EXTENDED ABSTRACT

Impact of sample size on yield estimation in commercial vineyards

Carlos Poblete-Echeverria[1], Rubén Íñiguez[2,3], Ignacio Barrio[2,3], Javier Tardáguila[2,3]
*Corresponding author: mail

[1] South African Grape and Wine Research Institute (SAGWRI), Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
[2] Televitis Research Group, University of La Rioja, 26006 Logroño, Spain
[3] Institute of Grapevine and Wine Sciences (University of La Rioja, Consejo Superior de Investigaciones Científicas, Gobierno de La Rioja), 26007 Logroño, Spain

Keywords: yield, sampling strategy, variability, precision viticulture

INTRODUCTION

The accurate estimation of yield is a fundamental for suitable viticulture, playing a pivotal role in the planning of logistics, the allocation of resources and the formulation of commercial strategies. The capacity to make yield projections enables producers to anticipate market demands and optimize operations, enhancing both efficiency and sustainability (Komm & Moyer, 2015). However, the intrinsic spatial and temporal variability of yield in vineyards presents considerable challenges to the collection of representative and precise data (e.g., Bramley and Proffitt 1999; Clingeleffer et al. 2001; Carrillo et al. 2015). Historically, the yield estimation has been dependent on manual techniques, such as the counting of bunches on a restricted sample of vines and the extrapolation of results to the entirety of the vineyard. Despite their widespread use, these techniques have notable limitations. They are costly in terms of labour and time, and they are prone to sampling errors due to the lack of representativeness, particularly in large or spatially heterogeneous vineyards (Dami, 2011).

To address these shortcomings, advanced techniques have been developed that integrate historical data and refined sampling designs. For example, Araya-Alman et al. (2019) proposed a methodology based on historical yield patterns to identify key sampling areas, thereby reducing estimation errors. Similarly, Oger et al. (2021) investigated the potential for optimising the number of sampled vines to minimise errors, emphasising the value of systematic approaches for more accurately capturing spatial variability. The adoption of emerging technologies has further transformed yield estimation practices. Torres-Sánchez et al. (2021) demonstrated how unmanned aerial vehicles (UAVs) equipped with high-resolution cameras and photogrammetric point cloud analysis can provide detailed spatial data on cluster distribution across extensive vineyard areas. In a further development of the techniques described above, Meyers et al. (2011) devised a dynamic spatial optimisation model which enhances the representativeness of the samples taken while reducing the costs of the operation. Moreover, Nuske et al. (2011) implemented computer vision algorithms for automated cluster detection and counting, thereby eliminating the necessity for destructive sampling. The integration of machine learning has expanded the scope of these technologies. Palacios et al. (2023) applied computer vision and machine learning to enable early yield predictions across different grapevine varieties, facilitating adaptive management strategies early in the growth cycle. Similarly, Íñiguez et al. (2024) developed deep learning models capable of detecting grape clusters even under complex occlusion conditions, improving data accuracy in challenging environments.

In this context, all these new methodologies and technologies can help to optimise the sampling strategy. To achieve a satisfactory yield estimation from punctual measurements, the number of measurements must reflect the expected yield variance at the desired scale. However, field measurements represent a significant effort in terms of labour and time, logistics and cost associated with equipment and the technology used. Therefore, a proper definition of sampling size is a key aspect of the success of the yield estimation approach used.

RESEARCH OBJECTIVES

The objective of this study is to evaluate the impact of sample size on the accuracy of grape yield estimation, focusing on how smaller sample sizes affect estimation errors related to yield components (in this case the number of bunches per vine). Additionally, the study aims to provide practical recommendations for improving yield component sampling, considering the inherent spatial variability of the vineyard blocks. It also seeks to assess manual bunch counting as a tool for yield prediction, examining its strengths and limitations while identifying areas for refinement to enhance both accuracy and efficiency. By achieving these goals, the research intends to establish a framework that integrates traditional methods with modern, data-driven approaches, offering valuable insights for optimizing yield estimations in commercial viticulture.

MATERIAL AND METHODS

Experimental sites

The study was conducted during the 2023 growing season in three dryland commercial vineyards of *Vitis vinifera* cv. Tempranillo in the Rioja wine appellation, Spain. The vineyard blocks were selected for its contrasting topographical and agronomic features.

– Block 1: Total area of 16.3 ha, with 1.02 ha studied. Double Royat cordon training system, 2.5 m × 1.2 m spacing, north-south orientation, 545 m altitude, and 1.7% slope.

– Block 2: Total area of 22.21 ha, with 1.1 ha studied. Double Royat cordon training system, planting configuration of 2.5 m × 1.1 m, north-south orientation, 522 m altitude, and 8.7% slope.

– Block 3: Total area of 7.94 ha, with 1.06 ha studied. Double Royat cordon training system, planting configuration of 2.5 m × 1.1 m, north-south orientation, 507 m altitude, and 3.5% slope.

Sampling strategy

To assess the impact of sample size on yield estimation components yield, manual bunch counting was performed in the three vineyards blocks, as reference measurement. Each vineyard block was divided into subplots termed "Mini-plots" and "Micro-plots" to ensure representativeness of the sampled area. The mini-plots, corresponding to the study area described earlier (approximately one hectare of the total plot area), comprised 25 to 40 rows of vines, depending on the shape of the vineyard blocks. Each micro-plot was subdivided into five rows within each mini-plot, providing an additional level of detail for analysis.

Sampling was conducted systematically to balance spatial distribution and operational feasibility. In the mini-plots, the first vine of each panel was sampled, starting from the first vine to the right of the post. Border panels were excluded to minimize edge effects. In the micro-plots, all vines within each row were sampled, except for those in the border panels. For each sampled vine, the number of bunches was manually recorded, and the data were carefully documented for subsequent statistical analysis.

– Block 1: 450 vines in the mini-plot and 360 vines in the micro-plot, totalling 810 vines monitored (40 rows in the mini-plot and 5 rows in the micro-plot, with approximately 12 panels sampled per row).

– Block 2: 350 vines in the mini-plot and 420 in the microplot, totalling 770 vines monitored (20 rows in both the mini-plot and micro-plot, with between 5 and 20 panels sampled per row).

– Block 3: 250 vines in the mini-plot and 350 in the micro-plot, totalling 600 vines monitored (25 rows in the mini-plot and 5 rows in the microplots, with approximately 10 panels sampled per row).

In total, 1,050 vines were sampled in the mini-plots and 1,130 in the micro-plots, generating a total dataset of 2,180 vines across the three vineyard blocks. This systematic sampling approach ensured the inclusion of spatial variability across different vineyard sections, providing robust data for statistical analysis and precise yield estimation.

Simulations

Simulations were conducted to assess the impact of sampling size on the accuracy of yield estimation in vineyards. Specifically, the error associated with different sampling intensities (1%, 3%, 5%, and 10%) was evaluated for estimating the total grape bunch production in three vineyard plots. Manual bunch counts per vine were used as the ground truth. For each plot, 100 random sampling simulations without replacement were performed at each sampling intensity. In each simulation, the total production was estimated by extrapolating the mean bunch count of the sampled vines to the total number of vines in the plot. A customize code was written in MATLAB to perform the simulations.

The estimation error (EE) was calculated as a percentage using the following equation:

$$EE = (EP - AP)/AP \times 100 \qquad (1)$$

where EP is the estimated production (Kg), and AP is the actual production (Kg). EP and AP are defined as cumulative values considering the total number of vines monitored.

The EE values were visualized using bar plots to illustrate the variability in estimates as a function of sampling intensity. This analysis provides insights into the relationship between sampling size and estimation accuracy, enabling the identification of the minimum sampling size required to achieve reliable yield estimates with low error margins.

Statistical analysis

Descriptive statistics were used to summarize the EE values for the different sampling sizes (1%, 3%, 5%, and 10%). For each sampling scenario, the mean and standard deviation of the errors were calculated over 100 iterations per plot. These metrics provided a comparative overview of the variability and accuracy of the yield component estimates. The analysis focused on evaluating how sampling size influenced the precision of the predictions. All computations and visualizations were performed using MATLAB and Python libraries, including pandas and matplotlib.

RESULTS

Variability analysis

The variability in bunch counts among the three vineyard blocks was high reaching a 74.3% of CV in the Block2, underscoring the heterogeneous nature of grape production. Descriptive statistics (Table 1) show disparities in mean bunch counts, ranges, and relative variability as indicated by the CV. Block 1 presented the highest mean bunch count (9.82 bunches per vine) and a relatively controlled CV% of 39.6. Despite this, the wide range of 22 bunches (1–23) highlights notable localized variability. Block 2, presented the lowest mean (6.43 bunches per vine), exhibited the highest variability (CV = 74.3%), indicating substantial heterogeneity likely influenced by environmental or agronomic factors. Its wide range (1–23) further emphasizes the need for intensive sampling. Block 3 showed intermediate values, with a mean of 7.85 bunches per vine and a CV of 49.1%. While its narrower range of 19 bunches (1–20) suggests more uniform production compared to Block 2, significant variability persists.

Graphical analyses using histograms with smoothed density curves (Fig. 1a) further illustrate this variability. Block 1 demonstrates a relatively symmetrical distribution around the mean (~10 bunches) but with notable dispersion, reflecting moderate variability. Block 2 displays a more dispersed and asymmetric distribution centred around lower values (~5 bunches), consistent with its high CV. Plot 3 exhibits a slightly more balanced distribution around its mean (~8 bunches) but still retains variability, as indicated by its range. Fig. 1b shows an example of the spatial variability of the number of bunches per vine in Block 1, where we can see again a confirmation of the high level of variability with a relatively random spatial pattern. These findings underscore the substantial spatial variability present in all study blocks and highlight the necessity of robust sampling strategies to account for heterogeneity and ensure accurate yield estimations.

Sample size effect

Simulation results for all three blocks demonstrated a clear inverse relationship between sampling size and yield estimation error (Figure 2). Sample sizes from 1% to 10% of the block area were assessed, with errors calculated as deviations from actual yield data. Smaller sampling sizes consistently resulted in higher estimation errors, often exceeding ±40% at 1% sampling. This effect highlights the difficulty of adequately capturing vineyard variability with minimal samples. Errors narrowed significantly as the sampling size increased, reaching below ±20% at 3% sampling and below ±15% at 5%. At the largest sampling size (10%), errors stabilized below ±10%, yielding highly accurate predictions across all blocks.

While general trends were consistent, Block 2 showed greater variability in estimation errors compared to Blocks 1 and 3, particularly at smaller sampling sizes. This reflects its higher inherent heterogeneity, necessitating more robust sampling strategies. In contrast, Blocks 1 and 3 exhibited relatively more uniform error distributions, with Block 3 achieving slightly better accuracy at larger sampling sizes. These differences present a clear indication of the importance of tailoring sampling strategies to each vineyard's specific spatial variability.

Larger sample sizes significantly enhance accuracy but must balance against operational constraints like labour and time. Sampling sizes of at least 5% emerge as a practical compromise for manual data acquisition, providing reliable yield estimates while maintaining efficiency. Automatic data acquisition with cameras mounted in robots or other vehicles supported by AI models for bunch counting can provide powerful support for improving the sample size in order to reach a higher level of accuracy in yield estimation.

CONCLUSION

This study demonstrates the critical role of sample size in achieving accurate yield estimations in commercial vineyards. Sample sizes exceeding 10% of the plot area significantly reduced estimation errors, providing reliable predictions for vineyard management. With the levels of spatial variability presented in the studied blocks, a sample size of 1%, commonly used in viticulture, can lead to an error range of ±60% which implies serious problems when this information is used for strategic decision-making.

While manual bunch counting proved effective, its accuracy is highly dependent on the spatial variability of the block and sampling strategy. The results obtained in this study give us useful insights into optimizing vineyard sampling strategies accounting for spatial variability, in order to achieve accurate yield estimations. Future research should focus on integrating traditional methods with emerging technologies such as remote sensing and machine learning to enhance prediction accuracy increasing the sample size. This approach holds potential for improving precision viticulture practices, aligning with industry goals of sustainability and efficiency.

ACKNOWLEDGEMENTS

Research founding FPI Grant 591/2021 by Universidad de La Rioja, Gobierno de La Rioja.

REFERENCES

Komm, B., & Moyer, M. M. (2015). Vineyard yield estimation. Washington State University Extension.

Dami, I. (2011). Methods of crop estimation in grapes. Ohio State University Extension.

Bramley, R. G. V., & Proffitt, T. (1999). Managing variability in viticultural production. The Australian Grapegrower & Winemaker.

Araya-Alman, M., et al. (2019). A new localized sampling method to improve grape yield estimation of the current season using yield historical data. Precision Agriculture.

Oger, B., et al. (2021). How to better estimate bunch number at vineyard level?. OenoOne.

Torres-Sánchez, J., et al. (2021). Grape cluster detection using UAV photogrammetric point clouds as a low-cost tool for yield forecasting in vineyards. Sensors.

Meyers, J. M., Sacks, G. L., Van Es, H. M., & Vanden Heuvel, J. E. (2011). Improving vineyard sampling efficiency via dynamic spatially explicit optimisation. Australian Journal of Grape and Wine Research, 17(3), 311–319.

Nuske, S., et al. (2011). Yield estimation in vineyards by visual grape detection. IEEE/RSJ International Conference.

Palacios, F., et al. (2023). Early yield prediction in different grapevine varieties using computer vision and machine learning. Precision Agriculture.

Íñiguez, R., et al. (2024). Deep learning modelling for non-invasive grape bunch detection under diverse occlusion conditions. Computers and Electronics in Agriculture.

TABLE AND FIGURE

Table 1. Descriptive statistics of bunch counts for the three vineyard blocks.

Block	N	Mean	Median	SD	Variance	Range	Min	Max	CV%
1	383	9.82	10	3.89	15.15	22	1	23	39.6
2	223	6.43	5	4.78	22.85	22	1	23	74.3
3	367	7.85	7	3.85	14.85	19	1	20	49.1

Summary includes the number of vines sampled (N), mean, median, standard deviation (SD), variance, range, minimum (Min), maximum (Max), and coefficient of variation (CV%) for each plot, highlighting differences in variability and yield heterogeneity.

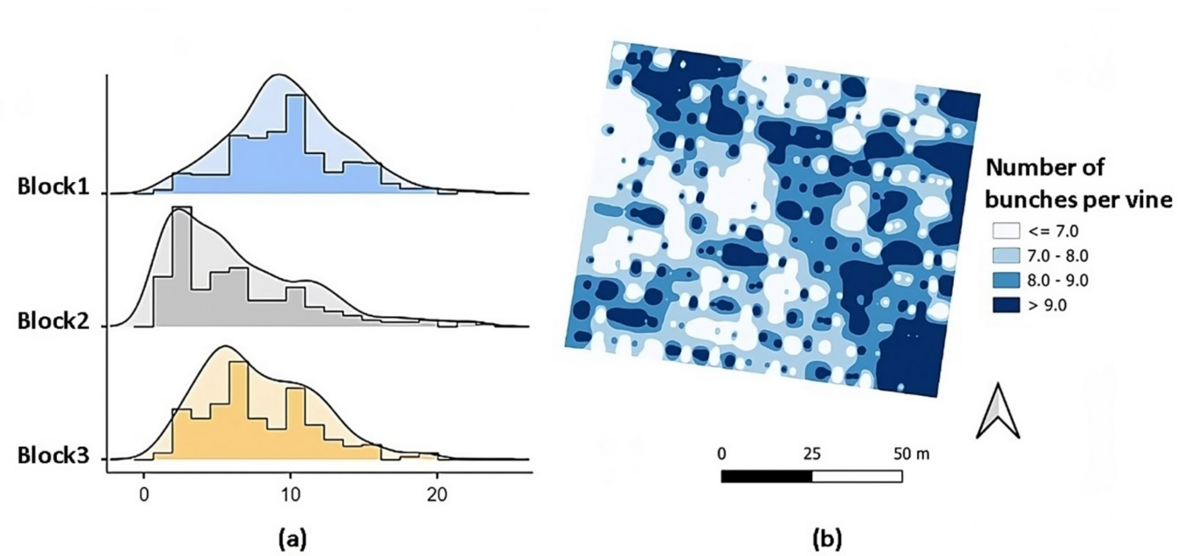

Figure 1. (a) Distribution of bunch counts across the three vineyard plots. Histograms combined with smoothed density curves illustrate the variability in bunch counts for Block 1 (blue), Block 2 (grey), and Block 3 (yellow), highlighting differences in distribution and heterogeneity; (b) Spatial Variability of the Number of Bunches per Vine in Block 1.