Machine-learning methods for the identification of key predictors of site-specific vineyard yield and vine size

James A Taylor, Terence R Bates, Rhiann Jakubowski, Hazaël Jones

AJEV | American Journal of Enology and Viticulture

Research Report
DOI: 10.5344/ajev.2022.22050

Machine-Learning Methods to Identify Key Predictors of Site-Specific Vineyard Yield and Vine Size

James A. Taylor [1]* Terence R. Bates,[2] Rhiann Jakubowski,[2] and Hazaël Jones[1]

Abstract

Background and goals

Lake Erie Concord growers have access to high-resolution spatial soil and production data, but lack protocols and information on the optimum time to collect these data. This study examines the type and timing of sensor information to support in-season management.

Methods and key findings

A three-year study in a 2.6 ha vineyard collected yield, pruning mass, canopy vigor, and soil data, including yield and pruning mass from the previous year, at 321 sites. Stepwise linear regression and random forest regression approaches were used to model site-specific yield and pruning mass using historical spatial production data, multi-temporal in-season canopy vigor, and soil data. The more complex yield elaboration process was best modelled with non-linear random forest regression, while the simpler development of pruning mass was best modelled by linear regression.

Conclusions and significance

Canopy vigor in the weeks preceding bloom was the most important predictor of the current season's yield and should be used to generate stratified sampling designs for crop estimation at 30 days after bloom. In contrast, pruning mass was not well-predicted by canopy vigor; even late-season canopy vigor, which is widely advocated to estimate pruning mass in viticulture. The previous year's pruning mass was the dominant predictor of pruning mass in the current season. To model pruning mass going forward, the best approach is to start measuring it. Further work is still needed to develop robust, local site-specific yield and pruning mass models for operational decision-making in Concord vineyards.

Key words: Concord, proximal canopy sensing, random forests

Introduction

High-resolution agri-data sets, especially from proximal, terrestrial-mounted sensing systems, are available to vineyard managers, but not yet widely adopted commercially (Tardaguila et al. 2021). Following trends in precision agriculture in other cropping systems, spatial canopy vigor data and apparent soil electrical conductivity (EC_a) data have tended to be the main types of data collected (Arnó et al. 2009, Matese and Di Gennaro 2015). These data helped build systems for zonal management (sub-block) to promote differential management (Martínez-Casasnovas et al. 2012, Targarakis et al. 2013, Bonilla et al. 2014) and have also been linked to production attributes, particularly grape yield and quality attributes (e.g., Lamb et al. 2008, Hall et al. 2011, Bonilla et al. 2015). With a few exceptions, most attempts to link ancillary canopy and soil data to vineyard production have focused on data collection at specific phenological stages. For example, the use of imagery around veraison, when vegetative vine development ceases in favor of reproductive (yield) development, has supported estimates of vine size (e.g., Dobrowski et al. 2003, Drissi et al. 2009, Hall et al. 2011, Kazmierski et al. 2011). This is based on the assumption that at veraison, the maximum vine size for the season has been achieved, but the process of senescence, which decreases vine photosynthetically active biomass, does not yet affect the canopy sensor response. However, from an in-season operational point of view, vine size information at veraison in many systems is too late in the season to perform operations that significantly alter crop load (vine balance) via canopy thinning. Avenues to effective vine management for targeted production goals (especially quality) are limited if they depend on information and decision-making at or after veraison.

For effective operational decision-making in-season, producers require information earlier in the season. Early- to mid-season canopy sensor data has been linked to crop production, although published

[1]ITAP, University of Montpellier, Institut Agro Montpellier, INRAE, Montpellier, France; [2]Cornell University, School of Integrative Plant Science, Horticulture Section, Cornell Lake Erie Research and Extension Laboratory, Portland, NY.

*Corresponding author (james.taylor@inrae.fr)

Manuscript submitted Aug 2022, accepted Dec 2022, published Feb 2023

results have been variable and concentrated on wine production systems in warm to hot climates (e.g., Pastonchi et al. 2020, Kasimati et al. 2021, Yu et al. 2021, Sams et al. 2022). These studies have also tended to focus only on univariate analyses, rather than formal multivariate model development, between in-season canopy sensor data and production attributes. Yield elaboration in grapes is known to be a multi-annual process, with primordia development for the yield in year *n* affected by vine conditions in year *n*-1 (Pratt 1971, Laurent et al. 2021). Despite this well-known effect, current site-specific vineyard yield and quality models do not include year *n*-1 data.

The biennial fruiting effect in *Vitis sp.* is of particular importance in systems where a production driver is limiting. Typically, this is either water in non-irrigated, hot climate production or temperature in cool climate production, although poor management can lead to unbalanced vines in any production system. Concord (*Vitis labruscana* Bailey) juice grape production in the Lake Erie American Viticulture Area (AVA) (https://www.ecfr.gov/current/title-27/chapter-I/subchapter-A/part-9/subpart-C/section-9.83 [accessed June 2022]), a cool climate region, operates under such a temperature limitation, and the importance of managing crop load to achieve a sustainable and profitable annual production level is well understood (Bates et al. 2021). If the fruit load set is too large for the vine size (i.e., the leaf area available to generate photosynthate), growers will often perform crop thinning (or be advised to crop thin) to ensure berry maturity at harvest and to protect the return crop the following year. Production parameters, notably the berry growth curve, and production practices dictate that crop estimation and subsequent thinning practices are best performed at ~30 days after bloom in this AVA (mid- to late-July) (Bates 2003, 2017). Therefore, to make good crop thinning decisions, growers need information on the amount of fruit set (yield potential), the vine size at this stage and, additionally, the spatial variability of both these attributes, which do not necessarily follow the same spatial patterning (Bates et al. 2018, Taylor et al. 2019). However, Lake Erie Concord grapegrowers do not have this information currently.

The absence of the right information in mid-July invariably leads to uncertainty in crop thinning decision-making. Action or inaction at this point has potential consequences. Removing fruit in areas where the crop load is good immediately affects (decreases) profit, while not acting to remove fruit in overcropped areas has potential quality control implications at harvest (delivery of immature fruit) and affects the return crop and potential yield/profit the following year. However, once the fruit is set, by dropping fruit the growers are reducing yield and potential income, which they are often reluctant to do. Promoting decision-making and good practices around crop load management relies on good information at the right time and, if it is to be done in a differential manner, good spatial information as well. At the moment, the Lake Erie Concord juice grape industry has no protocols or industry recommendations regarding the best type(s) of data and the best timing(s) of data collection to provide timely in-season crop load information.

Vegetative and reproductive development of any individual vine depends on the environment in which it is grown. It will be influenced by micro and macro-climatic effects and interactions with the soil and local terroir. The vine's vegetative and reproductive development are also interdependent to an extent. However, both processes are influenced by different external factors at different times, so their relationship is not necessarily a direct linear relationship. For example, a large vine in a fertile part of a vineyard may have a low fruit load in a given year due to adverse weather conditions during development of floral primordia in the previous year. Vines will also naturally compensate and redistribute resources between vegetative and reproductive organs based on local, seasonal conditions. Thus, yield elaboration is complex. Canopy development also depends on multiple, variable environmental conditions, in particular, access to soil water and to thermal units. In this reality, and with increasingly larger access to spatial agri-data sets, the recent rapid rise in machine-learning (ML) algorithms, particularly non-linear methods, should provide insight into how to use new spatial agri-data to improve operational decision-making in vineyards.

Machine-learning algorithms have been widely applied to the issue of yield prediction in agriculture (Chlingaryan et al. 2018). In viticulture, ML has predominantly been applied in image processing for either berry or bunch counting (e.g., Liu et al. 2020, Kierdorf et al. 2022, Palacios et al. 2022) to assist with mid-season yield estimates. However, ML approaches are not limited to image analysis, but can be used to identify preferred predictors (variables) within models, and thus reduce data requirements (Xu et al. 2021), especially in situations where auto-correlated spatio-temporal information is available (Nyéki et al. 2021). However, such applications have not yet been reported in viticulture.

Therefore, the primary aim of this paper is to compare common linear and non-linear ML approaches to site-specific modeling of grape yield and vine size in Concord vineyards, where vine size is defined as the pruned mass of first-year wood on the vine. By using site-specific, spatial historical information on crop load (yield and vine size in the previous year), spatial soil maps, and spatio-temporal canopy information throughout the growing season, the intent is to provide clear information to growers on the optimal type and timing of sensor data, in an operational setting, which will be required to provide the best information to inform site-specific decision-making in these vineyard systems. It was not the intent to develop or test the robustness and transferability of these models, as each vineyard system is likely to require local calibration to make effective predictions (Ballesteros et al. 2020).

Materials and Methods

Site description

All data were collected from a 2.6 ha (6.4 ac) Concord vineyard located at the Cornell Lake Erie Research and Extension Laboratory (CLEREL) (42°22′N; 79°29′W; WGS84). The block is located on a north-facing slope with east-west oriented rows, which differ from the north-south norm in this region. Vines are planted with the industry standard spacing of 2.44 m between vines and 2.74 m between rows (8 ft vine × 9 ft row spacing, in the local vernacular), trained to a single-wire bilateral cordon (~1.83 m or 6 ft), and cane-pruned to 100 to 120 nodes/vine. The trellis is supported by wooden posts after every third vine. The block is managed using commercial best practices (Jordan et al. 1980, Weigle et al. 2020) and is reserved for applied commercially-oriented trials by the Lake Erie Regional Grape Program. The vineyard is not irrigated and there was no in-season canopy management (hedging) or yield thinning performed during the study (2018 to 2021).

Data collection

Sampling scheme

To simplify sampling and record-keeping and to better mimic commercial conditions, the sampling design was a semi-regular grid based on rows and 'panels' (three-vine groupings between wooden posts) within rows. Excluding the end rows and the end panels, where production conditions are different, every second row was sampled, with every second panel sampled within these rows. Row lengths differed slightly (irregularly shaped block), but there were 22 rows sampled with 14 to 15 panels per row, resulting in 321 samples within the vineyard block (Figure 1).

Yield data

Yield data in 2018, 2019, 2020, and 2021 were collected during normal grape harvest operations with an OXBO YieldTracker system on an OXBO 6030 mechanical grape harvester (Oxbo International Corp.). Data from the yield monitor were geolocated with an Ag Leader 7500 WAAS corrected GPS receiver and collected with an Ag Leader 1200 InCommand field computer. In 2018, the harvester was also equipped with an Advance Viticulture Grape Yield Monitor (GYM) system (sensor and data logger) linked to a WASS-corrected Ag Leader 7500 GNSS receiver. The GYM is an effective yield monitoring system in this region (Taylor et al. 2016). A comparison of the Ag Leader and GYM yield sensor data and maps showed a strong correlation between the two sensing systems in 2018 ($r = 0.70$, data not shown). The OXBO YieldTracker yield maps in all four seasons (2018 to 2021) showed coherent patterning and were considered to be a good representation of the spatial yield variance in the block. In all years, the sensor yield data were adjusted to reflect the mean tonnage delivered from the field to the processing plant. The three target years had different mean yield profiles: 2019 was an average year (6.8 Mg/ha) and 2020 was lower yielding (5.4 Mg/ha), resulting (with favorable conditions) in the establishment of an above-average yield in 2021 (11.2 Mg/ha).

Pruning mass (PM) data

The mass of first-year pruned canes was collected and weighed for the entire panel at each of the designated 321 sample locations in the vineyard. A panel is the distance between two posts in the vineyard row, which typically contains three vines and is ~7.3 m (or 24 ft) long. Measurements of the panel associated with each sample point, rather than the individual vines at each sample point, were performed to avoid short-scale stochastic variance effects and were in line with local recommendations for mapping PM (Taylor and Bates 2012, Taylor et al. 2017).

Soil sensing data

In May 2019 and 2020 and June 2021, the vineyard was surveyed with a DualEM 1s sensor (DUALEM Inc.) mounted on a PVC pipe-based sled and towed behind an all-terrain vehicle (ATV). The sensor travelled along the center of every second interrow (~1.35 m from the line of the vine trunks and their supporting wires). EC_a was recorded at two depths of ~0.5 m and ~1.6 m (shallow and deep, respectively). Sensor data were recorded with a GeoSCOUT X field data logger with an internal GPS receiver (Holland Scientific). The high-resolution soil maps in all years were very similar ($r > 0.95$, data not shown), which was expected because the vineyard is in a cool-climate region and in spring (May/June), the soil is typically near field capacity following high precipitation from snowfall, with little evapotranspiration over the winter months. Therefore, if the data are correctly collected, the maps should reflect stable textural differences across the block.

Figure 1 Location of the midpoint of the sampled panels within the 2.6 ha study block at the Cornell Lake Erie Research and Extension Laboratory, Portland, NY.

Phenology data

The experimental station tracks the dates of the main phenological stages for the region, including budbreak, bloom, veraison, and maturity/ripening profiles leading up to harvest. Dates of budbreak, bloom, and veraison were recorded at the 50% achievement date (Table 1). These dates were used to synchronize the calendar dates of the canopy surveys to the phenological stages.

Canopy sensing data

Canopy surveys were performed using the CropCircle ACS-430 (Holland Scientific Inc.) to sense the side curtain of the canopy, mounted on an ATV as described (Taylor et al. 2017). The ACS-430 is a three-band active multispectral sensor that collects reflectance information in the red (670 nm), red-edge (730 nm), and near-infrared (780 nm) regions of the electromagnetic spectrum. Two sensing systems were used and oriented to either side of the ATV to image both left and right (different rows) as the sensing platform passed down the interrow. Every second row was traversed by the ATV. Therefore, the sensors captured data from one side of every canopy row, i.e., both the sampled and non-sampled rows in the vineyard. For early-season surveys, before the side curtain of the canopy had started to develop, sensors were oriented at the high-wire cordon (~1.8 m height) then progressively lowered as shoots lengthened until a minimum height of 0.8 m. There were eight, 13, and 18 campaigns carried out in 2019, 2020, and 2021, respectively, generating a relatively dense time-series of data, especially in the latter years.

Data analysis

Pruning mass data existed as manual measurements at each sample point; however, the yield, soil EC_a, and canopy-sensing data were collected from a moving vehicle at 1 Hz and generated irregular data points. To collate the PM and sensor data, the sensor data were interpolated onto the 321 sample sites using block kriging (7 m^2) with a local variogram structure using Vesper shareware (Minasny et al. 2005). The choice of block size reflected the panel area from which the PM measurements were derived.

For each data type, histograms of the data were generated and nonsensical values, e.g., yield <0 Mg/ha or normalized differences vegetative index (NDVI) >1 and NDVI <0, were removed in a first step, before a manual light-touch data-cleaning was performed to remove outlying points. In all cases, less than 3% of data were removed in this step. For the EC_a data, both the shallow and deep responses were interpolated. For the CropCircle response, the three bands, red (R), red-edge (RE), and near-infrared (NIR), were interpolated individually (i.e., three interpolations performed at each date), before the interpolated bands were used to construct seven different vegetative indices (VIs) using combinations of the three bands (Table 2). This made reconstruction of the various VIs relatively simple. An alternative, more laborious process would be to calculate each VI from the cleaned band data and then interpolate each individual VI (i.e., seven interpolations at each date). The band interpolation approach was preferred here. The manually-measured PM and interpolated yield data were used to create Crop Load values for each site for 2018 to 2020.

After interpolation and processing, a spreadsheet was generated with yield and PM for four years (2018 to 2021), crop load (2018 to 2020), soil EC_a deep and shallow (2019 to 2021), and the seven VIs at multiple dates from 2019 to 2021 (Table 3), which were all co-located on the center of the panel (three-vine section) in the vineyard that was the basic sampling unit. This formed the data set used in the modeling exercise.

Modeling

Stepwise multivariate linear regression (S-MLR) was selected as the linear modeling approach to be tested, while random forest regression (RFR) was used for the non-linear approach. A stepwise approach to linear regression was used to avoid over-fitting with the large number of

Table 1 Day of the year (and date) for three key phenological stages in 2019 to 2021 at the Lake Erie Research and Extension Laboratory. Bloom +30 is same date in July from the June date.

Year	Budbreak	Floraison (bloom)	Veraison
2019	128 (08/05)	171 (20/06)	238 (26/08)
2020	136 (15/05)	166 (14/06)	234 (21/08)
2021	110 (20/04)	158 (07/06)	232 (20/08)

Table 2 Vegetative indices (VIs) calculated from the three available bands of the CropCircle 430 canopy sensor.

Name	Abbreviation	Formula	Reference
Normalized differences vegetation index	NDVI	(NIR-R)/(NIR+R)	Rouse et al. 1974
Simplified difference vegetation index	DifVI	NIR – R	Adapted from Richardson and Wiegand 1977
Simple ratio (or plant cell density/relative veg. index)	SR (PCD/RVI)	NIR/R	Jordan 1969
Normalized differences red-edge	NDRE	(NIR-RE)/(NIR+RE)	Barnes et al. 2000
Modified simple ratio	MSR	R/sqrt((NIR/R)+1)	Chen 1996
Red-edge chlorophyll index	RECI	(NIR/RE)-1	Gitelson et al. 2003
MERIS terrestrial chlorophyll index	MTCI	(NIR-RE)/(RE-R)	Dash and Curran 2004

highly-correlated spatio-temporal VI data layers available in the models. For both approaches, four basic model constructions were tested. These were:

Model 1: Predictions using only historical vine production data (yield, PM, and crop load from the previous year, i.e., year *n*-1) and preseason soil information (deep and shallow EC_a). This tests the hypothesis that vegetative and reproductive development in year *n* is predominantly driven by the previous season's (year *n*-1) yield and PM.

Model 2: Predictions using spatiotemporal in-season canopy observations from early- to late-season surveys. This tests the hypothesis that the evolution of the vine canopy in year *n* is the main driver of yield and PM in year *n*, i.e., it is in-season development, and not year *n*-1 development, that drives production.

Model 3: Combines the predictors from Models 1 and 2 to predict yield and PM. This tests the hypothesis that yield and PM in year *n* is influenced by production in year *n*-1 and vine development throughout the season in year *n*.

Model 4: Presents a simplified version of Model 3, where canopy information is limited to a single survey just prior to the date of crop estimation in these vineyard systems (Bloom date + 30 days). This considers that multi-temporal surveys are not always feasible and the best time to generate information from a single survey is likely to be when canopy development is approaching maturity (full vine size), and just before growers need information to inform crop estimation.

RFR modeling

Random forest algorithms can be used for either classification or regression (Breiman 2001). In this study, to predict continuous vineyard variables (yield and PM), the RFR approach was used. Briefly, the random forest algorithm is a combination of decision trees (Rokach and Maimon 2005). Each tree is generated from values taken randomly from the inputs available, making each tree slightly different. The result of the ML algorithm comes from the average result of many trees (the number of trees is a parameter of the algorithm).

The RFR was run for each Model type (M1 to M4), respecting the availability of predictor variables for each Model type. For model training, regardless of Model type, 10 iterations were performed, with the data set randomly separated for each iteration into a training and a test data set, with 70% of points assigned to the training set and the remaining 30% to the test data set (equivalent to 224 and 97 sites, respectively). The output of the RFR for each Model type was used to calculate the score of explained variance (EV) between the observed (y) and predicted ($\hat{y}$) test data (Equation 1) and mean absolute error (MAE) (Equation 2) as indicators of model performance.

$$\text{Explained Variance} = 1 - \frac{Var(y-\hat{y})}{Var(y)} \quad \text{(Eq. 1)}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\hat{y}-y|}{n} \quad \text{(Eq. 2)}$$

The order and the power of each predictor variable selected in the RFR was also extracted and the first five most powerful predictors recorded. RFR was implemented in Python using the package Scikit-learn (mainly RandomForestRegressor and metrics) (Pedregosa et al. 2011) with the following fixed parametrization: number of estimators (trees) = 150, maximum number of features the RF considers

Table 3 Dates of canopy sensing surveys during the three years of the study translated into phenological time (before or after budbreak, floraison, and veraison) to indicate the asynchronicity of vine phenology between years. DABB, days after budbreak; DBF, days before floraison (bloom); DAF, days after floraison; DBV, days before veraison; DAV, days after veraison.

Date of canopy surveys	Timing relative to phenology		
	2019	2020	2021
06 May			16 DABB
10 May			20 DABB
14 May			24 DABB
16 May	8 DABB		
21 May			17 DBF
26 May		18 DBF	
27 May			11 DBF
31 May	20 DBF		
01 June		13 DBF	
03 June			4 DBF
07 June			Floraison
09 June		5 DBF	
10 June	10 DBF		
15 June		1 DAF	
16 June			9 DAF
17 June	3 DBF		
24 June	4 DAF		17 DAF
26 June		12 DAF	
29 June			22 DAF
01 July		17 DAF	
09 July		25 DAF	32 DAF
15 July		31 DAF	
19 July	29 DAF		
20 July			31 DBV
24 July		28 DBV	
27 July			24 DBV
01Aug	25 DBV		
03 Aug		18 DBV	17 DBV
11 Aug			9 DBV
16 Aug			4 DBV
21 Aug		Veraison	
30 Aug	4 DAV		
03 Sept		13 DAV	
07 Sept			18 DAV
14 Sept		24 DAV	
16 Sept			27 DAV

to split a node = 40, minimum sample leaves in a node = 1 leaf. These values were selected using a sensitivity analysis based on curve fitting to identify suitable values for these data and models.

S-MLR modeling

Full linear models using all relevant predictors for each Model type (M1 to M4) were constructed in R (R Core Team 2022). The step function in the olsrr package (Hebbali 2020) was used to generate the most parsimonious model using a forward step approach and a threshold value of p = 0.01 to accept a new predictor into the model. Model evaluation was achieved by using a cross-validation with the same training and test data sets established for the RFR approach applied independently to the yield and to the PM-dependent variables. For each training-test pair (10 iterations), the S-MLR model was constructed on the training set and then applied to the test set. For each Model and dependent variable, the number and order of predictors selected in each of the iterations were recorded. The dominant predictor selected at each stepwise iteration, along with the number of times it was selected among the 10 iterations, was then extracted. The EV (Equation 2) from the observed and modeled test data for the 10 iterations was calculated. This provided an equivalent estimate of the variance explained by each Model type.

Mapping

Maps of selected dependent and independent variables used in the modeling were generated by performing local block kriging with a local variogram for the high-resolution sensor data (yield, soil EC_a, VIs), and using block kriging with a global variogram for the manual observations (PM), again using a 7 m^2 block. All interpolation was performed in the Vesper freeware (Minasny et al. 2005). Post-interpolation, but prior to mapping, data values were standardized [0,1] across all layers using Equation 3 so that they could be presented on a common legend.

$$y_{std} = \frac{y - y_{min}}{y_{max} - y_{min}} \quad \text{(Eq. 3)}$$

Where y_{std} is the standardized value for a given attribute and y_{min} and y_{max} are respectively the minimum and maximum values of y within the data (vineyard block).

Results

The direct observations (Table 1) and subsequent transformations (Table 3) show the differences in phenology at given dates (days of the year). Budbreak was the most variable phenological stage, with 26 days difference between 2020 and 2021. However, as the season progressed, the dates of floraison and veraison tended to get closer between years. There were common dates for surveys between years, such as 9 July in 2020 and 2021, that showed phenological differences with a seven-day difference from floraison (Table 3). This illustrated a potential need to time data collection relative to phenology, particularly for temporal canopy surveys, and not by date (day of the year), when determining preferred times for data acquisition in vineyard systems.

Table 4 Explained variance from cross-validation of four different models using different available inputs applied to two different regression approaches (stepwise-multivariate linear regression [S-MLR] and random forest regression [RFR]) across three years (2019 to 2021). The models were recalibrated for each year before cross-validation using relevant available variables. The best-performing model in each year is indicated in bold; RFR results are in italics.

Predicted variable	Model type	Year		
		2019	2020	2021
Yield	M1 - S-MLR	0.000	0.428	0.280
	M1 - RFR	*0.006*	*0.508*	*0.539*
	M2 - S-MLR	0.387	0.565	0.457
	M2 - RFR	*0.558*	*0.685*	*0.577*
	M3 - S-MLR	0.484	0.670	0.538
	M3 - RFR	***0.592***	***0.712***	***0.619***
	M4 - S-MLR	0.149	0.465	0.275
	M4 - RFR	*0.254*	*0.554*	*0.543*
Pruning mass	M1 - S-MLR	**0.732**	0.644	0.621
	M1 - RFR	*0.642*	*0.611*	*0.587*
	M2 - S-MLR	0.127	0.164	0.089
	M2 - RFR	*0.126*	*0.237*	*0.176*
	M3 - S-MLR	0.730	**0.659**	**0.627**
	M3 - RFR	*0.644*	*0.651*	*0.581*
	M4 - S-MLR	**0.732**	0.651	0.621
	M4 - RFR	*0.639*	*0.625*	*0.585*

Table 5 Mean average error (MAE) (Mg/ha for yield and kg/vine) from cross-validation of four different models that used different inputs (M1 to M4) applied to two different regression approaches (stepwise-multivariate linear regression [S-MLR] and random forest regression [RFR]) across three years (2019 to 2021). The models were recalibrated for each year using the relevant available variables. The best-performing model in each year is indicated in bold; RFR results are in italics. The higher yield MAE in 2021 is associated with a much higher mean yield in this year.

Predicted variable	Model type	Year		
		2019	2020	2021
Yield	M1 - S-MLR	0.442	1.159	2.818
	M1 - RFR	*0.430*	*1.056*	*2.210*
	M2 - S-MLR	0.350	1.024	2.347
	M2 - RFR	*0.278*	*0.836*	*2.015*
	M3 - S-MLR	0.316	0.892	2.213
	M3 - RFR	***0.267***	***0.800***	***1.899***
	M4 - S-MLR	0.397	1.114	2.831
	M4 - RFR	*0.350*	*0.968*	*2.197*
Pruning mass	M1 - S-MLR	**0.082**	0.131	**0.142**
	M1 - RFR	*0.093*	*0.143*	*0.148*
	M2 - S-MLR	0.146	0.217	0.221
	M2 - RFR	*0.146*	*0.206*	*0.211*
	M3 - S-MLR	**0.082**	**0.128**	**0.142**
	M3 - RFR	*0.092*	*0.136*	*0.151*
	M4 - S-MLR	**0.082**	0.131	**0.142**
	M4 - RFR	*0.093*	*0.140*	*0.149*

The EV (Equation 1) and MAE (Equation 2) were calculated for all model iterations (two dependent variables (Yield and PM) × four Model types (M1 to M4) × two regression approaches (S-MLR and RFR) (Tables 4 and 5). For yield modeling (Table 4), the RFR approaches consistently outperformed the equivalent S-MLR approach, with Model 3 (M3) generating the best results from the cross-validation approach. An analysis of the key predictors selected in the M3 RFR approach (Table 6) clearly showed a preference for canopy sensing information in the week before floraison, with this information selected in the top two strongest model predictors in all three years. DifVI appeared to be the most commonly selected VI across the years at this stage, although it was not the only VI with a strong prediction power in any given year, e.g., RECI at four DBF (days before bloom) was selected in 2021. The yield in year *n*-1 was only of importance in 2021 and neither PM nor EC_a were among the top five most powerful predictors in any year. The M1 model, using only historic information, had very poor prediction in 2019 for both linear and non-linear approaches. This is not to discount the value of these layers, especially the soil EC_a maps that often help interpret spatial production patterns, but rather to note that they were not particularly useful for this purpose. Given the lack of predictive power of the soil EC_a layers and the (expected) inter-annual similarities in the layers, obtaining annual soil EC_a scans is unlikely to to be of any real production benefit to growers.

For the PM modeling, linear modeling (S-MLR) performed better than the non-linear (RFR) approach, with M1, M3, and M4, all of which contained the PM in year *n*-1, performing in a similar manner (EV > 0.730). This is because the previous year's PM was the dominant predictor of PM in the current season (Table 6). Model 2, using only in-season canopy data, generated poor prediction fits for both linear and non-linear approaches (EV < 0.237 for all years). Model 3 had slightly better fits (higher EV, lower MAE) than M1 and M4, based on the inclusion of some canopy sensor data in the modeling; however, there was no clear trend in model predictors identified across the three years to indicate a preferred VI to collect, or a preferred date of VI collection (Table 6). To complement the information in Table 6, which shows only predictors from the best-performing models, the top predictors for all model iterations (Models 1 to 4 with S-MLR and RFR for PM and yield) are provided (Supplemental Table 1). These predictors should be considered together with the information in Tables 4 and 5 to determine the quality of prediction from each model type.

Discussion

The principal objective was to compare the accuracy of linear and non-linear algorithms to model site-specific grapevine yield and PM using various, mainly sensor-based, ancillary data layers. The non-linear RFR model worked better to predict yield, while S-MLR was best at modeling PM. Yield determination in grapes is a complex process, starting with primordia development during the previous season and modified by environmental and plant conditions, such as cluster number, cluster size (berries/cluster), and berry weight all the way through to the final harvest. It is a non-linear process and therefore is better modeled using a non-linear algorithm. In contrast, vine PM directly reflects the vegetative vigor of the vine during the season, which in turn is directly influenced by water and nutrient availability/uptake and, indirectly, by crop load. Water and nutrient availability to the vine is itself a result of seasonal conditions in non-irrigated cool climate vineyards. Since this trial involved no differential or variable rate management to the soil or vines to externally influence PM and the crop load was "moderate", general management created no extreme effects. Therefore, PM in this vineyard should be a simple response to seasonal growing conditions, i.e., it is a more straightforward, linear process. Consequently, the simpler linear model was still able to effectively model this vegetative development.

There were four model constructs (M1 to M4), using different potential combinations of input variables and evaluated using linear and non-linear approaches. These input variables were key data layers related to production in the previous year (yield, PM, crop load) and the current season (soil and canopy). The choice of these constructs was based on potential access to these data by growers, with M3 being

Table 6 The key predictors and timing of data acquisition (expressed in phenological time) in each year for the best-performing models identified from Tables 4 and 5. For the random forest regression (RFR), the first five predictors are shown, followed by their predictive power from the cross-validation in parentheses. For the stepwise multi-linear regression (S-MLR), the order reflects the stepwise progression, with the dominant predictor at each step given along with the number of times (out of 10) it was selected during cross-validation. Acronyms for vegetative indices (VIs) are the same as in Table 2.

Variable	Model	Year	Principal (ordered) predictors
Yield	M3 - RFR	2019	DifVI_03DBF (0.115), SR_20DBF (0.0719), NDVI_20DBF (0.0549), MSR_03DBF (0.0531), SR_03DBF (0.0449)
		2020	DifVI_13DBF (0.1667), DifVI_05DBF (0.1015), SR_05DBF (0.0547), NDVI_05DBF (0.0403), NDVI_18DBF (0.0341)
		2021	RECI_04DBF (0.0911), Yield_2020 (0.0738), DifVI_04DBF (0.0469), NDRE_04DBF (0.038), RECI_11DBF (0.0245)
Pruning mass (PM)	M3 - S-MLR	2019	PM_2018 (10)
		2020	PM_2019 (10), RECI_05DBF (10), MSR_13DAV (6)
		2021	PM_2020 (10), Various VIs at various dates…

the universal model that used all potential data sources. Given the complete nature of the inputs used, it is unsurprising that M3 produced the best results for the yield modeling. However, relative to M3, M1 and M4 performed poorly in site-specific yield prediction. These models considered no (M1) or only one (M4) mid-season predictor. Model 2, which used only multi-temporal canopy data, outperformed M1 and M4 and had EVs and MAEs that approached those achieved by M3 in all three years. This similarity in yield prediction between M2 and M3 was expected, because the dominant predictors selected by the non-linear RFR model were VIs (Table 6). Of these predictors, VIs collected in the three weeks leading up to floraison (early-season canopy sensing) were identified as key predictors of yield. Several different types of VIs were selected across the three years; however, the DifVI index was the most common higher-order predictor in the data set. This is in accordance with an industry-wide survey that assessed various VIs against PM in Concord vineyards in this region (Taylor and Bates 2021). However, the choice of DifVI mostly generated only a marginal gain in prediction quality due to the strong collinearity between the different VIs. When canopy data was limited to only a late-season (veraison) survey (M4), yield predictions were poor. These results clearly indicated that in this cool-climate, juice grape system, it is early-season canopy vigor, instead of mid/late-season vigor, that reflects yield development and final yield. Growers should target canopy sensing pre-floraison in these Concord production systems. The spatial pattern of canopy vigor around the time of crop estimation (30 days after floraison) was less representative of yield patterns in the vineyard block in all three years (lower quality of prediction with M4; Tables 4 and 5).

The 2019 yield prediction models that relied on 2018's year *n*-1 data (M1 and M4) performed poorly compared to other models in 2019, or to the equivalent models in the other years (2020 to 2021). The initial reason for this was unclear, so these data and models were verified. The maps (Figure 2) showed that there was a potential management effect in the southern part of the block, with greater (blue) vigor at veraison that translated into greater yield as well. This was an unintentional spatial management effect that will have confounded the model assumptions. Additionally, there was a significant amount of vine renewal work performed spatially in 2018 that may also have impacted the site-specific predictive ability of the 2018 year *n*-1 data sets. By the end of the 2019 growing season, the vines had stabilized and these management effects had been removed or lessened, with the M1 RFR model explaining ~50% of the yield variation in 2020 and 2021. These results highlighted the effect that variable management in a vineyard will have on production

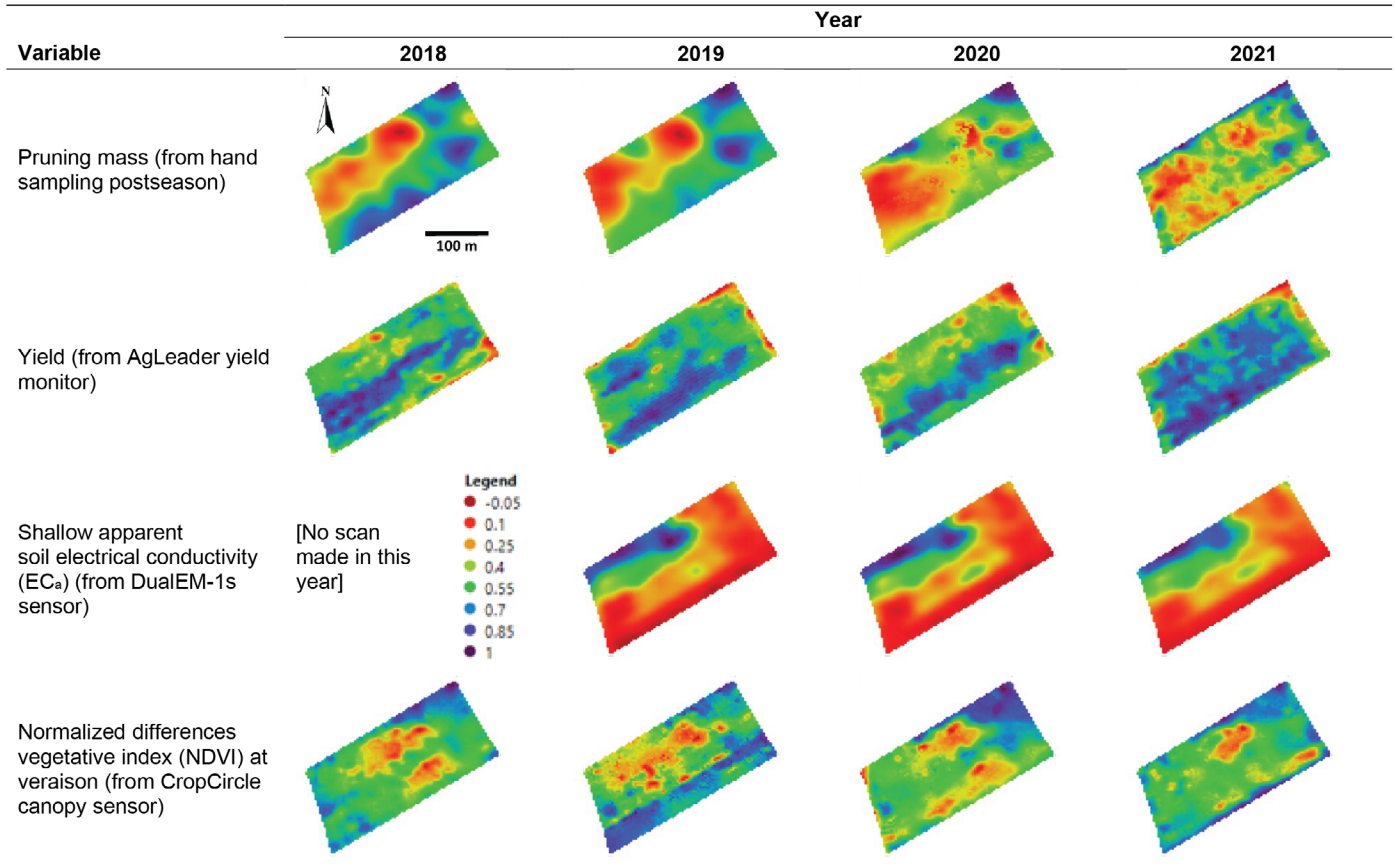

Figure 2 Maps of some key dependent and independent model variables to illustrate spatio-temporal patterning in the block. All data are presented on a common standardized (0 - 1) legend based on the maximum and minimum values in each layer.

modeling. It is also worth noting that explaining 50% of the variance in site-specific yield with a MAE of <3 Mg/ha would still be of value to growers in a management context if further work can demonstrate that the models are robust. However, the objective here was to identify trends and useful predictors for such models, rather than to develop a robust prediction model itself.

For the PM modeling, the results were very different. The year *n*-1 PM data were very dominant as a predictor of the current season's site-specific PM. Vine size and PM in these systems is variable and its dynamics are related to crop load, with under-cropped vines gaining PM, while over-cropped vines will lose PM (Bates et al. 2021). Balanced vines will remain in a stable PM state. In general, the vines in this study block were balanced, with Ravaz index values (Ravaz 1911) in the low- to mid-20s for 2018 to 2019 and <15 in 2020, which should result in little change in site-specific PM from year to year (Taylor and Bates 2013). The strength of the previous year's PM in the PM models reflects this. As this vineyard block has been well managed (well-balanced vines), it is not possible with these data to infer if this relationship will hold true in 'unbalanced' vineyards, where the crop load is low (<10) or high (>30). The relative failure of M2, which used only multi-temporal, in-season canopy information, and the lack of a clear trend in VI predictors in any year (Table 6), was unexpected (EV <0.2 in all three years; Table 4), given that late-season canopy vigor maps have been related to PM in these systems previously (Taylor et al. 2017). This previous work did recognize that PM is highly variable (vine-to-vine) (Taylor and Bates 2012) and that errors (differences) in co-located sensor and manual observations are to be expected. The protocol of Taylor et al. (2017) for relating PM to sensor-based NDVI data did allow for up to 15% of the data to be removed before modeling to improve model fits. In this study, no data were removed or 'cleaned' prior to modeling, but the sample size was 10-fold larger than that of Taylor et al. (2017), and it was expected that this 'noise' in the data would be accounted for in the modeling. However, this did not occur. Further work is needed to better understand the modeling limits here, but the clear indication is that relying only on VIs to model PM will be problematic. If vineyard blocks are well-managed (i.e., maintained at an appropriate crop load), then the clear advice to growers would be to generate a high-quality PM map (from a combination of sensor surveys and manual observation) and to use this map going forward to predict PM. Subsequent years would likely only need minimal manual sampling to update and correct the map.

The results from the yield modeling clearly showed that the most effective information for understanding yield came from proximal canopy sensing performed one to three weeks immediately before floraison (bloom). It is recommended that canopy surveys for yield prediction and for identifying stratified sampling designs for crop yield estimation at 30 DAB should be done at this phenological stage. Prebloom canopy sensing for use in postbloom crop estimation has the added advantage of providing time for the data to be processed and interpreted before crop estimation is performed. The modeling showed that late-season canopy sensing or historical (year *n*-1) production data were less relevant than pre-floraison canopy information to predict spatial in-season yield. In contrast, the best way to predict PM was to start measuring it. Canopy sensing at any phenological stage was not a good direct predictor of PM. Using late-season/veraison canopy vigor and targeted PM measurements for local calibration (Taylor et al. 2017) is one way to start to obtain spatial PM data (and to start to build a temporal history). However, growers have yet to adopt such an approach widely and more automated, grower-friendly measures of vine size (PM or leaf area index) remain a priority for the industry.

From an operational perspective, the quality of the models generated here can be considered to be suitable for commercial management purposes. The MAE of the best yield model varied between years, with differences in mean annual yields, but predictions were 2 to 8% relative error across the three years (absolute errors of 0.3 to 1.9 Mg/ha or 0.1 to 0.8 tons/ac). The best PM modeling was also consistent, but not as good, with 15 to 20% relative error (0.08 to 0.14 kg/vine or 0.2 to 0.3 lbs/vine). Having identified preferred data types and timings of acquisition for site-specific modeling of yield and PM, further work is needed to understand how robust, local models can be developed that are adaptable/transferable between different production systems.

Conclusion

Sensor- and manually-collected data clearly showed that the spatial pattern of the current year's yield potential is represented by the spatial pattern of canopy vigor in the weeks leading up to bloom, i.e., early-season vigor relates to yield potential (and final yield, without crop interventions). Prebloom canopy vigor surveys should be used for directed mid-season crop estimates (30 days postbloom) and to model yield. The spatial patterning of vine PM in balanced vineyards is known to be stable and was shown to be best represented by historic spatial PM information, rather than by spatio-temporal canopy vigor or by spatial soil information. Therefore, the best way to model and manage PM is to start measuring it. This still involves manual observations so more automated ways of PM mapping are required, although veraison canopy vigor mapping remains one way of approximating vine size. Growers should prioritize canopy vigor mapping prebloom and around veraison to provide the most useful information for crop load management. Complex site-specific processes, such as local yield development, were best described by a non-linear model, while local, in-season vegetative growth (PM), a less complex interaction, was best fitted using linear modeling approaches.

Supplemental Data

The following supplemental materials are available for this article at ajevonline.org:

Supplemental Table 1 The key predictors and timing of data acquisition (expressed as phenological time) in each year from all models generated in the study. For the random forest regression (RFR), the first five predictors are shown, with the prediction power from the cross-validation given in parentheses. For the stepwise multi-linear regression (S-MLR), the order reflects the stepwise progression, with the dominant predictor at each step given along with the number of times (out of 10) it was selected in the cross-validation process. Acronyms for vegetative indices (VIs) are the same as in Table 2. Acronyms for phenological stages are the same as in Table 3.

References

Arnó J, Martínez Casasnovas JA, Ribes Dasi M and Rosell JR. 2009. Review. Precision viticulture. Research topics, challenges and opportunities in site-specific vineyard management. Span J Agric Res 7:779-790. DOI: 10.5424/sjar/2009074-1092

Ballesteros R, Intrigliolo DS, Ortega JF, Ramírez-Cuesta JM, Buesa I and Moreno MA. 2020. Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. Precis Agric 21:1242-1262. DOI: 10.1007/s11119-020-09717-3

Barnes E, Clarke TR, Richards SE, Colaizzi PD, Haberland J, Kostrzewski M et al. 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. *In*: Proceedings of the 5th International Conference on Precision Agriculture, pp. 1-16. Bloomington, MN.

Bates TR. 2003. Concord crop adjustment: Theory, research, and practice. Lake Erie Vineyard Notes 6:1-11.

Bates TR. 2017. Mechanical crop control in New York "Concord" vineyards target desirable crop load levels. Acta Hortic 1177:259-264. DOI: 10.17660/ActaHortic.2017.1177.37

Bates T, Dresser J, Eckstrom R, Badr G, Betts T and Taylor J. 2018. Variable-rate mechanical crop adjustment for crop load balance in 'Concord' vineyards. *In* Proceedings of the 2018 IoT Vertical and Topical Summit on Agriculture. pp. 1-4. Tuscany, Italy. DOI: 10.1109/IOT-TUSCANY.2018.8373046

Bates TR, Jakubowski R and Taylor JA. 2021. Evaluation of the Concord crop load response for current commercial production in New York. Am J Enol Vitic 72:1-11. DOI: 10.5344/ajev.2020.20026

Bonilla I, Martinez de Toda F and Martínez-Casasnovas JA. 2014. Vineyard zonal management for grape quality assessment by combining airborne remote sensed imagery and soil sensors. *In:* Proceedings of SPIE 9239, Remote Sensing for Agriculture, Ecosystems, and Hydrology XVI. 92390S. DOI: 10.1117/12.2068017

Bonilla I, Martínez de Toda F and Martínez-Casasnovas JA. 2015. Unexpected relationships between vine vigor and grape composition in warm climate conditions. J Int Sci Vigne Vin 49:127-136. DOI: 10.20870/oeno-one.2015.49.2.87

Breiman L. 2001. Random forests. Machine Learning 45:5-32. DOI: 10.1023/A:1010933404324

Chen JM. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal application. Can J Remote Sens 22:229-242. DOI: 10.1080/07038992.1996.10855178

Chlingaryan A, Sukkarieh S and Whelan B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Comput Electron Agric 151:61-69. DOI: 10.1016/j.compag.2018.05.012

Dash J and Curran PJ. 2004. The MERIS terrestrial chlorophyll index. Int J Remote Sens 25:5403-5413. DOI: 10.1080/0143116042000274015

Dobrowski SZ, Ustin SL and Wolpert JA. 2003. Grapevine dormant pruning weight prediction using remotely sensed data. Aust J Grape Wine Res 9:177-182. DOI: 10.1111/j.1755-0238.2003.tb00267.x

Drissi R, Goutouly J-P, Forget D and Gaudillere J-P. 2009. Nondestructive measurement of grapevine leaf area by ground normalized difference vegetation index. Agron J 101:226-231. DOI: 10.2134/agronj2007.0167

Gitelson AA, Viña A, Arkebauer TJ, Rundquist DC, Keydan G and Leavitt B. 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys Res Letters 30:1248. DOI: 10.1029/2002GL016450

Hall A, Lamb DW, Holzapfel BP and Louis JP. 2011. Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. Precis Agric 12:103-117. DOI: 10.1007/s11119-010-9159-4

Hebbali A. 2020. olsrr: Tools for Building OLS Regression Models. R package version 0.5.3. https://CRAN.R-project.org/package=olsrr

Jordan CF. 1969. Derivation of leaf-area index from quality of light on forest floor. Ecology 50:663-666. DOI: 10.2307/1936256

Jordan TD, Pool RM, Zabadal TJ and Tompkins JP. 1980. Cultural practices for commercial vineyards: New York State College of Agriculture and Life Sciences. Misc Bulletin 111:69.

Kasimati A, Espejo-Garcia B, Vali E, Malounas I and Fountas S. 2021. Investigating a selection of methods for the prediction of total soluble solids among wine grape quality characteristics using normalized difference vegetation index data from proximal and remote sensing. Front Plant Sci 12:683078. DOI: 10.3389/fpls.2021.683078

Kazmierski M, Glémas P, Rousseau J and Tisseyre B. 2011. Temporal stability of within-field patterns of ndvi in non irrigated Mediterranean vineyards. J Int Sci Vigne Vin 45:61-73. DOI: 10.20870/oeno-one.2011.45.2.1488

Kierdorf J, Weber I, Kicherer A, Zabawa L, Drees L and Roscher R. 2022. Behind the leaves: estimation of occluded grapevine berries with conditional generative adversarial networks. Front Artif Intell 5:830026. DOI: 10.3389/frai.2022.830026

Lamb DW, Weedon MM and Bramley RGV. 2008. Using remote sensing to predict grape phenolics and colour at harvest in a Cabernet Sauvignon vineyard: Timing observations against vine phenology and optimising image resolution. Aust J Grape Wine Res 10:46-54. DOI: 10.1111/j.1755-0238.2004.tb00007.x

Laurent CM, Oger B, Taylor JA, Scholasch T, Metay A and Tisseyre B. 2021. A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture. Eur J Agron 130:126339. DOI: 10.1016/j.eja.2021.126339

Liu S, Zeng X and Whitty M. 2020. A vision-based robust grape berry counting algorithm for fast calibration-free bunch weight estimation in the field. Comput Electron Agron 173:105360. DOI: 10.1016/j.compag.2020.105360

Martínez-Casasnovas JA, Agelet-Fernandez J, Arnó J and Ramos MC. 2012. Analysis of vineyard differential management zones and relation to vine development, grape maturity and quality. Span J Agric Res 10:326-337. DOI: 10.5424/sjar/2012102-370-11

Matese A and Di Gennaro SF. 2015. Technology in precision viticulture: A state of the art review. Int J Wine Res 7:69. DOI: 10.2147/IJWR.S69405

Minasny B, McBratney AB and Whelan BM. 2005. VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. http://www.usyd.edu.au/su/agric/acpa

Nyéki A, Kerepesi C, Daróczy B, Benczúr A, Milics G, Nagy J et al. 2021. Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods. Precis Agric 22:1397-1415. DOI: 10.1007/s11119-021-09833-8

Palacios F, Melo-Pinto P, Diago MP and Tardaguila J. 2022. Deep learning and computer vision for assessing the number of actual berries in commercial vineyards. Biosyst Eng 218:175-188. DOI: 10.1016/j.biosystemseng.2022.04.015

Pastonchi L, Di Gennaro SF, Toscano P and Matese A. 2020. Comparison between satellite and ground data with UAV-based information to analyse vineyard spatio-temporal variability. OENO One 54:919-934. DOI: 10.20870/oeno-one.2020.54.4.4028

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. 2011. Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825-2830

Pratt C. 1971. Reproductive anatomy in cultivated grapes- A review. Am J Enol Vitic 22:92-106. DOI: 10.5344/ajev.1971.22.2.92

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ravaz M. 1911. L'effeuillage de la vigne. Ann d L'Ecole Natl d'agriculture Montpellier 11:216-241.

Richardsons AJ and Wiegand CL. 1977. Distinguishing vegetation from soil background information. Photogramm Eng Rem S 43:1541-1552.

Rokach L and Maimon O. 2005. Decision trees. *In* Data Mining and Knowledge Discovery Handbook. pp.165-192. Springer, Boston, MA. DOI: 10.1007/0-387-25465-X_9

Rouse JW, Jr., Haas RH, Schell JA and Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *In* Proceedings of the Third ERTS-1 Symposium. pp.309–317. Washington, DC.

Sams B, Bramley RGV, Sanchez L, Dokoozlian NK, Ford CM and Pagay V. 2022. Characterising spatio-temporal variation in fruit composition for improved winegrowing management in California Cabernet Sauvignon. Aust J Grape Wine Res 28:407-423. DOI: 10.1111/ajgw.12542

Tagarakis A, Liakos V, Fountas S, Koundouras S and Gemtos TA. 2013. Management zones delineation using fuzzy clustering techniques in grapevines. Precis Agric 14:18-39. DOI: 10.1007/s11119-012-9275-4

Tardaguila J, Stoll M, Gutiérrez S, Proffitt T and Diago MP. 2021. Smart applications and digital technologies in viticulture: A review. Smart Agric Technol 1:100005. DOI: 10.1016/j.atech.2021.100005

Taylor JA and Bates TR. 2012. Sampling and estimating average pruning weights in Concord grapes. Am J Enol Vitic 63:559-563. DOI: 10.5344/ajev.2012.12069

Taylor JA and Bates TR. 2013. Temporal and spatial relationships of vine pruning mass in Concord grapes. Aust J Grape Wine Res 19:401-408. DOI: 10.1111/ajgw.12035

Taylor JA and Bates TR. 2021. Comparison of different vegetative indices for calibrating proximal canopy sensors to grapevine pruning weight. Am J Enol Vitic 72:279-283. DOI: 10.5344/ajev.2021.20042

Taylor JA, Sánchez L, Sams B, Haggerty L, Jakubowski R, Djafour S et al. 2016. Evaluation of a grape yield monitor for use mid-season and at-harvest. J Int Sci Vigne Vin 50:57-63. DOI: 10.20870/oeno-one.2016.50.2.784

Taylor JA, Link K, Taft T, Jakubowski R, Joy P, Martin M et al. 2017. A protocol to map vine size in commercial single high-wire trellis vineyards using "off-the-shelf" proximal canopy sensing systems. Catalyst 1:35-47. DOI: 10.5344/catalyst.2017.16009

Taylor JA, Dresser JL, Hickey CC, Nuske ST and Bates TR. 2019. Considerations on spatial crop load mapping. Aust J Grape Wine Res 25:144-155. DOI: 10.1111/ajgw.12378

Weigle TH, Muza A, Brown B, Dunn A, Hed B, Helms M et al. 2020. 2020 New York and Pennsylvania Pest Management Guidelines for Grapes. Cornell University, Ithaca, NY.

Xu H, Zhang X, Ye Z, Jiang L, Qui X, Tian Y et al. 2021. Machine learning approaches can reduce environmental data requirements for regional yield potential simulation. Eur J Agron 129:126335. DOI: 10.1016/j.eja.2021.126335

Yu R, Brillante L, Torres N and Kurtural SK. 2021. Proximal sensing of vineyard soil and canopy vegetation for determining vineyard spatial variability in plant physiology and berry chemistry. OENO One 55:315–333. DOI: 10.20870/oeno-one.2021.55.2.4598