IVES Technical Reviews
vine & wine

https://doi.org/10.20870/IVES-TR.2024.8212

How to better estimate bunch number at 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

In viticulture, early-season estimation of average number of grape bunches per vine is crucial for vineyard planning, investments and marketing[1]. Although sampling is widely carried out to estimate grape bunches per vine, there is no clearly established sampling protocol that can be used as a reference when performing these estimations. Each practitioner therefore has their own sampling protocol. In this work, the effect of differences between sampling protocols in terms of estimation errors were investigated to give some recommendations on the best possible sampling practices.

Differences in sampling protocols

Practitioners from several organisations were asked about the sampling practices they apply to estimate the average bunch number per vine, and significant variations were observed. The main differences are related to:

- The counting protocol, which can include or omit missing vines.
- The arrangement of the sampled vines, which can be grouped within sampling sites of varying sizes (a varying number of consecutive vines sampled together along a row)
- The total number of vines sampled per field

The subsequent sections provide practitioners with some recommendations based on experimental results regarding the estimation of the average number of bunch per vine at field level.

Estimation must focus on the number of bunches and ignore missing vines

When performing sampling estimation, the sample size must be tailored to the variable of interest. Therefore, two variables with different properties should be estimated using distinct sampling protocols. In viticulture, for example, when following a sampling protocol in which missing vines are counted as vines with 0 bunches, two yield components are estimated simultaneously: the proportion of missing vines and the number of bunches. Given the inherently distinct nature of these two yield components, however, they should be sampled independently with specific protocols.

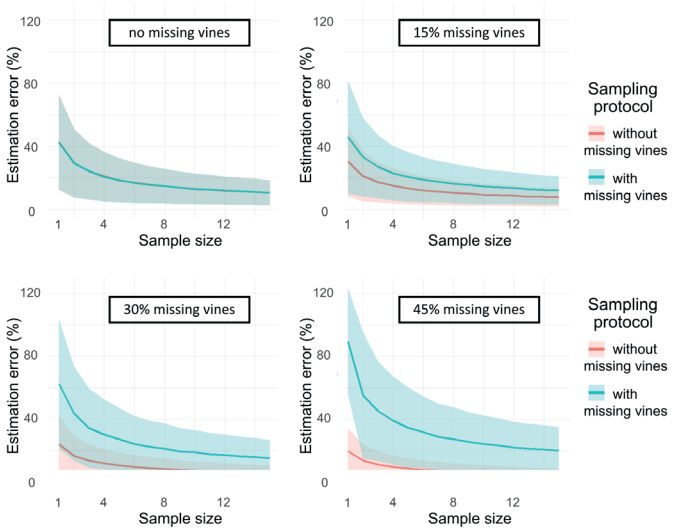


FIGURE 1. Average estimation error of the number of bunches per vine plant (line) and its standard deviation (coloured area). taking (blue) and not taking (red) missing vines into account. The protocols are compared across four different fields with 0%, 15%, 30%, or 45% of missing vines (simulated data).

For the four fields in Figure 1, the estimation error logically decreases as the sample size (number of vines actually sampled) increases, regardless of sampling protocol. The figure also shows that, with an increase in number of missing and dead vines in a field, considering them during the sampling of the average number of bunches per vine (in blue) leads to higher estimation errors. This phenomenon illustrates the need to independently estimate each of the two yield components to reduce estimation errors, especially when the proportion of missing vines in the fields is significant. For the purpose of clarity, the fields shown here are theoretical, but the results are consistent with findings obtained in real fields[2 3].

Observations should be spread over several measurement sites

Yield components are often spatially organised, meaning that closely spaced vines are more likely to exhibit similar properties than more widely spaced ones. In the case of the number of bunches per vine, two closely planted vines are more likely to have a similar number of bunches. By sampling consecutive vines at a single measurement site, there is a risk of overestimating (or underestimating) the number of bunches at the field level if that area has slightly more (or slightly fewer) bunches than the average. It is therefore preferable to distribute observations across multiple measurement sites, at least two or three, to limit this risk.

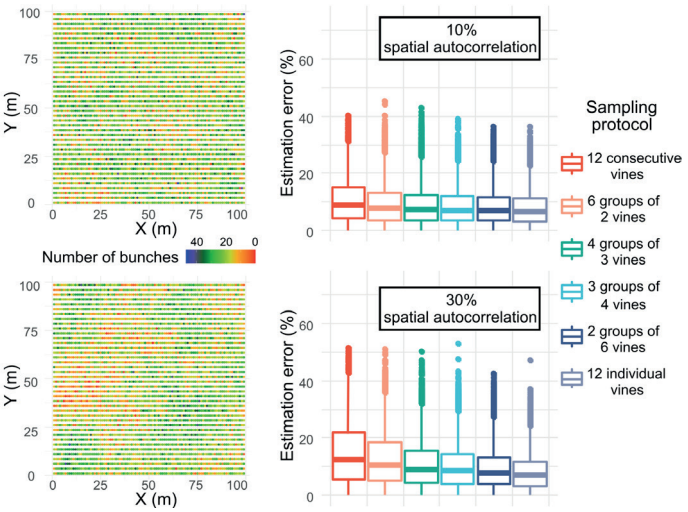


FIGURE 2. Estimation error related to the arrangement of the sampled vines in two fields (simulated data).

Figure 2 represents two fields. In the upper field, the number of bunches per vine is weakly spatially organised: vines with many bunches coexist with nearby vines with few bunches. In the lower

The translation of this article into English was offered to you by Moët Hennessy. Moët Hennessy

field, the number of bunches per vine is more structured: in a close area, most vines have similar bunch numbers. In each field, six sampling protocols distributing 12 vines within 1, 2, 3, 4, 6, or 12 measurement sites are compared.

In both fields, estimation errors are higher when the vines are grouped within a single measurement site (red boxplot) than within at least two measurement sites. Figure 2 illustrates how, with an equal number of observed vines, the more the sampled vines are distributed over a large number of measurement sites, the lower the estimation errors. This phenomenon is even more pronounced when the number of bunches per vine is spatially autocorrelated, as evidenced by the larger estimation errors in the field with 30% of spatial autocorrelation.

Tailoring sample size to observed variability

The final part of this technical review deals with the relationship between estimation error, sample size and sample variability. The sample variability is expressed here in the form of the coefficient of variation (CV). This coefficient is simply calculated by dividing standard deviation of the sample by its mean (Eq. 1):

$$CV = \frac{StandardDeviation}{Mean}$$

This is a standardised way of quantifying values dispersion within the sample. The statistical details of the method can be found in the article associated with this technical review[2].

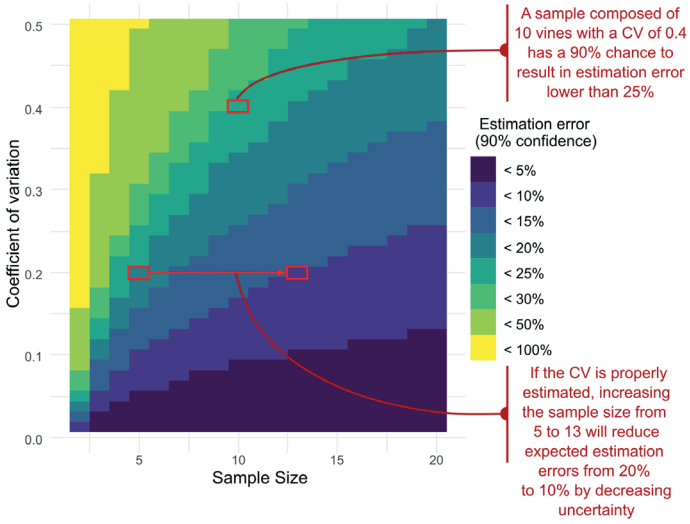


FIGURE 3. Expected estimation error of a sample size based on its size (number of sampled vines) and coefficient of variation.

Figure 3 illustrates how, for a given sample size, estimation errors vary with sample coefficient of variation (CV). It also shows how the variability (represented by the CV) of observations made on a few vines can be used to define an appropriate sample size in order to reach the expected error of estimation. The lower the variability of observations within a sample, the higher the confidence in the estimation. A highly variable sample (high CV) reflects significant field heterogeneity, which is associated with a high risk of error. Therefore, when high variability is observed during sampling, it is advisable to increase the sample size to mitigate high estimation errors. Additionally, increasing the sample size provides more significant gains when sample variability is high. Provided the CV of the sample does not change, increasing the number of observations from 5 to 13 (Figure 3) reduces the upper bound of the confidence interval for estimation errors from ~39% to ~22% if the CV is 0.4, and only from ~19% to ~11% if the CV is 0.2.

By computing its CV and using information provided by Figure 3, it is possible to assess in real-time the quality of a sample. The practitioner can then decide whether to continue sampling by identifying the sample size that, at an equal coefficient of variation, would give the desired precision. This method could also help detect the fields where the sample size required to achieve a desired level of error is unattainable, allowing sampling to be stopped and to focus estimation efforts on other fields. It should be noted that if the new measurements change the coefficient of variation, it will then be necessary to move vertically on the graph to account for this change. For ease of reading, a tabular form of this figure is freely available in an appendix (see Zenodo reference[4]). This appendix also introduces 99%, 95%, 75%, and 50% confidence levels to best fit desired confidence.

Conclusions

The present study highlights what or what not to do when estimating the average number of bunches per vine via sampling at the field level. The first recommendation involves implementing a sampling protocol specific to each yield component. While it may be tempting to simultaneously sample the number of missing vines and the average number of bunches per vine, these two components have different properties and thus they each require specific protocols. The second recommendation refers to the organisation of the measurement sites. While it is possible to group them into sampling sites made of several consecutive vines to accommodate operational constraints, there should be a minimum of 2 or 3 measurement sites to avoid increasing estimation errors. Ultimately, this work demonstrates that defining a single sample size that is applied uniformly across all the fields is counterproductive. Instead, the sample size should be determined based on the observed heterogeneity, choosing the best compromise between estimation error and sampling effort, in terms of time and costs. In the event that the coefficient of variation can be calculated in real-time during sampling, the resources described in this article could be used to determine the precision associated with a sample and to adapt protocols along the way. ■

Acknowledgements: *This work was financed by the Occitanie region..*

1 Laurent, C., Oger, B., Taylor, J. A., Scholasch, T., Metay, 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

2 Oger, B., Laurent, C., Vismara, P., & Tisseyre, B. (2023a). 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

3 Oger, B., Zhang, Y., Gras, J. P., Valloo, Y., Faure, P., Brunel, G., & Tisseyre, B. (2023b). High spatial resolution dataset of grapevine yield components at the within-field level. *Data in Brief, 50*, 109580. https://doi.org/10.1016/j.dib.2023.109580

4 Oger, B., Laurent, C., Vismara, P., & Tisseyre, B. (2024). Sampling to estimate the average number of bunches per vine: table of confidence intervals for estimation errors. *Zenodo*. https://doi.org/10.5281/zenodo.10594995

This work is licensed under a Creative Commons Attribution 4.0 International License.