataset, clustering results of these varieties would require further confirmation.

Table 2 presents the result of the ANCOVA on the proportion of DMV per field using vine age, variety mortality (based on Figure 2) and available ancillary variables. As expected, and according to Figure 2, the primary factor explaining the differences in the proportion of DMV in the fields was vine age. The older the vineyard field, the more DMV there were. In total, vine age accounted for 47.4 % of the variance explained by the ANCOVA, which was followed by variety, categorised as either "high", "intermediate" or "low" mortality. This result was also expected, since this effect was already highlighted by the regression shown in Figure 2 and the resulting

TABLE 1. Classification of variety mortality over time. Low, intermediate and high mortality are defined by comparing the regression coefficient of the variety to the regression coefficient, including all varieties and its standard error. Asterisk (*) indicates varieties represented only once or twice in the dataset.

Low mortality	Intermediate mortality	High mortality
Merlot	Grenache	Syrah
Viognier *	Marselan	Aranel *
Chasan *		Caladoc *
Petit Verdot *		Mourvèdre *
		Terret blanc *

TABLE 2. ANCOVA of the proportion of DMV. Explanatory variables are vineyard age, variety mortality type, field average soil resistivity and field average GLCV.

Effect	Degree of freedom	F	p-value	Generalised Eta Squared
Vine age	1	18.887	0.0003	0.474
Variety mortality type	2	4.415	0.0250	0.296
Field average soil resistivity	1	0.211	0.6500	0.010
Field average Fcover	1	4.837	0.0390	0.187

classification. Overall, variety type represented 29.6 % of the variance explained by the ANCOVA. Additionally, the ANCOVA revealed that the proportion of DMV tended to be higher when the field had a low Fraction of Vegetation Cover (FCover). The Fcover represented 18.7 % of the variance explained by the ANCOVA. This result was also expected (Vélez *et al.*, 2020). DMV necessarily leads to a local decrease in biomass. Beyond a certain level, this was naturally detected by a decrease in the average Fcover values at the field level. Note that the average soil electrical resistivity of the field did not explain any differences in the proportions of DMV between the fields in these conditions. Despite being associated with soil properties and developmental conditions, such as soil water availability, soil resistivity was an integrative indicator that was not *a priori* sufficient to explain the differences in vine mortality between the fields in these conditions.

2. Sampling estimation of DMV

The first part of the results provided information on the relationship between the proportion of DMV in the fields and readily available data, such as the properties of these fields and ancillary biomass and soil data. This second part focuses on estimating this same proportion through field observations made by sampling.

Figure 3 represents the estimation errors of the proportion of DMV resulting from a row-based random sampling (RBRS). Considering a field with r different rows, the figure shows results with $r \times 1000$ repetitions (1000 repetitions with $n = 2$ row RBRS, 1000 repetitions with $n = 1$ rows RBSR, 1000 repetitions with $n = 3$ rows RBSR, […], 1000 repetitions with $n = r$ rows RBSR.). Each repetition was then plotted based on relative sampled area and its resulting estimation error (Equation 3). Light pixels contain a large number of repetitions, while darker pixels contain fewer repetitions.

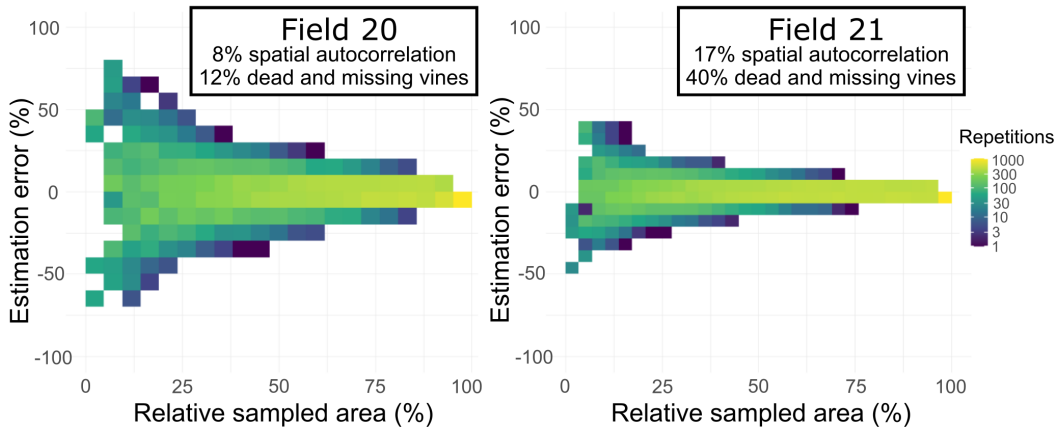

FIGURE 3. Distribution of errors when estimating the proportion of DMV with row-based sampling depending on the relative sampled area. Results are shown for Fields 20 (left) and 21 (right).

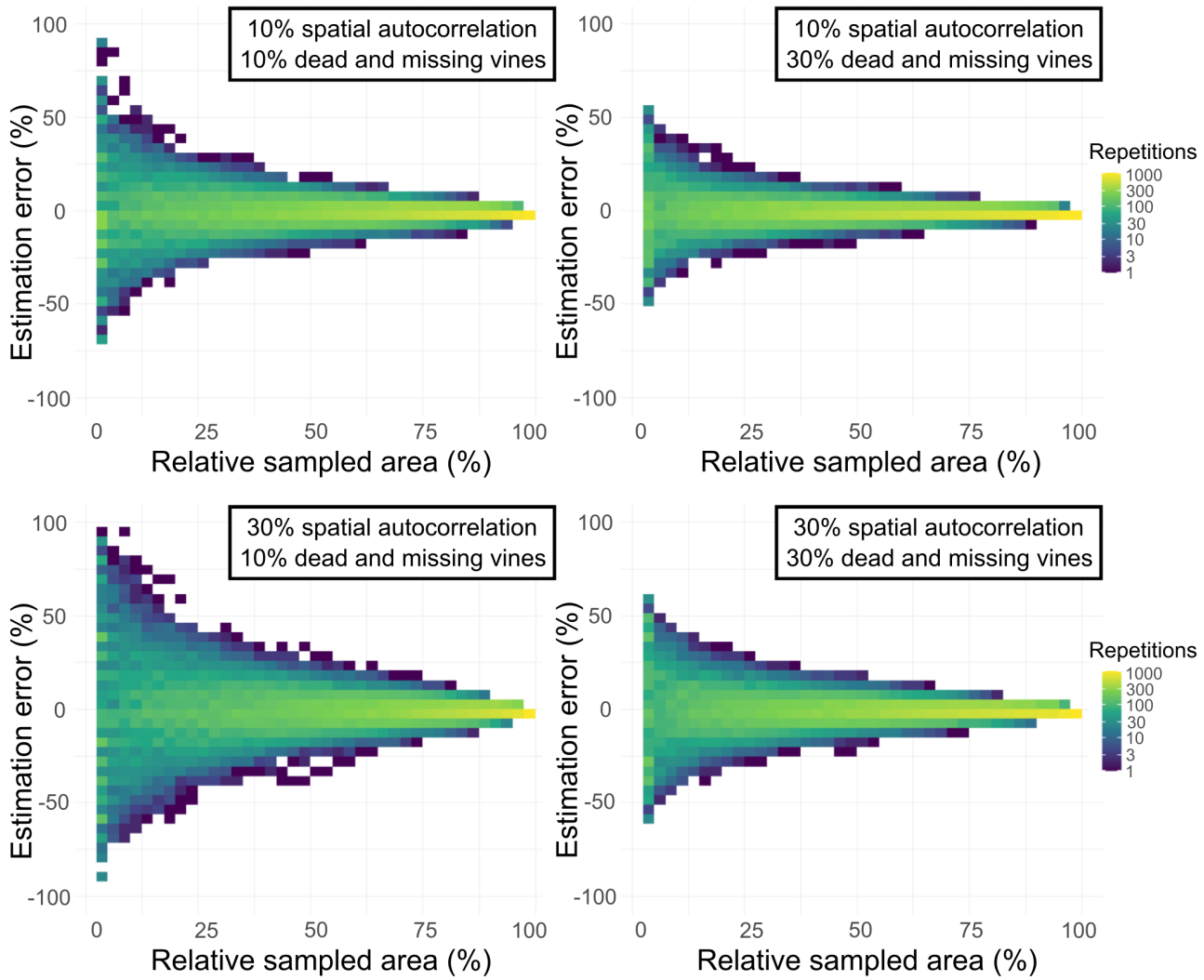

FIGURE 4. Distribution of estimation errors when estimating the proportion of DMV with row-based sampling depending on the relative sampled area. Results are shown for four simulated fields with different amounts of DMV and different levels of spatial autocorrelation.

Logically, for both fields, the magnitude of variation of estimation errors was larger when a small area of the field was sampled. As the relative sampled area increased, the magnitude of variation of errors decreased until the relative sampled area reached the whole field (100 % of the field area). At this point, the error was systematically zero, as the entire field had been sampled and the DMV exhaustively counted. The overall distribution thus takes the form of a narrowing beam centered on 0 %. Regardless of the relative sampled area, more repetitions were closer to a 0 % error than to the border of the beam. This result shows that the estimation based on random sampling was unbiased; i.e., that the distribution of the estimation errors was centred on 0. These characteristics regarding the distribution of estimation errors were consistent across all 29 real fields and all simulated fields in this study (results not shown).

The results in Figure 3 show that, although the fields were geographically close and the general shape of the two error distributions was similar, a higher magnitude of variation of errors was observed for Field 20 than for Field 21; *i.e.*, the "beam of errors" was wider for Field 20. This result may seem surprising at first glance since the number of missing vines was higher for Field 21. Beam width (*i.e.*, the estimation uncertainty) can be expected to be greater when the proportion of DMV is higher. The following section delves deeper into these aspects through a specific study conducted on simulated fields in order to better analyse how the estimation error was affected by some field characteristics.

Figure 4 shows the estimation errors of the proportion of DMV obtained through RBRS in simulated fields. The fields under consideration had 40 rows and the distribution of errors resulted from 40,000 repetitions for each graph. Figure 4 focuses on four chosen simulated fields whose number of DMV and level of spatial autocorrelation were controlled. The distributions obtained were very similar to those obtained from the real data (Figure 3).

In Figure 4, the effect of the proportion of DMV is clearly visible. The ranges of errors in the graphs on the left (10 % DMV) are larger than in those on the right (30 % DMV). An increase in the proportion of DMV led to a decrease in

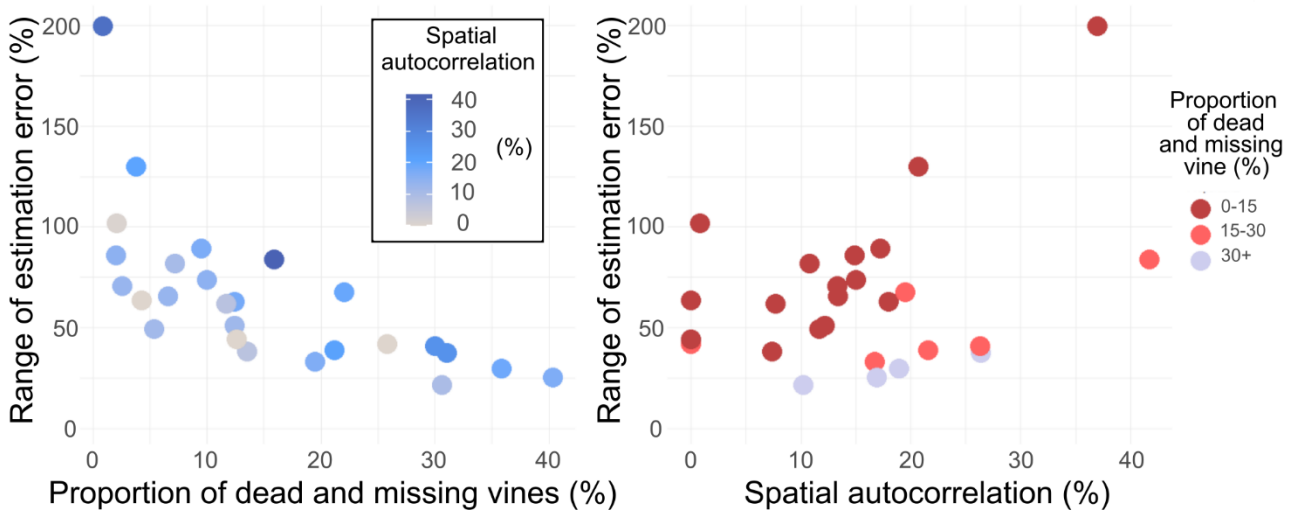

FIGURE 5. Effect of the proportion of DMV (left) and the spatial autocorrelation (right) on estimation error distribution for real fields. Estimation error distribution is represented by the range of estimation error when 50 % of the field is sampled.

the dispersion of errors regardless of the % of the sampled field area. The effect of spatial autocorrelation is also visible. The ranges of errors in the top graphs (10 % spatial autocorrelation) are lower than in the bottom graphs (30 % spatial autocorrelation). This result shows that, for the same relative sampled area, the estimation uncertainty was higher when there was more spatial autocorrelation. Based on extensive simulations (results not shown) in which spatial autocorrelation ranged from 1 % to 99 %, these results can be generalised to any spatial autocorrelation value and to proportions of DMV up to 50 %.

Figure 5 is composed of two scatter plots representing the effect of the proportion of DMV (left) and its spatial autocorrelation (right) on estimation errors for all of the real fields of the study. To be able to compare the error distributions observed in Figures 3 and 4, an indicator called the "Range of estimation error" was defined. This indicator corresponds to the difference between the maximum and minimum values when 50 % of the field area is sampled. This represents a significant sampling effort and may be somewhat unrealistic, but it was only used here for the comparison of fields based on the dispersion of estimation errors.

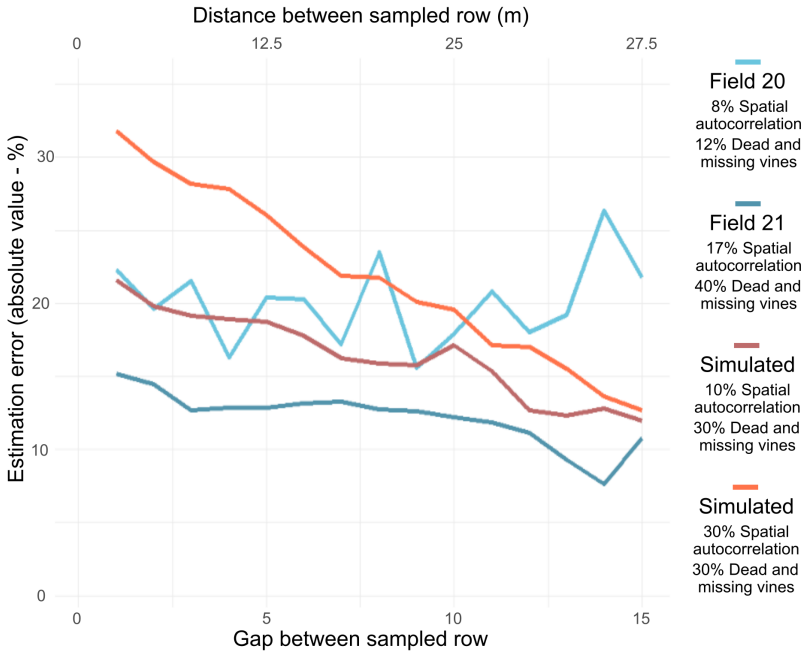

FIGURE 6. Average absolute estimation errors of the proportion of DMV with 10,000 random samples based on two rows. The errors are represented as a function of the gap (number of rows) or of the distance between the two sampled rows.

In Figure 5, a clear decreasing trend is observed in the proportion of DMV plot (left). The higher the proportion of DMV, the narrower the range of estimation errors. This phenomenon was consistent with the simulated data in Figure 4. In the Spatial autocorrelation plot (Figure 5, right), the result is more nuanced, because the proportion of DMV has a dominant effect on the range of estimation error. However, by dividing the measurements into three broad classes, based on the proportion of DMV (0-15 %, 15 -30 % and > 30 %), a similar trend can be observed within each class. The range of estimation error tended to increase as spatial autocorrelation increased. This tendency was stronger when the proportion of DMV in the field was low (bright red line). Overall, these two results for real fields confirm the results that were obtained using the simulated data (Figure 4). The results of the statistical test using a regression model confirm the significance of the relationship between the range of estimation error and both the proportion of DMV and the spatial autocorrelation, with *p*-values of $2.39e^{-06}$ and 0.0012 respectively.

Figure 6 shows the average absolute estimation errors obtained from 10,000 sampling repetitions based on $n = 2$ RBRS for the two real fields and two simulated fields. For clarity, the same fields as those chosen for Figures 3 and 4 have been used here. The absolute estimation errors were expressed as a function of the distance that separated the two sampled rows. Figure 6 focuses on the impact of the distance between sampled rows on estimation errors. Since all fields had the same row spacing (2.5 m), the gap between rows can also be expressed as a distance.

Field 21 and both simulated fields showed the same trend of absolute errors: error decreased with increasing distance between the two sampled rows. This effect was the most pronounced for the simulated field with 30 % spatial autocorrelation, with a decrease of average absolute estimation error of 32 % to 10 %. In the simulated fields, this effect was found to be more pronounced when spatial autocorrelation was high. This result was consistent across numerous simulations (results not shown). With all other parameters fixed, the higher the spatial autocorrelation, the greater the estimation errors when the two sampled rows were closer together. When spatial autocorrelation was high, DMV tended to cluster within specific areas. Therefore, the rows that were close to each other might not be independent, which may increase the risk of obtaining a sample that does not accurately represent the whole field. This trend was difficult to detect with real data since DMV spatial autocorrelation is not well known in advance. Moreover, in this context, each field had unique properties (shape, variety, age, proportion of DMV, etc.), thus it was even more difficult to assess the impact of a single factor, such as spatial autocorrelation.

Field 20 did not share the same trend: the average absolute estimation errors did not clearly decrease. Two factors can explain this difference. Firstly, the field had a low spatial autocorrelation (8 %), which means that the DMV were not very spatially structured. Therefore, the gap between the two sampled rows had less influence on the estimation. Secondly, this field only had 20 rows. Having fewer rows means that fewer row combinations were possible and the effective number of repetitions was lower than for other field that had more rows. The reduced number of possible combinations also explained the lack of curve smoothness for this field.

DISCUSSION

In light of the results from Figures 4 and 5, it is clear that obtaining an accurate estimation of the proportion of DMV (dead and missing vines) is easier for fields in which the proportion of DMV is high. Conversely, the assessment of this proportion within a field characterised by a lower number of DMV appears to be more challenging. Sampling rare events is known to be more challenging than estimating more common events. For example, in a field with only one missing vine, underestimating the exact proportion of DMV vines is highly likely, because the probability of sampling that single vine is low. Moreover, if the sampling does happen to observe this unique missing vine, it will necessarily lead to an overestimation of the proportion of DMV. These overestimations and underestimations, which are related to rare events, are exacerbated by the fact that, for practical reasons, errors are expressed as a percentage of the proportion of DMV. These are the phenomena that explain the patterns observed in Figures 4 and 5.

From a practical point of view, it may be counterproductive to try to precisely estimate the proportion of DMV when there are very few DMV in the field. Having a precise estimation of a low proportion of DMV can be costly and may require a high sampling effort. In the specific context of yield estimations, where each component of the yield must be estimated (proportion of DMV, number of clusters per vine and average cluster weight), an additional factor reinforces this premise. An estimation of DMV is expected to have less impact on the total yield estimation when the proportion of DMV is low. Using the classical relationship of yield estimation (Laurent *et al*., 2021), a rapid estimation shows that a 10 % estimation error on DMV will only result in a 1.25 % error for the total yield estimation for a field with 10 % DMV. Meanwhile, the same 10 % estimation error in DMV will result in a 6.67 % error for the total yield estimation for a field with 40 % DMV. It is therefore more opportune to deploy significant sampling efforts when the expected proportion of DMV is high, because it is easier to estimate and has more impact on the yield estimation. To this end, explanatory factors for vine mortality presented in the first part of the results (Figure 1 and the ANCOVA results in Table 2) offer valuable insights that could assist in defining *a priori* estimates. Age and variety are two field properties that are always known and that can provide important initial information. An effect on vine vigour that was related to ancillary data was also revealed by the results of these analyses, although it was less significant than the age and variety effects. It highlighted that when the latter two effects were removed, there was a correlation between a decrease in vigour and the proportion of DMV. While these results need to be confirmed using more comprehensive datasets that

represent a greater diversity of grape production systems, they represent an interesting tool when comparing fields of the same variety and planted in the same year (assuming planting density and management are equivalent). More broadly, any information that can help to obtain *a priori* estimates would also provide insights that would contribute to defining the required sampling effort. Other ancillary data, even if imprecise or weakly correlated with vine mortality, could thus be valuable.

The results presented in Figures 4 and 5 also reveal that, like other viticulture variables (Oger *et al*., 2023), the proportion of DMV was more challenging to estimate when the distribution of DMV was spatially structured. One possible explanation for this can be found in Figure 6, which shows higher errors when the sampled rows are spatially close. Spatial autocorrelation, in fact, promoted an uneven distribution of DMV across the fields. Consequently, adjacent rows tended to exhibit similar proportions of DMV, potentially leading to overestimations or underestimations when both were sampled. Comparable results were found for most of the studied fields (results not displayed). Nonetheless, this effect was not systematically observed. While strong spatial autocorrelation is expected to increase the likelihood of encountering fields with significantly varying numbers of DMV per row, this did not invariably hold true. Several factors must be considered to interpret this result. First, in vineyards, the variability of DMV and its spatial autocorrelation can be broken down into two orientations: one aligned along the vine rows (intra-row) and another perpendicular to the rows (inter-row). This distinction is supported by vineyard management practices, where viticultural operations are typically performed row by row. From an epidemiological perspective, for example, grapevine trunk diseases often spread anisotropically along the rows, facilitated by pruning activities (Zanzotto *et al*., 2013). Because the sampling method in this study examined entire rows, it was unaffected by the intra-row organisation of DMV. On the other hand, inter-row spatial autocorrelation, which affects the number of DMV from one row to another, explains the observed effects on sampling estimation errors. The spatial autocorrelation measured in this study does not distinguish inter-row and intra-row spatial autocorrelation, nor does it address possible anisotropy. As a reminder, DMV are Boolean (absence-presence) data. Geostatistical analyses, which aim to study spatial autocorrelation, require the conversion of this Boolean data into quantitative data. The simplest approach to this transformation consists of averaging the number of DMV over a certain elementary surface area. This involves defining a regular grid that, when applied to the block, allows for the calculation of a mean number of DMV regularly distributed over the field. Further research focusing specifically on the spatial organisation of DMV and methods for measuring DMV autocorrelation could shed new light on the results presented here. This would also help define the best way to compute the proportion of DMV positions by using a grid with dimensions or shapes that differ from the 5 x 5 m squares employed in this study.

In the scientific literature, of the few studies that have focused on vine mortality (Fussler *et al*., 2008, Merot *et al*., 2023) and on estimating the proportion of DMV in vineyard fields (Robbez-Masson & Foltete, 2005; Desassis, 2005; Matese *et al*., 2019), none have dealt with sampling methods, even though sampling remains the primary means for growers to monitor mortality in their fields. The results presented in this article, based on both real and simulated data, provide new information that can be used to improve the estimation of this often poorly assessed phenomenon in commercial vineyard fields. However, some limitations must be considered. Despite being based on both real and simulated fields that encompassed a wide variety of conditions, the results may not be representative of all viticultural production systems. Additionally, only random sampling was applied here; it would be advisable to validate these findings with other sampling protocols, such as snowball sampling (Griffith *et al*., 2016) and stratified sampling (Wulfsohn, 2010), which could be used to estimate the frequency of a binary qualitative variable using NDVI maps. In this article, only full rows were considered as the spatial sampling support. This choice was based on the fact that counting DMV is simple and quick, and can generally be done in a short time along the entire row. However, there may be situations where counting DMVs across an entire row is impractical, such as when other observational activities (like disease monitoring, yield estimation and maturity assessment) are being carried out, and there is a need for optimisation by combining the DMV estimation with (an)other variable(s) of interest. In such cases, it might be more appropriate to use smaller sampling sites, consisting of just a few vines each. These smaller sites could help to reduce the impact of spatial autocorrelation, provided they are spaced out. However, a drawback is that a larger distance (over a longer time) must be covered to obtain the same sampled area. Another drawback is that it can be counterproductive to consider the same sampling protocol for two variables with different characteristics and distributions. Oger *et al*. (2023) showed that it was, for example, not relevant to combine the estimation of DMVs with a protocol designed to estimate the number and average weight of bunches per vine. It should also be noted that operator-related measurement errors were not considered here: the results relating to the sampling estimation errors assumed that the operator performing the observations perfectly discriminated between DMVs and productive vines. Furthermore, it is worth noting that these results are compatible with an automated detection tool (Fernandes *et al*., 2021). Whether the estimation is carried out by manual sampling or by other equipment, such as a vehicle equipped with cameras and image analysis tools, the results of this study remain relevant when choosing the rows to be observed. However, when the identification of DMVs is not totally accurate, combining measurement errors and estimation errors should be considered. Finally, these findings highlight the need for new sources of data that could provide *a priori* estimates of the proportion of DMV to define a suitable sampling protocol.

The results of this study indicate that an ideal sampling protocol should be based on available data and knowledge of the fields, with which an initial rough estimate can be made to distinguish different types of fields based on the expected proportion of DMV (for example, < 10 %, 10-30 %, > 30 %). If the expected proportion of DMV is high, a significant sampling effort should be deployed to accurately estimate the actual DMV proportion. For instance, in Field 21 (Figure 3), observing 40% of the plot would result in an error of less than 10 % in 90 % of cases. If the expected proportion of DMV is low, sampling can be skipped, as it would add little precision to current knowledge. In intermediate cases, quick sampling can refine the estimate, but precise estimation requires a substantial sampling effort. For example, for Field 20, observing 70 % of the field was necessary to achieve an error of less than 10 % in 90 % of cases. When sampling is based on complete row observations, attention should be given to row spacing to avoid spatial autocorrelation of the measurements. This is especially important for fields in which DMV are spatially structured (*e.g.*, those affected by grape trunk diseases) and the spread of disease is spatial (Agustí-Brisach *et al.*, 2015).

CONCLUSION

This article provides new information that is useful for estimating the proportion of dead and missing vines (DMV) through sampling. Based on an exhaustive set of data acquired from a 20 ha vineyard, it was shown that an average of 2.2 % of the total number of vines in each field over 10 years of age were affected by vine mortality each year. Age, variety and, to a lesser extent, vegetation indices averaged at the field level provided a first rough estimate of the proportion of DMV in vineyard fields and could therefore be used to guide sampling efforts. In this regard, the article introduces a classification of various grapevine varieties commonly cultivated in the South of France based on their mortality rates. In a second step, the study highlighted the link between sampling effort and estimation errors. The results show how the number of DMV and their spatial organisation affect the uncertainty associated with sampling; for example, it seems more opportune to aim for a low estimation error where the DMV proportion is high, and special attention must be paid to the distance between sampling rows (sites) when the DMV are spatially autocorrelated. These findings highlight the value of an initial estimate based on the known properties of the fields, in order to define a sampling effort adapted to the objectives pursued. In available fields, sampling 50 % of the total area resulted in estimation errors of up to 30 % in the worst cases, but was generally sufficient for obtaining estimates with errors of less than 10 %. Overall, this article provides guidelines for DMV sampling that could quickly find practical applications.

ACKNOWLEDGEMENTS

This work was financed by the Occitanie region.

The authors would like to thank Pauline Faure, Romain Girardot, Thomas Crestey and Yoann Valloo from “Le Mas Numérique” for the collection and provision of the data used in this article.

The authors would also like to thank Dr James Taylor for his proofreading.

REFERENCES

Agustí-Brisach, C., León, M., García-Jiménez, J., & Armengol, J. (2015). Detection of grapevine fungal trunk pathogens on pruning shears and evaluation of their potential for spread of infection. Plant Disease, 99(7), 976-981. https://doi.org/10.1094/PDIS-12-14-1283-RE

Ancelin, J., Poulain, S., & Peneau, S. (2022). jancelin/centipede: 1.0 (1_0). Zenodo. https://doi.org/10.5281/zenodo.5814960

Bellehumeur, C., Legendre, P., & Marcotte, D. (1997). Variance and spatial scales in a tropical rain forest: changing the size of sampling units. Plant Ecology, 130(1), 89–98. https://doi.org/10.1023/a:1009763830908

Borgo, M., Pegoraro, G., & Sartori, E. (2016). Susceptibility of grape varieties to esca disease. BIO Web of Conferences, 7, 01041. https://doi.org/10.1051/bioconf/20160701041

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. Statistical Science, 16(2). https://doi.org/10.1214/ss/1009213286

Chanussot, J., Bas, P., & Bombrun, L. (2005). Airborne remote sensing of vineyards for the detection of dead vine trees. In Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS’05. https://doi.org/10.1109/igarss.2005.1526490

Chuche, J., & Thiéry, D. (2014). Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: a review. Agronomy for Sustainable Development, 34(2), 381–403. https://doi.org/10.1007/s13593-014-0208-7

Desassis, N., Monestiez, P., Bacro, J. N., Lagacherie, P., & Robez-Masson, J. M. (2005). Mapping unobserved factors on vine plant mortality. In Springer eBooks (pp. 125–136). https://doi.org/10.1007/3-540-26535-x_11

Di Gennaro, S. F., & Matese, A. (2020). Evaluation of novel precision viticulture tool for canopy biomass estimation and missing plant detection based on 2.5D and 3D approaches using RGB images acquired by UAV platform. Plant Methods, 16(1). https://doi.org/10.1186/s13007-020-00632-2

Fernandes, M., Scaldaferri, A., Fiameni, G., Teng, T., Gatti, M., Poni, S., Semini, C., Caldwell, D., & Chen, F. (2021). Grapevine Winter Pruning Automation: On Potential Pruning Points Detection through 2D Plant Modeling using Grapevine Segmentation. In 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 13–18. https://doi.org/10.1109/cyber53097.2021.9588303

Fontaine, F., Gramaje, D., Armengol, J., Smart, R.E., Nagy, Z.A., Borgo, M., Rego, C., Corio-Costet, M.-F. (2016). Grapevine trunk diseases. A review, 1st edition. OIV, Paris, France. http://www.oiv.int/public/medias/4650/trunk-diseasesoiv-2016.pdf

Fussler, L., Kobes, N., Bertrand, F., Maumy, M., Grosman, J., & Savary, S. (2008). A Characterization of Grapevine Trunk Diseases in France from Data Generated by the National Grapevine Wood Diseases Survey. Phytopathology, 98(5), 571–579. https://doi.org/10.1094/phyto-98-5-0571

Gambetta, G. A., Herrera, J. C., Dayer, S., Feng, Q., Hochberg, U., & Castellarin, S. D. (2020). The physiology of drought stress in grapevine: towards an integrative definition of drought tolerance. Journal of Experimental Botany, 71(16), 4658–4676. https://doi.org/10.1093/jxb/eraa245

Gräler, B., Pebesma, E. J., & Heuvelink, G. B. (2016). Spatio-temporal interpolation using gstat. R J., 8(1), 204.

Griffith, D. A., Morris, E. S., & Thakar, V. (2016). Spatial autocorrelation and qualitative sampling: the case of snowball type sampling designs. Annals of the American Association of Geographers, 106(4), 773–787. https://doi.org/10.1080/24694452.2016.1164580

Kassambara, A. (2023). Rstatix: pipe-friendly framework for basic statistical tests.

Lade, S. B., Štraus, D., & Oliva, J. (2022). Variation in Fungal Community in Grapevine (Vitis vinifera) Nursery Stock Depends on Nursery, Variety and Rootstock. Journal of Fungi, 8(1), 47. https://doi.org/10.3390/jof8010047

Laurent, C., Oger, B., Taylor, J. A., Scholasch, T., Métay, A., & Tisseyre, B. (2021). A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture. European Journal of Agronomy, 130, 126339. https://doi.org/10.1016/j.eja.2021.126339

Matese, A., Cinat, P., Romboli, Y., Berton, A., & Di Gennaro, S. F. (2019). Missing plant detection and biomass estimation from 3D models generated from UAV in a vineyard. In Precision agriculture'19 (pp. 165-172). Wageningen Academic. https://doi.org/10.3920/978-90-8686-888-9_19

Merot, A., Coulouma, G., Smits, N., Robelot, E., Gary, C., Guerin-Dubrana, L., Poulmach, J., Burgun, X., Pellegrino, A., & Fermaud, M. (2023). A systemic approach to grapevine decline diagnosed using three key indicators: plant mortality, yield loss and vigour decrease. OENO One, 57(1), 133–149. https://doi.org/10.20870/oeno-one.2023.57.1.5575

Oger, B., Laurent, C., Vismara, P., & Tisseyre, B. (2023). How to better estimate bunch number at vineyard level? OENO One, 57(3), 27–39. https://doi.org/10.20870/oeno-one.2023.57.3.7404

Oger, B., Laurent, C., Vismara, P., & Tisseyre, B. (2021). Is the optimal strategy to decide on sampling route always the same from field to field using the same sampling method to estimate yield? OENO One, 55(1), 133–144. https://doi.org/10.20870/oeno-one.2021.55.1.3334

Panissod, C., Dabas, M., Hesse, A., Jolivet, A., Tabbagh, J., & Tabbagh, A. (1998). Recent developments in shallow-depth electrical and electrostatic prospecting using mobile arrays. Geophysics, 63(5), 1542-1550.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30(7), 683-691

Primicerio, J., Caruso, G., Comba, L., Crisci, A., Gay, P., Guidoni, S., Genesio, L., Aimonino, D. R., & Vaccari, F. P. (2017). Individual plant definition and missing plant characterization in vineyards from high-resolution UAV imagery. European Journal of Remote Sensing, 50(1), 179–186. https://doi.org/10.1080/22797254.2017.1308234

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org/.

Riou, C., Agostini, D., Aigrain, P., Barthe, M., Robert, M. D., Gervais, J., Jobard, E., Lurton, L., Moncomble, D., & Prêtet-Lataste, C. (2016). Action plan against declining vineyards: An innovative approach. BIO Web of Conferences, 7, 01040. https://doi.org/10.1051/bioconf/20160701040 (https://doi.org/10.1051/bioconf/20160701040)

Robbez-Masson, J., & Foltête, J. (2005). Localising missing plants in squared-grid patterns of discontinuous crops from remotely sensed imagery. Computers & Geosciences, 31(7), 900–912. https://doi.org/10.1016/j.cageo.2005.02.013 (https://doi.org/10.1016/j.cageo.2005.02.013)

Suarez, D. L., Celis, N., Anderson, R. G., & Sandhu, D. (2019). Grape rootstock response to salinity, water and combined salinity and water stresses. Agronomy, 9(6), 321. https://doi.org/10.3390/agronomy9060321 (https://doi.org/10.3390/agronomy9060321)

Tang, J., Woods, M., Cossell, S., Liu, S., & Whitty, M. (2016). Non-Productive Vine Canopy Estimation through Proximal and Remote Sensing. IFAC-PapersOnLine, 49(16), 398–403. https://doi.org/10.1016/j.ifacol.2016.10.073

Thompson, S. K. (2012). Sampling (Vol. 755). John Wiley & Sons.

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data?. Precision Agriculture, 19, 957-971.

Tondriaux, C., Costard, A., Bertin, C., Duthoit, S., Hourdel, J., & Rousseau, J. (2018). How can remote sensing techniques help monitoring the vine and maximize the terroir potential?. In E3S Web of Conferences (Vol. 50, p. 02007). EDP Sciences.

Van Zoeren, J., Martinson, T. E., Caldwels, D., & Walter-Peterson, H. (2020). Missing Parts: The cost of missing cordons, canes and vines. Appellation Cornell, 2020-1

Vaudour, E., Leclercq, L., Gilliot, J., & Chaignon, B. (2017). Retrospective 70 y-spatial analysis of repeated vine mortality patterns using ancient aerial time series, Pléiades images and multi-source spatial and field data. International Journal of Applied Earth Observation and Geoinformation, 58, 234–248. https://doi.org/10.1016/j.jag.2017.02.015 (https://doi.org/10.1016/j.jag.2017.02.015)

Vélez, S., Barajas, E., Rubio, J. A., Vacas, R., & Poblete-Echeverría, C. (2020). Effect of Missing Vines on Total Leaf Area Determined by NDVI Calculated from Sentinel Satellite Data: Progressive Vine Removal Experiments. Applied Sciences, 10(10), 3612. https://doi.org/10.3390/app10103612

Wulfsohn, D. (2010). Sampling techniques for plants and soil. Advanced Engineering Systems for Specialty Crops: A Review of Precision Agriculture for Water, Chemical, and Nutrient Application, and Yield Monitoring, 3-30.

Zanzotto, A., Gardiman, M., Serra, S., Bellotto, D., Bruno, F., Greco, F., & Trivisano, C. (2013). The spatiotemporal spread of esca disease in a Cabernet Sauvignon vineyard: a statistical analysis of field data. Plant Pathology, 62(6), 1205-1213. https://doi.org/10.1111/ppa.12034