AJEV Papers in Press. Published online January 17, 2023.

AJEV PAPERS IN PRESS • AJEV PAPERS IN PRESS

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Research Article

Machine-Learning Methods for the Identification of 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]

[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 24 Aug, 2022, accepted 20 Dec, 2022

Key words: Concord, proximal canopy sensing, random forests

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 intends to provide clearer information regarding 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 spatial historical production data, multi-temporal in-season canopy vigor and soil data. The more complex yield elaboration process was best modelled with

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 for pruning mass estimation 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.

Introduction

High-resolution agri-data sets, especially from proximal, terrestrial mounted sensing systems, are available for vineyard managers but not yet widely commercially adopted (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 (Arno et al. 2009, Matese and Di Gennaro 2015). These data have helped to build systems for zonal management (sub-block) to promote differential management (Martinez-Casasnovas et al. 2012, Targarkis et al. 2013, Bonilla et al. 2014). These data 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 tends to cease

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

in favor of reproductive (yield) development, have supported estimations of vine size (e.g. Dobrowski et al. 2003, Drissi et al. 2009, Kazmierski et al. 2011, Hall 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 the photosynthetically active biomass of the vine, is yet to have a significant effect on 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 will significantly alter crop load (vine balance) via canopy thinning. Avenues to effective vine management for targeted production (especially quality) goals are limited by 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 the results published 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, site-specific vineyard yield and quality models have yet to be proposed that 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

vines in any production system. Concord (*Vitis labruscana* Bailey) juice grape production in the Lake Erie American Viticulture Area (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 level of production is well understood (Bates et al. 2021). If the fruit load set is too high 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, Bates 2017). Therefore, to make good crop thinning decisions, growers need information on the amount of fruit set (yield potential) and the vine size at this stage and, in addition, they need information on the spatial variability of both these attributes that do not necessarily follow the same spatial patterning (Bates et al. 2018, Taylor et al. 2019). However, Lake Erie concord grape growers do not currently have this information.

The absence of the right information in mid-July invariably leads to uncertainty in the crop thinning decision-making. Action and 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 mature fruit) and affects the return crop and potential yield/profit in the following year. However, once the fruit is set, by dropping fruit the growers are reducing yield and potentially income, which in general they are loathed and risk-adversed to do. Promoting decision-making

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

and good practices around crop load management is very reliant on having 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 will be very dependent 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 will also be interdependent to an extent. However, both processes are influenced by different external factors at different times, meaning that their relationship will not necessarily be 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 the development of the floral primordia in the previous year. Vines will also naturally compensate and redistribute resources between vegetative and reproductive organs based on local, seasonal conditions. The implication is that yield elaboration is complex. Canopy development is also dependent 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 algorithms, particularly non-linear methods, should provide better insights into how to use these new spatial agri-data to improve operational decision-making in vineyards.

Machine-learning (ML) 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 situations for either berry or bunch counting (e.g. Liu et al. 2020,

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Kierdorf et al. 2022, Palacios et al. 2022) to assist with yield estimation mid-season. However, machine-learning approaches are not limited to image analysis, and can be used to identify preferred predictors (variables) within models and to 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 in viticulture have not been reported to date.

Therefore, the primary aim of this paper is to compare common linear and non-linear machine-learning approaches to site-specific modelling 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 aid site-specific decision-making in these vineyard systems. It is not the intent to develop or to test the robustness and transferability of these models, as each vineyard system is likely to require some level of local calibration to have effective prediction models (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) (42.3766, -79.4861, WGS84). The block is located on a north facing slope with E-W oriented rows, which differs from the N-S norm in this region. Vines are planted on the industry standard spacing of 2.44 m between vines and 2.74 m between rows (8 ft vine x 9 ft row spacing in the local vernacular), trained to a single-wire bi-lateral cordon (~1.83 m or 6 ft), and cane pruned to 100-120

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 (2018-21).

Data collection. *Sampling scheme*. To simplify sampling and record-keeping (and mimic conditions closer to commercial situations) the sampling design was a semi-regular grid based on rows and 'panels' (3-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 (irregular shaped block) but there were 22 rows sampled with 14-15 panels per row resulting in 321 samples within the vineyard block (Fig. 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., Lynden, WA). Data from the yield monitor were geo-located with an Ag Leader 7500 WAAS corrected GPS receiver (Ag Leader, Ames, IA, USA) 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) (Joslin, South Australia) linked to a WASS-corrected Ag Leader 7500 GNSS receiver. The GYM has been previously shown to be 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

YieldTracker yield maps in all four seasons (2018-21) 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), 2020 was lower yielding (5.4 Mg/ha) and resulted (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 3 vines and is ~7.3 m (or 24 ft) in length. Measurements at the panel associated with each sample point, rather than the individual vine at each sample point, were performed to avoid short-scale stochastic variance effects and in line with local recommendations for mapping PM (Taylor and Bates 2012, Taylor et al. 2017).

Soil sensing data. In May of 2019 and 2020 and June of 2021, the vineyard was surveyed with a DualEM 1s sensor ((DUALEM Inc., Mississauga, Ontario, Canada) mounted on a PVC pipe-based sled and towed behind an all-terrain vehicle. The sensor travelled along the center of every second inter-row (~1.35 m from the line of the vine trunks and their supporting wires). Apparent electrical soil conductivity (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, Lincoln, NE). It is noted that the high resolution soil maps in all years were very similar ($r > 0.95$, data not shown), which was expected given that that this is a cool-climate region and in spring (May/June) the soil is typically near field capacity

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

following high precipitation (mainly in the form of snowfall) and little evapotranspiration over the winter months. Therefore, if the data is correctly collected, the maps should reflect stable textural differences across the block.

Phenology data. The experimental station records 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, Lincoln, NE, USA) mounted on an All Terrain Vehicle (ATV) following the protocol established by Taylor et al. (2017) in these production systems to sense the side curtain of the canopy. The ACS-430 is a 3-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 inter-row. 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) and then progressively lowered as shoots lengthen until a minimum height of 0.8 m. There were 8, 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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 various 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 t/ha or 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 Infra-Red (NIR) were individually interpolated (i.e. three interpolations performed at each date), before the interpolated bands were used to construct seven different vegetative indices using combinations of the three bands (Table 2). This made reconstruction of the various vegetative indices (VIs) a relatively simple process. An alternative, more laborious process would be to calculate each vegetative index (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 at each site for 2018-20.

After interpolation and processing, a spreadsheet was generated with yield and PM for four years (2018-21), Crop Load (2018-20), Soil EC_a deep and shallow (2019-21) and the seven VIs at multiple dates from 2019-21 (see Table 3 in results for dates), which were all co-located on the

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

center of the panel (3-vine section) in the vineyard that was the basic sampling unit. This formed the dataset used in the modelling exercise.

Modeling. Stepwise Multivariate Linear Regression (S-MLR) was selected as the linear modelling 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 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 pre-season 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 spatio-temporal 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 both Model 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

development is approaching maturity (full vine size) and just before growers need information to inform crop estimation.

Random forest regression modelling. Random forest algorithms can be used for either classification or regression (Breiman 2001). In this study, with the intent to predict continuous vineyard variables (yield and PM), the random forest regression (RFR) approach was used. Briefly, the Random Forest algorithm is a combination of decision trees (Rokach et al. 2005). Each tree is generated from values taken randomly from the inputs available, making each tree slightly different. The result of the machine learning 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 (M1-4) respecting the availability of predictor variables for each Model type. For model training, regardless of Model type, 10 iterations were performed, with the dataset 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 Random Forest regression for each Model type was used to calculate the score of explained variance (EV) between the observed (y) and predicted ($\hat{y}$) test data (Eqn. 1) and mean absolute error (MAE) (Eqn. 2) as indicators of model performance.

$$\text{Explained Variance} = 1 - \frac{Var(y-\hat{y})}{Var(y)} \qquad \text{Equation 1}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\hat{y}-y|}{n} \qquad \text{Equation 2}$$

The order and the power of each predictor variable selected in the RFR was also extracted and the first five most powerful predictors recorded. Random Forest regression was implemented

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 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.

Stepwise Multivariate Linear Regression (S-MLR) modelling. Full linear models using all relevant predictors for each Model type (M1-4) 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. The number and order of predictors selected in each of the iterations, for each Model and dependent variable, were recorded. The dominant predictor selected at each step-wise iteration, along with the number of times it was selected among the 10 iterations, was then extracted. The EV (Eqn. 2) from the observed and modelled test data for the 10 iterations was calculated. This provided an equivalent estimation of the variance explained by each Model type.

Mapping. Maps of selected dependent and independent variables used in the modelling 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

freeware (Minasny et al. 2006). Post-interpolation but prior to mapping, data values were standardized [0,1] across all layers using Eqn. 3 so that they could be presented on a common legend.

$$y_{std} = \frac{y - y_{min}}{y_{max} - y_{min}} \qquad \text{Equation 3}$$

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 in Table 1 and subsequent transformations in 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 then veraison, tended to get closer between years. In Table 3, there were common dates for surveys between years, e.g. 9th of July, that showed phenological differences with a 7-day difference from floraison on this date between 2020 and 2021. This illustrated the potential need to consider the timing of data collection, particularly for the temporal canopy surveys, relative to phenology, and not the date (day of the year), when determining preferred times for data acquisition in vineyard systems.

Tables 4 and 5 show the calculated EV (Eqn 1) and MAE (Eqn 2) respectively for all the model iterations (2 dependent variables (Yield and PM) by 4 Model Types (M1-4) by 2 regression approaches (S-MLR and RFR)). For the yield modelling (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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

before floraison, with this information selected in the top two strongest model predictors in all 3 years. The 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 4 DBF (days before bloom) was selected in 2021. The yield in year *n*-1, was only of importance in 2021, and the PM or EC_a information did not appear in the top 5 most powerful predictors in any year. It is noted that the M1 model, using only historical information had a 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 to 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 be of any real production benefit to growers.

For the PM modeling, the linear modelling (S-MLR) performed better than the non-linear (RFR) approach, with M1, 3 and 4, that all 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 the 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 modelling; however, there was no clear trend in model predictors identified across the three years in regards to a preferred VI to collect or a preferred date of VI collection (Table 6). To complement the information in Table 6, which only shows predictors from the best performed models, the top predictors for all model iterations (Models 1-4 with S-MLR and RFR for PM and yield) are provided in the supplementary information (Table S1). These

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

predictors should be considered together with the information in Tables 4 and 5 on the quality of the prediction from each model type.

Discussion

The principal objective for this analysis was to compare how well a linear and a non-linear algorithm were in modelling site-specific grapevine yield and PM using various, mainly sensor-based, ancillary data layers. The non-linear Random Forest Regression (RFR) model worked better with yield prediction, while the Stepwise Multivariate Linear Regression (S-MLR) was the preferred approach for modelling PM. Yield determination in grapes is a complex process, starting with primordia development during the previous season and influenced by environmental and plant conditions on cluster numbers, cluster size (berries/cluster) and berry weight all the way through to the final harvest, i.e. it is a non-linear process, and is better modelled using a non-linear algorithm. In contrast, the vine PM is a direct reflectance of 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. As there was no differential or variable rate management to the soil or vines to externally influence PM, and the crop load was “moderate”, so that general management was not creating any extreme effects, the evolution of 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 constructs of models (M1-4), using different potential combinations of input variables, evaluated with the linear and non-linear approaches. These input variables were

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 were based on the potential access to these data by growers, with M3 being the universal model that used all potential data sources. It is unsurprising, given the complete nature of the inputs used, that M3 produced the best results for the yield modelling. However, M1 and M4 performed poorly in site-specific yield prediction, relative to M3. Both of these had no (M1) or only one (M4) mid-season predictor in the modelling. Model 2, which only used multi-temporal canopy data, outperformed M1 and M4, and had EVs and MAEs that were approaching those achieved by M3 in all three years. This similarity in yield prediction between M2 and M3 was expected given that 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, i.e. 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 of Taylor et al. (2021) that assessed various VIs against PM in Concord vineyards in this region. However, the choice of DifVI generally only generated 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 it is the early season canopy vigor in this cool-climate, juice grape system, and not the mid/late-season vigor, that reflects yield development and the 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 the 2018 year *n*-1 data (M1 and M4) performed poorly when compared to other models in 2019, or to the equivalent models in the other years (2020-21). The initial reason for this was unclear, and these data and models were verified. The maps (Fig 2) showed that there was a potential management effect in the southern part of the block, with higher (blue) vigor at veraison that translated into higher 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 that may also have locally (site-specifically) impacted the predictive ability of these year *n*-1 (2018) data sets in 2019. By the end of the 2019 growing season, the vines had 'stabilised' 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 issue that enabling variable management in a vineyard will have on production modelling. 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 and not the generation of robust, repeatable prediction models.

For the PM modelling, 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

404 state. In general, the vines in this study block were balanced, with Ravaz index values (Ravaz
405 1911) in the low to mid 20s for 2018-19 and < 15 in 2020, which should either result in little
406 change in site-specific PM from year to year (Taylor and Bates 2013). The strength of the previous
407 year's PM in the PM models reflects this. As this vineyard block has been well managed (well-
408 balanced vines) it is not possible with these data to infer if this relationship will hold true in
409 'unbalanced' vineyards where the Crop load is low (< 10) or high (>30). The relative failure of
410 M2, using only multi-temporal in-season canopy information, and the lack of a clear trend in VI
411 predictors in any year (Table 6) was unexpected (EV < 0.2 in all three years, Table 4), given that
412 late season canopy vigor maps have previously been related to PM in these systems (Taylor et al.
413 2017). This previous work did recognize that PM is highly variable (vine-to-vine) (Taylor et al.
414 2012) and that errors (differences) in co-located sensor and manual observations are to be
415 expected. The protocol of Taylor et al. (2017) for relating PM to sensor-based NDVI data did allow
416 for up to 15% of the data to be removed before modelling to improve model fits. In this study, no
417 data were removed or 'cleaned' prior to modelling, but the sample size was 10-fold larger than
418 that of Taylor et al. (2017) and it was expected that this 'noise' in the data would be accounted for
419 in the modelling. However, this does not seem to be the case. Further work is needed to better
420 understand the modelling limits here, but the clear indication is that relying only on VIs to model
421 PM will be problematic. If vineyard blocks are well-managed (i.e. maintained at a good Crop load
422 level) then the clear advice to growers would be to generate a high-quality PM map (from a
423 combination of sensor surveys and manual observations) and to use this map going forward to
424 predict PM. Subsequently years would likely only need minimal manual sampling to update and
425 correct the map.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

The results from the yield modelling clearly showed that the most effective information for understanding yield came from proximal canopy sensing in the period (1-3 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. Canopy sensing pre-bloom for use in post-bloom crop estimation has the added advantage of providing time for the data to be processed and interpreted before crop estimation is performed. The modelling showed that late-season canopy sensing or historical (year *n*-1) production data were less relevant than pre-floraison canopy information for spatial in-season yield considerations. For PM, the best way to predict it is to start measuring it. Canopy sensing at any phenological stage was not a good direct predictor of PM. Using late-season/veraison canopy vigour and targeted PM measurements for a local calibration (Taylor et al. 2017) is one way to start to obtain a spatial PM data (and to start to build a temporal history). However, growers have yet to widely adopt such an approach and more automated, grower-friendly means of vine size (PM or leaf area index) remain a priority for the industry to make routine vine size measurements.

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 2-8 % relative error across the three years (absolute errors of 0.3-1.9 Mg/ha or 0.1-0.8 tons/ac). The best PM modelling was also consistent but not as good, with 15-20 % relative error (0.08-0.14 kg/vine or 0.2-0.3 lbs/vine). Having identified preferred data types and timings of acquisitions for site-specific modelling 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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Conclusion

Sensor and manually observed 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 any crop interventions). Pre-bloom canopy vigor surveys should be used for directed crop estimations mid-season (30-days post-bloom) 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 historical, 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 and 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 pre-bloom and around veraison to have the best information for crop load management. A further conclusion was that complex site-specific processes, such as local yield development, were best described by a non-linear model, whilst local, in-season vegetative growth (PM), that is a less complex interaction, was best fitted using linear modelling approaches.

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, Ramirez-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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Barnes E, Clarke TR, Richards SE, Colaizzi P, Haberland J, Kostrzewski, M, Waller P, Choi C, Riley E and Thompson TL. 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, Bloomington, MN,1-15.

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.

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. 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany, IoT Tuscany 2018:1-4. DOI: 10.1109/IOT-TUSCANY.2018.8373046.

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. Remote Sens Agric Ecosyst Hydrol. XVI, 9239:92390S.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. OENO One, 49:127-136. DOI: 10.20870/oeno-one.2015.49.2.87.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

491 Breiman L. 2001. Random forests. Machine Learning 45:5-32. DOI:
492 10.1023/A:1010933404324.
493 Chen JM. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal
494 application. Can J Remote Sens 22:229-242. DOI: 10.1080/07038992.1996.10855178.
495 Chlingaryan A, Sukkarieh S and Whelan B. 2018. Machine learning approaches for crop
496 yield prediction and nitrogen status estimation in precision agriculture: a review. Comput
497 Electron Agr. 151:61-69. DOI: 10.1016/j.compag.2018.05.012.
498 Dash J and Curran PJ. 2004. The MERIS terrestrial chlorophyll index. Int J Remote Sens
499 25:5403-5413. DOI: 10.1080/0143116042000274015.
500 Dobrowski SZ, Ustin SL and Wolpert JA. 2003. Grapevine dormant pruning weight
501 prediction using remotely sensed data. Aust J Grape Wine Res 9:177-182. DOI: 10.1111/j.1755-
502 0238.2003.tb00267.x.
503 Drissi R, Goutouly J-P, Forget D and Gaudillere J-P. 2009. Nondestructive measurement
504 of grapevine leaf area by ground normalized difference vegetation index. Agron J 101:226-231.
505 DOI: 10.2134/agronj2007.0167.
506 Gitelson AA, Viña A, Arkebauer TJ, Rundquist DC, Keydan G and Leavitt B. 2003.
507 Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys Res
508 Letters 30:1248. DOI: 10.1029/2002GL016450.
509 Hall A, Lamb DW, Holzapfel BP and Louis JP. 2011. Within-season temporal variation
510 in correlations between vineyard canopy and winegrape composition and yield. Precis Agric
511 12:103-117. DOI: 10.1007/s11119-010-9159-4.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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. Fr Plant Sci 12:683078. DOI: 10.3389/fpls.2021.683078.

Kazmierski M, Glemas P, Rousseau J and Tisseyre B. 2011. Temporal stability of within-field patterns of ndvi in non irrigated mediterranean vineyards. J Int des Sci la Vigne du Vin 2011, 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. Fr Art Int 5:830026. DOI: 10.3389/frai.2022.830026.

Lamb D, Weedon M and Bramley R. 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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 Agr 173:105360. DOI: 10.1016/j.compag.2020.105360.

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.

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.

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

Nyéki A, Kerepesi C, Daróczy B, Benczúr A, Milics G, Nagy J, Harsányi E, Kovács AJ and Neményi M. 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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D and Brucher M, Perrot M and Duchesnay E. 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.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Richardson 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. Springer. Boston, MA. p.165-192.

Rouse JW, Haas RH Jr, 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

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

575 management in California Cabernet Sauvignon. Aust J Grape Wine Res 28:407-423. DOI:
576 10.1111/ajgw.12542.

577 Tagarakis A, Liakos V, Fountas S, Koundouras S and Gemtos TA. 2013. Management
578 zones delineation using fuzzy clustering techniques in grapevines. Precis Agric 14:18-39. DOI:
579 10.1007/s11119-012-9275-4.

580 Tardaguila J, Stoll M, Gutiérrez S, Proffitt T and Diago MP. 2021. Smart applications
581 and digital technologies in viticulture: A review. Smart Agric Technol 1:100005. DOI:
582 10.1016/j.atech.2021.100005.

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

585 Taylor JA, Link K, Taft T, Jakubowski R, Joy P, Martin M, Hoffman JS, Jankowski J and
586 Bates TR. 2017. A protocol to map vine size in commercial single high-wire trellis vineyards
587 using “off-the-shelf” proximal canopy sensing systems. Catalyst 1:35-47. DOI:
588 10.5344/catalyst.2017.16009.

589 Taylor JA, Dresser J, Hickey CC, Nuske ST and Bates TR. 2019. Considerations on
590 Spatial Crop Load Mapping. Aust J Grape Wine Res 25:144-155. DOI: 10.1111/ajgw.12378.

591 Weigle TH, Muza A, Brown B, Dunn A, Hed B, Helms M, and Loeb G. 2020. 2020 New
592 York and Pennsylvania Pest Management Guidelines for Grapes. Cornell University.

593 Xu H, Zhang X, Ye Z, Jiang L,QUI X, Tian Y, Zhu Y and Cao W. 2021. Machine
594 learning approaches can reduce environmental data requirements for regional yield potential
595 simulation. Eur J Agron 129:126335. DOI: 10.1016/j.eja.2021.126335.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Yu R, Brillante L, Torres N, 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.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

600 **Table 1** Recorded day of the year (and date) for three key phenological stages the three years of
601 the study (2019-21) at the Lake Erie Research and Extension Laboratory.

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)

602 Note : Bloom +30 is same date in July from the June date.

603

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

604 **Table 2** Vegetative indices (VIs) calculated from the three available bands of the CropCircle

605 430 canopy sensor.

Name	Abbreviation	Formula	Reference
Normalised 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
Normalised 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

606

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

607 **Table 3** Dates of canopy sensing surveys during the three years of the study translated into a
608 phenological time indication (before or after budbreak, floraison and veraison) to indicate the
609 asynchronicity of vine phenology between years. (Where DABB = Days After BudBreak, DBF =
610 Days Before Floraison (Bloom), DAF = Days After Floraison, DBV = Days Before Veraison and
611 DAV = Days After Veraison).

Date of Canopy	Timing relative to Phenology		
Surveys	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 June		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
01August	25 DBV		
03 August		18 DBV	17 DBV
11 August			9 DBV
16 August			4 DBV
21 August		Veraison	
30 August	4 DAV		
03 Sept.		13 DAV	
07 Sept.			18 DAV
14 Sept.		24 DAV	
16 Sept.			27 DAV

612

613

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

614 **Table 4** Explained Variance from cross-validation of four different specified models (using
615 different available inputs) (M1 - M4) applied to two different regression approaches (Stepwise-
616 Multivariate Linear Regression (S-MLR) and Random Forest Regression (RFR)) across three
617 years (2019-21). The models were recalibrated for each year before cross-validation using relevant
618 available variables. Best performed model in each year indicated in bold. RFR results 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*

619

620

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

621 **Table 5** Mean Average Error (MAE) (Mg/ha for yield and kg/vine) from cross-validation of four
622 different specified models (using different available inputs) (M1 – M4) applied to two different
623 regression approaches (Stepwise-Multivariate Linear Regression (S-MLR) and Random Forest
624 Regression (RFR)) across three years (2019-21). The models were recalibrated for each year using
625 the relevant available variables. Best performed model in each year indicated in bold. RFR results
626 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*

627
628

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

629 **Table 6** The key predictors and timing of data acquisition (expressed as phenological time) in
630 each year from the best performed models identified from Tables 4 and 5. For the Random Forest
631 Regression (RFR), the first five predictors are shown with the prediction power from the cross-
632 validation given in parentheses. For the Stepwise Multi-Linear Regression (S-MLR) the order
633 reflects the stepwise progression with the dominant predictor at each step given along with the
634 number of times (out of 10) it was selected in the cross-validation process. Acronyms for VIs are
635 the same as 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…

636

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

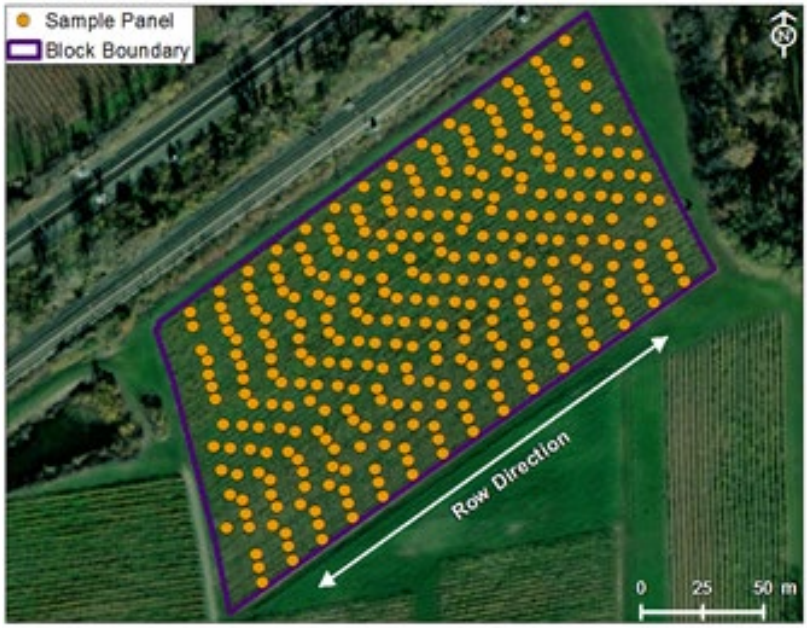
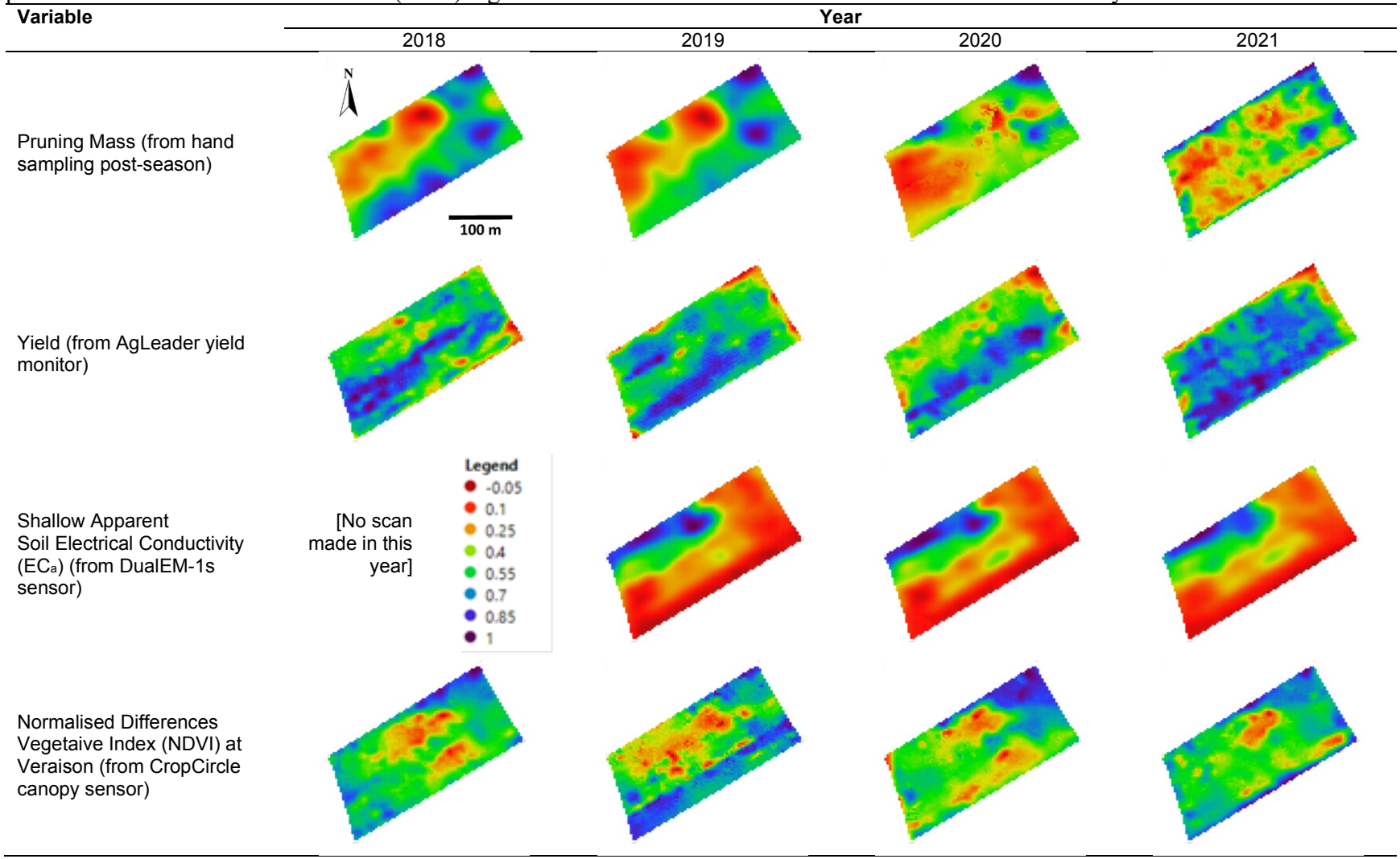


Figure 1 Location of the midpoint of the sampled panels within the 2.6 ha study block at the Cornell Lake Erie Research and Extension Laboratory, Portland, NY.

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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



American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

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 VIs are the same as Table 2. Acronyms for phenological stages are the same as for Table 3.

Predicted Variable	Model Type	Year	Order of Predictors
Yield	M1 - S-MLR	2019	Yield_2018 (6)
		2020	Yield_2019 (9); ShallowEC$_a$_2019; DeepEC$_a$_2019; PM_2019
		2021	Yield_2020 (10); PM_2020 (5)
	M1 – RFR	2019	Yield_2018 (0.2694); DeepEC$_a$_2019 (0.2579); CropLoad2018 (0.192); ShallowEC$_a$_2019 (0.1647); PM_2018 (0.1159)
		2020	ShallowEC$_a$_2020 (0.3317); Yield_2019 (0.2816); DeepEC$_a$_2020 (0.2238); CropLoad_2019 (0.097); PM_2019 (0.0659)
		2021	Yield_2020 (0.4298); DeepEC$_a$_2021 (0.1389); ShallowEC$_a$_2021 (0.1126); CropLoad_2020 (0.108)
	M2 - S-MLR	2019	MSR_3 DBF (10); SR_20 DBF (8); RECI_8 DABB (9); Various VIs at various dates
		2020	DifVI_5 DBF (10); MTCI_Veraison (9); MTCI_1 DAF (6)
		2021	DifVI_4 DBF (10); Various VIs 9 DAF (8)
	M2 – RFR	2019	DifVI_3 DBF (0.1193); SR_20 DBF (0.0738); NDVI_20 DBF (0.0598); MSR_13 DBF70619 (0.0563); MSR_20 DBF (0.0505)
		2020	DifVI_13 DBF (0.189); DifVI_5 DBF (0.1036); SR_5 DBF (0.0493); NDVI_5 DBF (0.0418); NDVI_18 DBF (0.0316)
		2021	RECI_4 DBF (0.1009); DifVI_4 DBF (0.051); NDRE_4 DBF (0.0406); RECI_11 DBF (0.0263); SR_4 DBF (0.0245)
Yield	M3 - S-MLR	2019	MSR_3 DBF (10); SR_20 DBF (8); DeepEC (7); RECI_8 DABB (5)
		2020	DifVI_5 DBF (8); MTCI_Veraison (8); DeepEC$_a$_2020 (5)
		2021	DifVI_4 DBF (10); Yield_2020 (10); Various VIs at various dates…

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Predicted Variable	Model Type	Year	Order of Predictors
		2019	DifVI_3 DBF (0.115); SR_20 DBF (0.0719); NDVI_20 DBF (0.0549); MSR_3 DBF (0.0531); SR_3 DBF0.0449
	M3 – RFR	2020	DifVI_13 DBF (0.1667); DifVI_5 DBF (0.1015); SR_5 DBF (0.0547); NDVI_5 DBF (0.0403); NDVI_18 DBF (0.0341)
		2021	RECI_4 DBF (0.0911); Yield_2020 (0.0738); DifVI_4 DBF (0.0469); NDRE_4 DBF (0.038); RECI_11 DBF (0.0245); SR_4 DBF (0.0205)
		2019	SR_29 DAF (10); NDRE_29 DAF90719 (9); Yield_2018 (9)
	M4 - S-MLR	2020	Yield_2019 (9); ShallowEC (8); DeepEC (8); MSR_090720 (8)
		2021	Yield_2020 (10); PM_2020 (7); CropLoad_2020 (2)
		2019	Yield2018 (0.149); SR_29 DAF (0.1447); DeepEC$_a$_2019 (0.1319); ShallowEC$_a$_2019 (0.1193); MSR_29 DAF (0.1154)
	M4 – RFR	2020	ShallowEC$_a$_2020 (0.2721); Yield_2019 (0.231); DeepEC$_a$_2020 (0.1647); CropLoad_2019 (0.0533); MSR_31 DAF (0.0486)
		2021	Yield_2020 (0.3894); ShallowEC$_a$_2021 (0.0877); CropLoad_2020 (0.0749); DeepEC$_a$_2021 (0.0572); MSR_22 DAF (0.0571)
		2019	PM_2018 (10)
	M1 - S-MLR	2020	PM_2019 (10); ShallowEC$_a$_2020 (10); DeepEC$_a$_2020 (10)
		2021	PM_2020 (10); CropLoad_2020 (2)
		2019	PM_2018 (0.4886); CropLoad_2018 (0.3138); Yield_2018 (0.0762); ShallowEC$_a$_2019 (0.0624); DeepEC$_a$_2019 (0.0591)
Pruning Mass	*M1 – RFR*	2020	CropLoad_2019 (0.4063); PM_2019 (0.2971); DeepEC$_a$_2020 (0.1247); ShallowEC$_a$_2020 (0.0926); Yield_2019 (0.0793)
		2021	PM_2020 (0.659); CropLoad_2020 (0.0842); Yield_2020 (0.067); DeepEC$_a$_2021 (0.053); ShallowEC$_a$_2021 (0.0479)
		2019	Various VIs at 4 DAV (8) or 25 DBV (2)
	M2 - S-MLR	2020	NDVI_13 DAV (10); Various VIs at various dates
		2021	SR/DifVI_4 DBF (9); Various VIs_160921 (7)
Pruning Mass		2019	MTCI_4 DAV (0.0494); DifVI_20 DBF (0.0485); DifVI_4 DAV (0.0442); RECI_4 DAV (0.0333); SR_4 DAV (0.0315);

American Journal of Enology and Viticulture (AJEV). doi: 10.5344/ajev.2022.22050
AJEV Papers in Press are peer-reviewed, accepted articles that have not yet been published in a print issue of the journal or edited or formatted, but may be cited by DOI. The final version may contain substantive or nonsubstantive changes.

Predicted Variable	Model Type	Year	Order of Predictors
	M2 – RFR	2020	SR_ 24 DAV (0.0591); MSR_13 DAV (0.0413); SR_18 DBF (0.0292); SR_13 DAV (0.0261); NDVI_13 DAV (0.0257)
		2021	MTCI_17 DAF (0.0286); SR_24 DBV (0.0211); SR_17 DAF (0.0206); MSR_17 DAF (0.0186); MTCI_24 DBV (0.0182)
	M3 - S-MLR	2019	PM_2018 (10)
		2020	PM_2019 (10); RECI_5 DBF (10); MSR_13 DAV (6)
		2021	PM_2020 (10); Various VIs at various dates…
	M3 – RFR	2019	PM_2018 (0.3693); CropLoad_2018 (0.2813); DifVI_20 DBF (0.0206); MTCI_4 DAV (0.0119); RECI_20 DBF (0.0105)
		2020	CropLoad_2019 (0.2431); PM_2019 (0.2354); SR_24 DAV (0.022); MSR_13 DAV (0.0174); MTCI_12 DAF (0.0126)
		2021	PM_2020 (0.2805); CropLoad_2020 (0.137); MTCI_17 DAF (0.0141); Yield_2020 (0.0125); MSR_17 DAF (0.0105); SR_24 DBV (0.0103)
	M4 - S-MLR	2019	PM_2018 (10)
		2020	PM_2019 (10); ShallowEC$_a$_2020 (10); SR_090720 (8)
		2021	PM_2020 (10); CropLoad_2020 (2)
	M4 – RFR	2019	PM_2018 (0.4658); CropLoad_2018 (0.2837); Yield_2018 (0.0411); SR_29 DAF (0.0329); ShallowEC$_a$_2019 (0.0293)
		2020	CropLoad_2019 (0.3704); PM_2019 (0.2813); DeepEC$_a$_2020 (0.0734); ShallowEC$_a$_2020 (0.0483); Yield_2019 (0.0461)
		2021	PM_2020 (0.6398); CropLoad_2020 (0.058); Yield_2020 (0.0415); DeepEC$_a$_2021 (0.0366); ShallowEC$_a$_2021 (0.0306)