Highlights

An exploratory analysis of the influence of resource use on the yield versus quality trade-off in Australian vineyards

Author

- Comparative analysis of resource use, quality and quantity in Australian winegrowing.
- Regional comparison of outcomes and resource use in Australian winegrowing regions.
- Baseline models for comparing wine crops.
- Analysis of national, decade long data source.

An exploratory analysis of the influence of resource use on the yield versus quality trade-off in Australian vineyards

Author[1,1,1]

Abstract

The global focus on sustainability in agronomic industries has changed the way in which these enterprises do business. When strategies for a sustainable winegrowing industry are assessed, there is a trade-off between balancing the amount of resources invested and the resultant yield verses quality produced. In this analysis we observe relationships between yield and quality through the use of linear models. An extensive amount of research into a variety of factors' effect on grape quality and yield exists; but due to the lack of long-term and in-depth data, individual effects are often studied in isolation. The dataset used for this analysis includes data collected for the past 10 years from a multitude of vineyards located over a diverse range of Australian winegrowing regions. We aim to use this broad dataset to describe the relationship of input resources to the output yield and quality of vineyards. There was an expected strong relationship between size and resource use, with the overall space of a vineyard and its access to resources greatly determining the upper limit of potential yield. However, size was also negatively related to the potential quality, with higher quality being connected to high resource inputs per area; rather than to the overall expenditure of resources.

Vineyard outputs were also augmented by regional and yearly affects. It is important to also consider a vineyard's business goal, region, external pressures and economies of scale. With regional constraints also contributing to deciding the best strategies to pursue when considering quality or quantity.

1. Introduction

The global focus on sustainability in agronomic industries has changed the way in which these enterprises do business. When strategies for a sustainable winegrowing industry are assessed, there is a trade-off between balancing the amount of resources invested and the resultant yield verses quality produced. This dilemma exists across agriculture through shared fundamental considerations such as water use and nitrogen levels (Hemming et al., 2020; Kawasaki and Uchida, 2016; ZHU et al., 2017). Quality in viticulture (the cultivation of grapes for wine production) is driven through its integration within the wine industry; with a wine's potential quality being initially defined through the chemical makeup of the grapes used in its production. The consideration of sustainability within viticulture is further complicated by environmental and socio-demographic pressures. In the Australian context, these include: biosecurity, climate and international market demands.

In this analysis we observe relationships between yield and quality through the use of linear models. An extensive amount of research into a variety of factors' effect on grape quality and yield exists; but due to the lack of long-term and in-depth data, individual effects are often studied in isolation (Abbal et al., 2016). The lack of consolidated datasets also restricts the ability to gain statistical insights at large scales and across multiple regions

(Keith Jones, 2002; Knight et al., 2019). The dataset used for this analysis includes data collected for the past 10 years from a multitude of vineyards located over a diverse range of Australian winegrowing regions.

We aim to use this broad dataset to describe the relationship of input resources to the output yield and quality of vineyards. The practical addition of this aim is a baseline for comparison - given a vineyard within Australia, one could extrapolate their comparative efficiency with regard to the tradeoff between invested resources, yield and quality. In achieving this we will also confirm the existence of a yield versus quality trade off within Australian winegrowing; one not prior confirmed explicitly across such varying regions, scales and climates.

2. Methods

We created four linear models to explore relationships between resource-use and vineyard outputs (see Table1). The data was sourced from Sustainable Winegrowing Australia and Wine Australia. Variables used included: yield, average sale price, region, water use, emissions, area harvested and year. After fitting to the data, each model was validated using k-fold cross validation.

2.1. Analysis

Before models were fit to the data, Pearson Correlation Coefficients were used to look at the existence of linear relationships between predictor variables. These relationships were summarised in correlation matrices to compare the level of interaction present between predictor variables. The relationships between the predictors and response variables were then modelled

Table 1: Summary of models; their predictors, covariates and variable interactions.

	Response	Predictors	Covariates	Interactions
Model 1	Yield	Water Used Scope 1 Emissions	Area Harvested Year GI Region	N/A
Model 2	$\frac{\text{Yield}}{\text{Area Harvested}}$	Water Used Scope 1 Emissions	Area Harvested Year GI Region	Area Harvested * Scope 1 Emissions Area Harvested * Water Use Year * Region
Model 3	Yield×Average Sale Price	Water Used Scope 1 Emissions	Area Harvested Year GI Region	N/A
Model 4	$\frac{\text{Yield}\times\text{Average Sale Price}}{\text{Area Harvested}}$	Water Used Scope 1 Emissions	Area Harvested Year GI Region	Area Harvested * Scope 1 Emissions Area Harvested * Water Use Year * Region

using General Linear Models. Both the Pearson Correlation Coefficients and General Linear Models were created using the R statistical programming language (R Core Team, 2021). General Linear Models were chosen as they offer the ability to produce statistical models that are explicit in the relationships between predictors and response variables. General Linear Models also allow the exploration of interactions between predictors and present easily comparable differences in the influence and magnitude of relationships. A variety of alternate methods were also explored, including: Splines, hierarchical regression, General Additive Models, and Generalised Linear Models. These alternative approaches were not used as final models due to offering no further insights or improvements in accuracy.

The response variables of the models were yield and quality. Yield was defined as the total tonnes of grapes harvested. For the purpose of this study, quality was defined by the financial value of winegrape crops' average sale price per tonne. The definition of quality was an important consideration, as quality can be defined in a variety of ways, for example analysing grapes': aroma, chemical composition and color. Using sale price as a defining trait of quality was due to the market value of winegrapes being reliant on grape quality and because Wine Australia explicitly defines grape quality through the use of discrete price brackets in their annual reports ; the generalisation made to reflect quality through using average price assumed a due diligence of those who purchased the grapes (Yegge, 2001). Both response variables were examined as totals and as scales of area harvested. Values were compared in this manner to observe how economies of scale affect the use of resources.

2.2. Significant Tests

2.3. Data

Data used in this analysis was sampled by Sustainable Winegrowing Australia and Wine Australia. Sustainable Winegrowing Australia is Australia's national wine industry sustainability program, which aims to facilitate grapegrowers and winemakers in demonstrating and improving their sustainability (SWA, 2022). Wine Australia is an Australian Government statutory authority governed by the Wine Australia Act 2013 (Win, 2019).

Data sampled by Wine Australia was collected via phone surveys and included: summary statistics such as yield and average price of sale per tonne; these values were summarised by region and grape varietal. Data recorded by Sustainable Winegrowing Australia was entered manually by winegrowers using a web based interface with some fields being optional, variables included: region, harvest year, yield, area harvested, water used and fuel used (diesel, petrol, biodiesel and LPG). To enable direct comparisons between fuels, they were converted to tonnes of Carbon Dioxide equivalent.

The inclusion of Wine Australia data was due to average sale price being an optional field in Sustainable Winegrowing Australia's dataset. Regional average prices from Wine Australia were filled into values that were missing from the Sustainable Winegrowing Australia data; the common practice of purchasing grapes at regional prices was an important consideration in this decision. Two subsets of data were then created for the analysis. The first subset contained all vineyards and was used for Models 1 and 3. The second subset contained vineyards which either recorded a value for average price of sale per tonne through Sustainable Winegrowing Australia, or were within a region with an average price of sale recorded by Wine Australia; this subset was used for Models 2 and 4. These subsets meant that the data would be limited to samples which had recorded values for the response variables (see Table1), where every sample had a recorded value for yield but not average price of sale per tonne.

The first subset of data was used for Model 1 and Model 2 (see Table1). This subset contained 5298 samples spanning the period from 2012 to 2022, covering 55 GI Regions and 1261 separate vineyards.

The second subset of data, was limited to vineyards that recorded a value for their average sale price of grapes per tonne. This subset was used for Model 3 and Model 4 (see Table1); and contained 2878 samples spanning the period from 2015 to 2022, covering 51 GI Regions and 944 separate vineyards. 1842 of the values for average price of sale per tonne were extracted from Wine Australia surveys with the remaining 1036 being from Sustainable Winegrowing Australia's dataset.

Additional variables were considered for analysis but were excluded due to being either underreported or had insignificant contributions to model accuracies. Variables explored but not used due to low reporting values included: fertiliser, and scope 2 emissions. Variables considered but ultimately removed due to a lack of significant contributions to models, included: the use of renewable energy, contractor use, and pressures such as frost, fire and disease. Data preprocessing was conducted prior to analysis using the Python programming language (G. van Rossum, 1995). Preprocessing included logarithmic transformations, centring and scaling by standard deviation. Variables such as scope 1 emissions, which required prior calculations were also computed using Python.

2.4. Total Emissions

The equation given from the Australian National Greenhouse Accounts Factors, shown as

$$tCO_2e = \frac{Q \times EC \times EF1 + EF3}{1000}, \tag{1}$$

was used to convert the quantity of fuel in litres, Q, using a prescribed Energy Content, EC, and emission factors of scope one, $EF1$, and scope three, $EF3$, to tonnes of Carbon Dioxide Emission equivalent, $tCO2e$ (Department of Climate Change, Energy, the Environment and Water, 2022). Emissions were calculated for total diesel, petrol, bio-diesel and LPG used.

2.5. Region

Differences in vineyard locations were captured through the use of Geographical Indicator Regions (GI Regions). Each GI Region has its own unique mixture of climatic and geophysical properties that describes a unique winegrowing region within Australia; these regions were predefined by Wine Australia (Halliday, 2009; Oliver et al., 2013; SOAR et al., 2008). Both Wine Australia and Sustainable Winegrowing Australia used the same GI Region format to describe location.

The site of a vineyard predetermines several physical parameters such as climate, geology and soil; making location a widely considered key determinant of grape yield and quality (Abbal et al., 2016; Agosta et al., 2012; Fraga et al., 2017). The climatic properties of each GI Region were summarised by using predefined classifications as per the Sustainable Winegrowing Australia (2021) user manual. The user manual describes climates by rainfall and temperature, creating supersets of Regions of similar climatic properties. The climatic groups were used to illustrate similarities and differences occurring in areas larger than GI Regions.

2.6. Model Validation

Models were validated using K-fold cross validation calculated through the R Caret Package (Kuhn, 2008). K-fold cross validation works by removing a subset of data from the sample used to train models and then predicts those variables to determine how sensitive the model is to changes in the sample data. For this analysis each model was validated using 10 folds, repeated 100 times.

Table 2: Summary statistics of each continuous variable.

Variable	Mean	Standard Deviation	Minimum	Maximum
Yield	7.757E+02	2.179E+03	1.000E+00	7.231E+04
Area Harvested	6.670E+05	1.337E+06	7.000E+02	2.436E+07
Water Used	7.471E+06	5.646E+08	1.000E+00	4.268E+10
Scope One Emissions	4.173E+04	8.571E+04	6.755E+00	2.110E+06
$\frac{\textbf{Yield}}{\textbf{Area}}$	1.009E+01	8.127E+00	4.000E-02	8.634E+01
Average Sale Price	1.477E+03	9.216E+02	1.600E+02	2.600E+04
$\frac{\textbf{Average Sale Price}}{\textbf{Area Harvested}}$	1.347E+02	5.711E+02	1.753E-01	2.979E+04

3. Results

3.1. Data

Each variable was logarithmically transformed and then centred around a mean of 0. The values of these variables were then divided by standard deviation creating a comparable ratio intrinsic to each variable. Table 2 shows the summary statistics of each variable, to contextualise these ratios to real values.

Table 3: Variable Pearson correlation values for logarithmically transformed values.

Variable	Yield	Area Harvested	Water Used	Scope One Emissions	$\frac{\text{Yield}}{\text{Area}}$	Average Sale Price	$\frac{\text{Average Sale Price}}{\text{Area Harvested}}$
Yield	1.00E+00	7.44E-01	-4.31E-03	7.29E-01	3.50E-01	-2.26E-01	-1.64E-01
Area Harvested	7.44E-01	1.00E+00	-5.33E-03	8.92E-01	7.85E-02	-1.18E-01	-2.04E-01
Water Used	-4.31E-03	-5.33E-03	1.00E+00	-1.93E-03	-5.60E-03	-3.56E-02	-2.67E-02
Scope One Emissions	7.29E-01	8.92E-01	-1.93E-03	1.00E+00	9.36E-02	-9.42E-02	-1.93E-01
$\frac{\text{Yield}}{\text{Area}}$	3.50E-01	7.85E-02	-5.60E-03	9.36E-02	1.00E+00	-4.85E-01	-1.70E-01
Average Sale Price	-2.26E-01	-1.18E-01	-3.56E-02	-9.42E-02	-4.85E-01	1.00E+00	4.73E-01
$\frac{\text{Average Sale Price}}{\text{Area Harvested}}$	-1.64E-01	-2.04E-01	-2.67E-02	-1.93E-01	-1.70E-01	4.73E-01	1.00E+00

3.2. Exploratory Analysis

Linear relationships between variables were explored using Pearson Correlation Coefficients. Values for these coefficients reflect the linear relation between two variables, on a scale between -1 and 1; the magnitude and sign of a coefficient indicates the strength of the relation, and whether the relation is positive or negative respectively. This was undertaken for data on the original scale and for data as a logarithmic transform. The logarithmic transformed data showed the strongest correlations, likely due to a skew caused by a greater number of smaller vineyards within the dataset (see Table 3). Transforming data prior to calculating the coefficients changes several things: The logarithmic transform of the data alters the interpretation of the coefficients to percentage change - a coefficient will be indicative of the change in percentage of one variable compared to the other; scaling by standard deviation also changes this interpretation to be a percentage of that variables standard deviation. Scaling by standard deviation also makes the Pearson Correlation Coefficient equal to the covariance of the two variables. With all this in mind, when considering the logarithmically transformed variables, a coefficient of 1 would indicate that: given the change of one variable by one percentage of its standard deviation, the other variable would change by one percent of its own standard deviation. The importance of this is the dimensionless nature of these relationships and that it can be translated directly to any vineyard's case that has a well known distribution.

To determine if a coefficient was indicative of a strong relationship, confidence intervals were used. P-values reflected the significance of a given correlation coefficient when considering its relation to sample size via its incorporation as an element of standard error. Strong relationships were found to be present as all P-values, except for the non-transformed values for water used, were considered significant ($P < 2.200E\text{-}16$).

3.3. General Linear Models

General Linear Models were used to describe how response variables related to predictors' values. Log transformed variables were used as inputs to these models as they resulted in higher R^2 values and described the relationships proportionally; reflecting coefficient values as percentages of a variable's standard deviation. Each model showed a strong relationship between the predictors and the response (see Table 4). Model accuracy was measured in R^2, as this allowed an easy comparison between their performances and their validation.

3.3.1. F-tests

To determine if predictors significantly related to a Model's response variable, F-tests were conducted. Aside from 3 variables, all F-tests across each model indicated a significant contribution at 95% confidence. The three exceptions were: scope 1 emissions in Model 3 (P=2.221E-01) and Model 4 (P=3.621E-01), and Model 2's interaction between area harvested and water

Table 4: Summary of models; their performance, F-statistics and Residual error.

	R^2	Adjusted R^2	F-Statistic	P-Value	Residual Standard Error	Residual Sum of Squares	Residual Mean of Squares
Model 1 Yield	9.072E-01	9.061E-01	7.753E+02	2.200e-16	3.065E-01	4.913E+02	1.000E-01
Model 2 Yield/Area	7.951E-01	7.770E-01	4.403E+01	2.200e-16	4.722E-01	1.085E+03	2.200E-01
Model 3 Value	9.753E-01	9.748E-01	1.885E+03	2.200e-16	1.589E-01	7.111E+01	3.000E-02
Model 4 Value / Area	9.669E-01	9.638E-01	3.095E+02	2.200e-16	1.904E-01	9.528E+01	4.000E-02

used (P=2.192E-01).

Scope 1 emissions was included in all models to directly compare the response variables as ratios of vineyard size to raw values. Even though not significant within models 3 and 4, when using the Pearson Correlation Coefficients scope 1 emissions was strongly correlated to every Model's response variable; this was especially so for Model 1 and 4 (Yield and average price per tonne as a ratio to area harvested, respectively).

3.3.2. T-tests

T-tests were used to determine if predictors significantly contributed to their models when accounting for other variables; this allowed a more granular examination of interactions and factors within categorical variables, showing which specific years and areas contributed significantly and which did not (the appendix contains a comprehensive list of these values).

For Models 1 (yield) and 3 (value) year played a pivotal role, with only one year in each model not being significant (2021/2022 and 2016/2017 respectively). Both Model 1 and 3 showed a majority of regions were significant with 32 of 54 regions being significant in Model 1, and 42 of 50 regions being significant in Model 3 at 95% confidence.

The number of combinations of year and region meant that Models 2 and 4 had many tests (424 and 243 respectively). Model 2 found 62.56% of these combinations were indicative of a significant contribution to the model at 95% significance. Model 4 was found to have 88.07% of its year/region combinations indicating a significant contribution. A likely reason for some combinations not being significant was a lack of samples in that particular region/year being present; with region sample sizes ranging from 1 to 1006. With regard to continuous variables: Model 1 and 2 showed all variables to be significant at 95% confidence when accounting for other variables. T-tests for Model 3 showed all continuous variables except scope 1 emissions were significant. Model 4 showed all variables aside from scope 1 emissions and water use to be significant; with scope 1 emissions and water use only being significant when considered as an interaction with area harvested but not when considered on their own.

3.3.3. Model Coefficients

The coefficients of each model describe the relationship of a predictor variable to its response when considering all other variables. Due to the transformations of the data, coefficients are individually interpreted in the same manner as the prior regression values were (see Section 3.2); unlike the regression values, coefficient ranges are not limited between -1 and 1.

We look at the coefficients of categorical and continuous variables separately. This is done as the categorical variables have many coefficients, one

Table 5: Summary of each Models coefficients for continuous variables

	Intercept	**Area Harvested**	**Water Used**	**Scope 1 Emissions**	**Area Harvested * Scope 1 Emissions**	**Area Harvested * Water Used**
Model 1	-3.318E-02	7.418E-01	8.660E-02	6.731E-02		
Model 2	-6.516E-01	5.774E-01	1.079E-02	8.498E-02	-4.971E-02	-5.346E-02
Model 3	1.808E-02	9.713E-01	-2.310E-02	-6.992E-03		
Model 4	6.702E-01	-7.354E-01	-6.732E-03	-5.645E-03	2.726E-02	7.515E-02

for each category, whilst continuous variables have only one. The coefficient for categorical variables is summarised in Figure 1; illustrating the difference in the range as well as affect region and year could have on each model. Comparatively, the continuous variables coefficients are summarised in Table 5. In terms of magnitude, GI region has the highest possible absolute value for each model. An important consideration is that region and year are binary, such that they are only equal to zero or the coefficient (as they will present as a value of 1 which will be multiplied by the coefficient); this means that, although region may have a strong relationship, it can be overshadowed by an extreme value of one of the continuous variables. The most notable difference between the continuous variables coefficients is the change from positive to negative values. This change occurs between the Models for Yield (Model 1 and 2) and the Models for value (Models 3 and 4); where all but the coefficient for area harvested had the opposite sign (see Table 5). These models also differ in an order of magnitude when looking at resource use, with the coefficients for yield being smaller than those for value.

3.3.4. Model comparisons: yield versus value

Directly comparing response variables, how crop value changes with yield, also allows an indirect comparison between the response variables and resource use. We do this through using known relationships of response variables to their predictors. These relationships are described by the coefficients. Resource use is described by the predictor variables (through water used and scope 1 emissions), because of this we can observe the response variables somewhat interchangeably with the predictors - although caution should be taken to view them sceptically and alongside the influence of their coefficients. As the predictors are known to have a strong positive correlation with each other, they will tend toward increasing and decreasing together (but not at the same rates). It is also important to consider the interactions of predictor variables when comparing the response variables that are ratios of area. Furthermore, these comparisons require the consideration of the covariates, in this case: area harvested, year and region.

Observing Figure 2 shows an almost discrete difference between vineyards in 'Hot' areas than other regions. Comparing Figure 2 to Figure 3 shows almost opposing trends. However, with area coming into play in Figure 3, many data points are scaled differently; specifically the vineyards from 'Hot' regions change to be found the bottom right tail end, indicating the production of large quantity of lower value grapes. An inconspicuous difference between the Figures, is that a large amount of the difference can be explained by rotation (being 90° clockwise from Figure 2 to 3). This is more visible when comparing both graphs to the map of regional averages for response

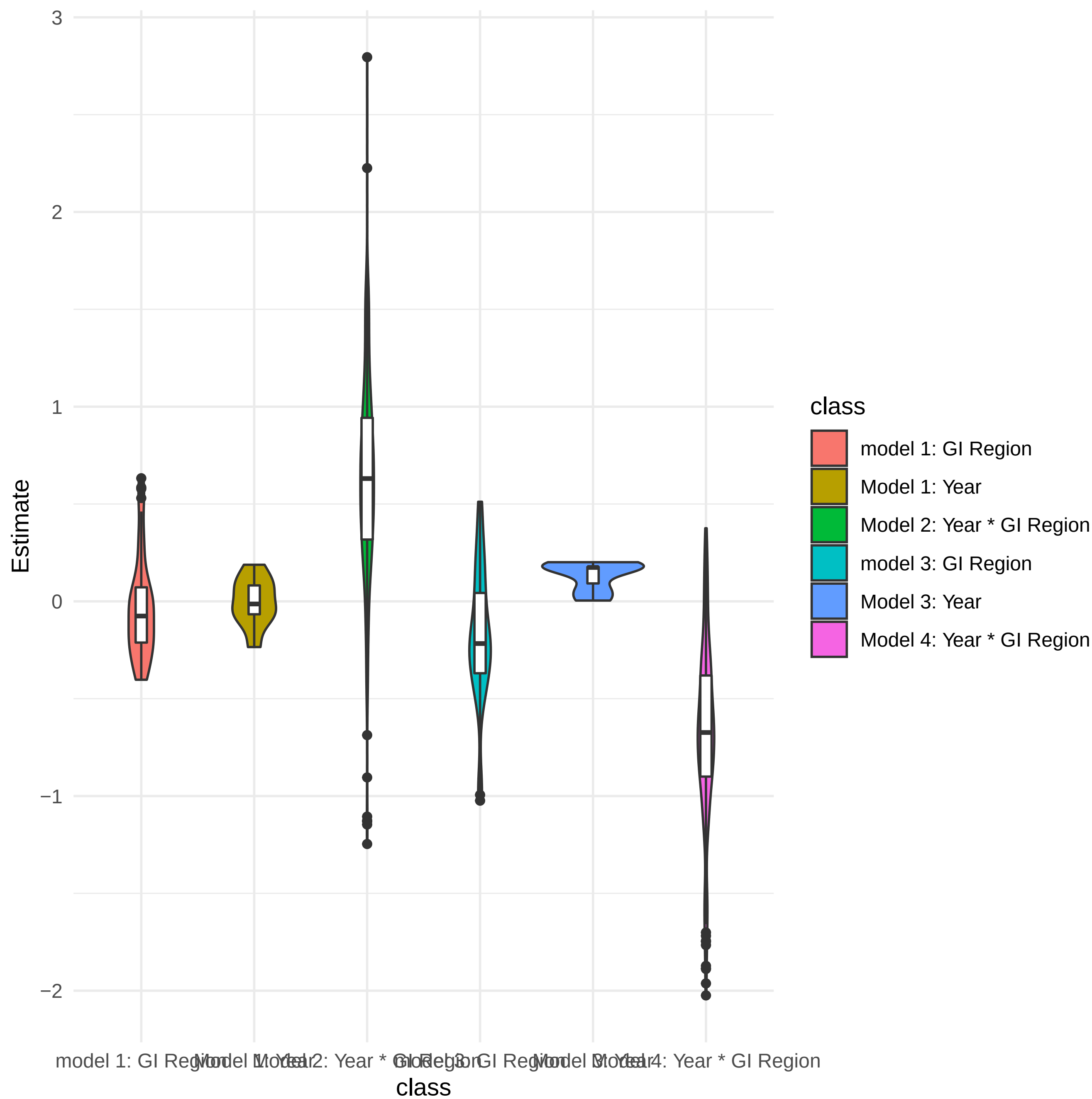

Figure 1: Violin plots of GI Region and Year coefficients for each model.

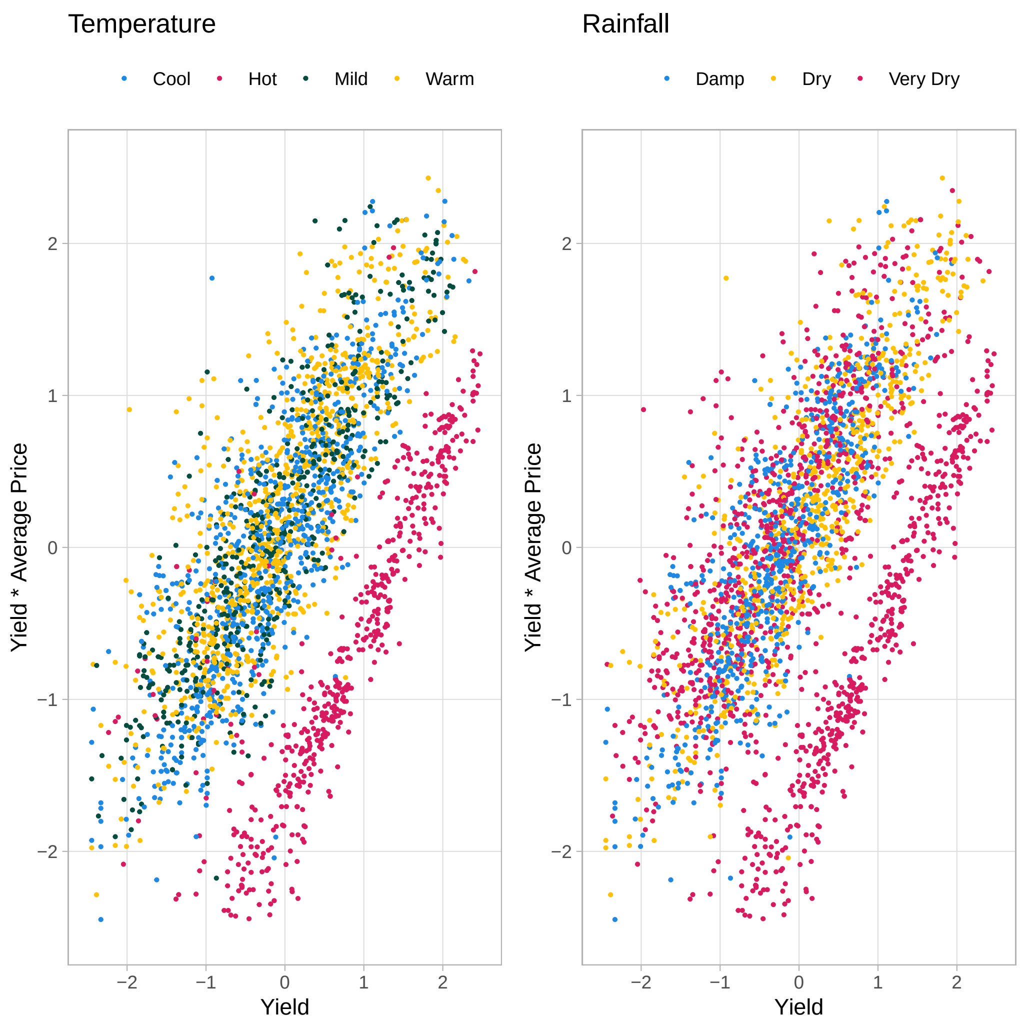

Figure 2: Scatter plot of vineyard yield against the product of yield and average price per tonne. The axes are in standard deviations with points coloured by climate.

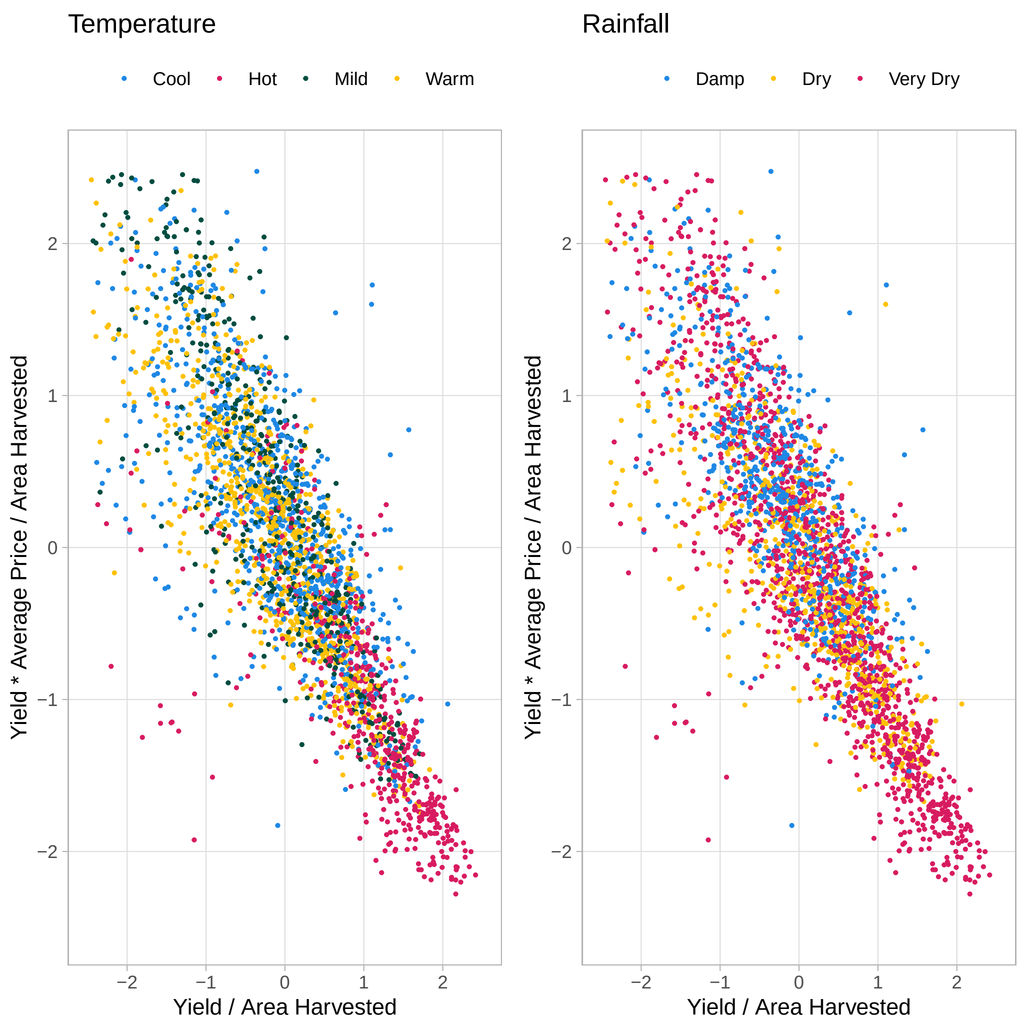

Figure 3: Scatter plot of vineyard yield against the product of yield and average price per tonne as ratios to area harvested. The axes are in standard deviations with points coloured by climate.

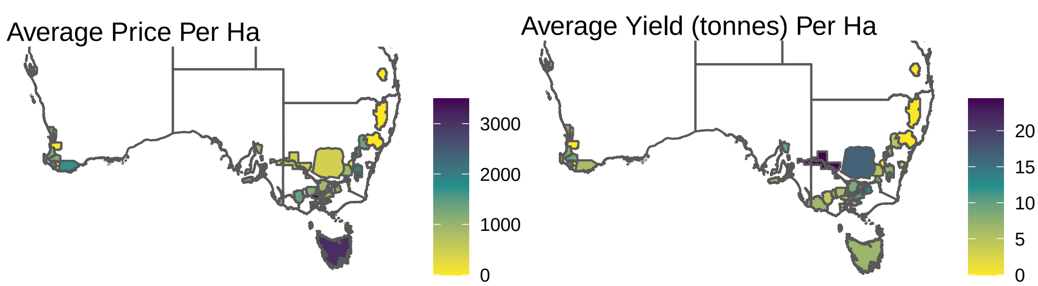

Figure 4: Map of regional average yield and value per hectare.

Table 6: Model validation using k-fold cross validation, for 10 folds repeated 100 times.

	Residual Mean Squared Error	**R2**	**Mean Average Error**
Model 1	3.087E-01	9.045E-01	2.165E-01
Model 2	5.104E-01	7.409E-01	3.493E-01
Model 3	1.652E-01	9.723E-01	1.008E-01
Model 4	2.235E-01	9.500E-01	1.279E-01

variables, see Figure 4. There is a notable change between regional averages when looking at yield versus value. Through the coefficients we can deduce that: this difference is also a difference between more resources used for the raw response variables; and a difference between overall resource use and the size of the vineyard when considering the response variables as a ratio to area. Noting, resource use and area harvested have a combined relationship through their interactions, and separate relationships as individual variables (see Table 5). A notable occurrence in Figure 3, is that the 'Very Dry' vineyards which produce lower yields and higher quality grapes are predominantly found in the Barossa Valley (a wine region known for its high quality Shiraz). This note is important as it shows climate is not exclusively the consideration, soil and other geographical phenomenon have considerable impacts on vineyard outcomes.

3.4. Model Validation

To validate the performance of these models k-fold cross validation was used. This was done using 10 folds, $k = 10$, repeated 100 times. The models performed similarly to their original counterparts (see Table 3.4).

4. Discussion

There was an expected strong relationship between size and resource use, with the overall space of a vineyard and its access to resources greatly determining the upper limit of potential yield. However, size was also inversely related to the potential quality, with higher quality being related to high resource inputs per area; rather than to the overall expenditure of resources. Vineyard outputs were also augmented by regional and yearly affects. Even given regional and yearly changes, there was a strong connection between smaller vineyards and higher quality. This could have been due to the easier management of smaller properties.

4.1. Resource use and yield versus quality

There are many on-the-ground decisions that influence both quality and yield. Comparing the R^2 values between Models 2 and 4 showed that the average price per tonne of grapes described a great deal of the relationship between resource use and yield when variables were considered as ratios of area (due to the discrepancy in R^2 between the two models, see Table 4). This descrepency is likely due to different vineyard prioritisation, which can be described by the type of quality and quantity a vineyard aims to target. Decisions such as the prioritisation of quality over quantity, are governed by complex physical and social forces, for example: international market demands, disease pressures and natural disasters (Abad et al., 2021; Cortez et al., 2009; Hall et al., 2011; I. Goodwin, et al., 2009; Kasimati et al., 2022; Oliver et al., 2013; Srivastava and Sadistap, 2018); with many of these occurrences being highlighted throughout the past decades vintage reports from Wine Australia (Wine Australia, 2019, 2021, 2022; Winemakers' Federation of Australia, 2013, 2014, 2015, 2016, 2017, 2018). It is also important to consider that these reports show that the warm inland regions have seen a decline in profit during this period, whereas regions targetting quality did not. Size becomes an important consideration, as it dictates the potential capacity to produce greater volumes of grapes. However, given the comparison of value per area, regions with larger vineyards (such as warmer in land regions) and larger vineyards in general, tend to underperform. When considering the 'Hot Very Dry' vineyards (see Figure 3) These vineyards would be very competitive with only a minor increase to sale price, possibly outperforming other regions.

The negative trend between size and average sales price could be a side effect of supply versus demand, especially when looking at the level of difference in production of some vineyards. Economies of scale likely played a role in determining yield but were only one consideration alongside resource use. Size was also less of a determining factor when considering quality. It is possible that the relationship of scope 1 emissions between yield and quality was closely tied to a vineyard's area; due to requiring more fuel to cover issues (such as fixing a broken irrigation pipe), where a larger area has the potential for issues to be further away. This is further cemented when noting that most irrigation systems are diesel based, with water use being a significant variable in each model and scope 1 emissions not; scope one emissions' lack of significance and contribution given its F-statistics, could be indicative that other vineyard activities requiring fuel are not as determining factors for a vineyards grape quality. The relationship between yield, value and area was not simply about efficiently producing the most grapes; sales price and by association grape quality, are integral to the profitability, and this is strongly linked to resource-use and thus the longevity and sustainability of a vineyard. There are important considerations unique to winegrowing compared to other agricultural industries. The vertical integration of winegrowing within the wine industry ties winegrowers to secondary and tertiary industries, such as wine production, packaging, transport and sales. This results in unique issues and considerations for each vineyard, where on-the-ground decisions are influenced by other wine industry's choices, such as the use of sustainable practices in vineyards as a requirement for sale in overseas markets; notably these interactions can be further complicated by some winegrowers being completely integrated into a wine company, while others are not (Knight et al., 2019). Incorporating decisions into the model could help describe the contributing factors to regional differences beyond resource consumption and regional differences but would require incredibly granular data and more sophisticated modelling.

4.2. Regional Differences

Some regions appeared to produce many low quality grapes at scale whilst others focussed on producing higher quality grapes in lower volumes. This behaviour can also be observed when reviewing Wine Australia's annual reports, where it is apparent that some GI regions, such as the Riverland, are known for producing large amounts of lower grade (low value per tonne) grapes Wine Australia (2022); Winemakers' Federation of Australia (2017). Comparatively other regions, such as Tasmania, only produce high quality grapes but in smaller quantities. The difference in pricing per tonne between the lowest and highest graded grapes can be greater than a hundred times the difference in value per tonne. Not all regions target only one grade of grape, with some producing a variety of differently graded grapes; such as the Yarra Valley, which produces grades from C to A.

Some regions are known for their quality and may have a bias in purchasers or bring greater demand regardless of similarities and differences in production of quality of grapes (Halliday, 2009). This effect could stifle the potential for market opportunities within lesser known regions. A further possibility is the existence of regional upper limits on potential quality, or that there are diminishing returns in some regions when pursuing quality or quantity; however these types of relationships may be obfuscated by knowledgeable winegrowers who avoid this pitfall.

Due to regional differences, different strategies are likely employed across different regions; such as some regions targeting mass production over quality. This is most notable when grouping regions by climate, especially when considering GI Regions in the 'Hot Very Dry' climate (see Figure 2). In alternative attempts at models it was found that without the direct incorporation of GI Region or year, predictions greatly under performed. The effect of climate in the models was never as significant as the more granular GI regions, and always led to less accurate models. Although not chosen over

GI region, climate was considered to be a large determinant of the ability to produce larger quantities of grapes, as well as a determinant in grape quality (Agosta et al., 2012). The more granular GI Region likely explained a broader mix of geographical phenomenon, such as soil, geology and access to water resources (Abbal et al., 2016; Carmona et al., 2011). The interaction between year and GI Region likely accounted for events such as bushfires, which would be impactful, but only at a local level, both in time and space.

4.3. Limitations

Limitations included overestimating yield for models 1 and 2, and underestimating crop value in models 3 and 4 (see appendix). The issue of model 1 and 2 over predicting yield, may have been due to preventative measures brought on by regional pressures such as fire, frost and disease. Where, more resources were required to prevent these issues from spreading within a region, thus disproportionately effecting some vineyards compared to others locally. This type of maintenance is not well captured especially when considering that some regions, especially those in warmer areas, are not as prone to disease as cooler climates and could potentially have lower operating costs per hectare. This could create a discrepancy in vineyards that utilised preventative measures in wetter regions, as opposed to those that did not, thus expending less fuel and energy but risking disease. When reviewing the differences between regions it is important to consider that vineyards in 'Hot Very Dry' areas can be hundreds of times the size of those in other regions. This limitation could be overcome by incorporating the profitability of vineyards, compare the financial success of working at different operational scales.

Variables such as the utilisation of renewable energy, contractors, and the occurrence of disease, fire and frost were originally explored to capture the discrepancies between similar vineyards that produced different yields and crop values. However, none of these variables were significantly connected to the response variables, and did not add to model accuracy; even when considered as interactions. The use of other methods, specifically splines, resulted in more normally distributed residuals but at a drastically reduced overall accuracy when comparing R^2 and Residual Square Error. Attempts to fully explain small variations was always overshadowed by the dramatic differences in regional trends.

Having more data for each region would also be an improvement, allowing greater comparison between regions. More variables may also help to discern vineyards that can produce larger volumes of grapes at higher prices. The use of semi transparent tools such as random forests and decision trees alongside more variables and data may help to uncover the reasons for values that were under or overestimated. These differences could be caused by the use of alternative sustainable practices in the field. And, while there is evidence to suggest that environmentally sustainable practices can reduce costs, increase efficiency, whilst improving the quality of grapes; more research is needed to link these benefits across different regions and climates (Baiano, 2021; Mariani and Vastola, 2015; Montalvo-Falcón et al., 2023).

5. Conclusion

In summary, vineyard yield and crop value is well-defined by the resources used. However, it is important to consider a vineyard's business goal, region, external pressures and economies of scale. Where, larger vineyards are likely to produce greater overall yields, and have higher yield per area. Smaller vineyards are likely to produce more value per area, and a higher quality of grape. It is likely that regional constraints also contribute to the best strategy to pursue when considering quality or quantity.

References

, 2019. Wine Australia Act 2013.

Abad, J., Hermoso de Mendoza, I., Marín, D., Orcaray, L., Santesteban, L.G., 2021. Cover crops in viticulture. A systematic review (1):
Implications on soil characteristics and biodiversity in vineyard. OENO One 55, 295–312. doi:10.20870/oeno-one.2021.55.1.3599.

Abbal, P., Sablayrolles, J.M., Matzner-Lober, É., Boursiquot, J.M., Baudrit, C., Carbonneau, A., 2016. Decision Support System for Vine Growers Based on a Bayesian Network. Journal of agricultural, biological, and environmental statistics 21, 131–151. doi:10.1007/s13253-015-0233-2.

Agosta, E., Canziani, P., Cavagnaro, M., 2012. Regional climate variability impacts on the annual grape yield in Mendoza, Argentina. Journal of Applied Meteorology and Climatology 51, 993–1009.

Baiano, A., 2021. An Overview on Sustainability in the Wine Production Chain. Beverages 7. doi:10.3390/beverages7010015.

Carmona, G., Varela-Ortega, C., Bromley, J., 2011. The Use of Participatory Object-Oriented Bayesian Networks and Agro-Economic Models for

Groundwater Management in Spain. Water resources management 25, 1509–1524. doi:10.1007/s11269-010-9757-y.

Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Using data mining for wine quality assessment, in: Discovery Science: 12th International Conference, DS 2009, Porto, Portugal, October 3-5, 2009 12, Springer. pp. 66–79.

Department of Climate Change, Energy, the Environment and Water, 2022. Australian National Greenhouse Accounts Factors.

Fraga, H., Costa, R., Santos, J.A., 2017. Multivariate clustering of viticultural terroirs in the Douro winemaking region. Ciência Téc. Vitiv. 32, 142–153.

G. van Rossum, 1995. Python tutorial, Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI),.

Hall, A., Lamb, D.W., Holzapfel, B.P., Louis, J.P., 2011. Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. Precision Agriculture 12, 103–117.

Halliday, J.C.J.C., 2009. Australian Wine Encyclopedia. Hardie Grant Books, VIC.

Hemming, S., de Zwart, F., Elings, A., Petropoulou, A., Righini, I., 2020. Cherry tomato production in intelligent greenhouses-sensors and ai for control of climate, irrigation, crop yield, and quality. Sensors (Basel, Switzerland) 20, 1–30. doi:10.3390/s20226430.

I. Goodwin,, L. McClymont,, D. Lanyon, A. Zerihun, J. Hornbuckle, M. Gibberd, D. Mowat, D. Smith, M. Barnes, R. Correll, 2009. Managing soil and water to target quality and reduce environmental impact.

Kasimati, A., Espejo-García, B., Darra, N., Fountas, S., 2022. Predicting Grape Sugar Content under Quality Attributes Using Normalized Difference Vegetation Index Data and Automated Machine Learning. Sensors 22. doi:10.3390/s22093249.

Kawasaki, K., Uchida, S., 2016. Quality Matters More Than Quantity: Asymmetric Temperature Effects on Crop Yield and Quality Grade. American journal of agricultural economics 98, 1195–1209. doi:10.1093/ajae/aaw036.

Keith Jones, 2002. Australian Wine Industry Environment Strategy.

Knight, H., Megicks, P., Agarwal, S., Leenders, M., 2019. Firm resources and the development of environmental sustainability among small and medium-sized enterprises: Evidence from the Australian wine industry. Business Strategy and the Environment 28, 25–39. doi:10.1002/bse.2178.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, Articles 28, 1–26. doi:10.18637/jss.v028.i05.

Mariani, A., Vastola, A., 2015. Sustainable winegrowing: Current perspectives. International Journal of Wine Research 7, 37–48.

Montalvo-Falcón, J.V., Sánchez-García, E., Marco-Lajara, B., Martínez-Falcó, J., 2023. Sustainability Research in the Wine Industry: A Bibliometric Approach. Agronomy 13. doi:10.3390/agronomy13030871.

Oliver, D., Bramley, R., Riches, D., Porter, I., Edwards, J., 2013. Review: Soil physical and chemical properties as indicators of soil quality in Australian viticulture. Australian Journal of Grape and Wine Research 19, 129–139. doi:10.1111/ajgw.12016.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

SOAR, C., SADRAS, V., PETRIE, P., 2008. Climate drivers of red wine quality in four contrasting Australian wine regions. Australian journal of grape and wine research 14, 78–90. doi:10.1111/j.1755-0238.2008.00011.x.

Srivastava, S., Sadistap, S., 2018. Non-destructive sensing methods for quality assessment of on-tree fruits: A review. Journal of Food Measurement and Characterization 12, 497–526.

Sustainable Winegrowing Australia, SWA., 2021. Sustainable Winegrowing Australia User Manual.

SWA, S.W.A., 2022. Sustainable Wingrowing Australia. https://sustainablewinegrowing.com.au/case-studies/.

Wine Australia, 2019. National Vintage Report 2019 .

Wine Australia, 2021. National Vintage Report 2021 .

Wine Australia, 2022. National Vintage Report 2022 .

Winemakers' Federation of Australia, 2013. National Vintage Report 2013 .

Winemakers' Federation of Australia, 2014. National Vintage Report 2014 .

Winemakers' Federation of Australia, 2015. National Vintage Report 2015 .

Winemakers' Federation of Australia, 2016. National Vintage Report 2016 .

Winemakers' Federation of Australia, 2017. National Vintage Report 2017 .

Winemakers' Federation of Australia, 2018. National Vintage Report 2018 .

Yegge, J.M., 2001. Influence of Sensory and Non-Sensory Attributes of Chardonnay Wine on Acceptance and Purchase Intent. Ph.D. thesis. ProQuest Dissertations Publishing.

ZHU, D.w., ZHANG, H.c., GUO, B.w., XU, K., DAI, Q.g., WEI, H.y., GAO, H., HU, Y.j., CUI, P.y., HUO, Z.y., 2017. Effects of nitrogen level on yield and quality of japonica soft super rice. Journal of Integrative Agriculture 16, 1018–1027. doi:10.1016/S2095-3119(16)61577-0.

Table .7: Summary of models, their predictors, covariates and variable interactions.

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{\text{Yield}}{\text{Area}}$	Average Price Per Tonne	$\frac{\text{Average Price per tonne}}{\text{Area}}$
Yield	1.000E+00	7.440E-01	-4.309E-03	7.290E-01	3.500E-01	-2.262E-01	-1.644E-01
Area	7.440E-01	1.000E+00	-5.331E-03	8.921E-01	7.854E-02	-1.178E-01	-2.042E-01
Water Used	-4.309E-03	-5.331E-03	1.000E+00	-1.929E-03	-5.600E-03	-3.562E-02	-2.669E-02
Scope One Emissions	7.290E-01	8.921E-01	-1.929E-03	1.000E+00	9.357E-02	-9.422E-02	-1.933E-01
$\frac{\text{Yield}}{\text{Area}}$	3.500E-01	7.854E-02	-5.600E-03	9.357E-02	1.000E+00	-4.849E-01	-1.698E-01
Average Price Per Tonne	-2.262E-01	-1.178E-01	-3.562E-02	-9.422E-02	-4.849E-01	1.000E+00	4.732E-01
$\frac{\text{Average Price per tonne}}{\text{Area}}$	-1.644E-01	-2.042E-01	-2.669E-02	-1.933E-01	-1.698E-01	4.732E-01	1.000E+00

Table .8: Pearson correlation coefficients for each logarithmically transformed variable.

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{\text{Yi}}{\text{Ar}}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086
$\frac{\text{Yield}}{\text{Area}}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625
$\frac{\text{Average Price per tonne}}{\text{Area}}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.076

Table .9: P-values for the non-transformed water used variable's Pearson correlation coefficients.

Variable	Water Used
Yield	7.538E-01
Area	6.981E-01
Scope One Emissions	8.883E-01
$\frac{\text{Yield}}{\text{Area}}$	6.836E-01
Average Price Per Tonne	5.600E-02
$\frac{\text{Average Price per tonne}}{\text{Area}}$	1.522E-01

Table .10: Summary statistics for each variable on the original scale..

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{\text{Yi}}{\text{Ar}}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086
$\frac{\text{Yield}}{\text{Area}}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625
$\frac{\text{Average Price per tonne}}{\text{Area}}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.076

Table .11: Model 1 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Year	9	7.060E+01	7.800E+00	8.353E+01	<2.20E-16
GI Region	54	1.507E+03	2.790E+01	2.972E+02	<2.20E-16
Area Harvested	1	3.211E+03	3.211E+03	3.419E+04	<2.20E-16
Water Used	1	1.040E+01	1.040E+01	1.103E+02	<2.20E-16
Scope One Emissions	1	6.600E+00	6.600E+00	7.056E+01	<2.20E-16

Table .12: Model 2 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Area Harvested	1	2.407E+03	2.407E+03	1.080E+04	<2.20E-16
Scope One Emissions	1	3.989E+01	3.989E+01	1.789E+02	<2.20E-16
Water Used	1	5.500E+02	5.500E+02	2.467E+03	<2.20E-16
Area Harvested*Scope One Emissions	1	6.921E+01	6.921E+01	3.104E+02	<2.20E-16
Area Harvested * Water Used	1	1.040E+00	1.040E+00	4.686E+00	3.045E-02 *
Year * GI Region	424	1.144E+03	2.700E+00	1.210E+01	<2.20E-16

Table .13: Model 3 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Year	6	1.324E+01	2.210E+00	8.748E+01	<2.20E-16 ***
GI Region	50	6.498E+02	1.300E+01	5.151E+02	<2.20E-16 ***
Area Harvested	1	2.142E+03	2.142E+03	8.491E+04	<2.20E-16 ***
Water Used	1	3.200E-01	3.200E-01	1.259E+01	3.947E-04 **
Scope One Emissions	1	4.000E-02	4.000E-02	1.492E+00	2.221E-01

Table .14: Model 4 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Area Harvested	1	2.066E+03	2.066E+03	5.700E+04	<2.20E-16
Scope One Emissions	1	6.000E-02	6.000E-02	1.569E+00	2.105E-01
Water Used	1	2.014E+02	2.014E+02	5.557E+03	<2.20E-16
Area Harvested*Scope One Emissions	1	5.246E+01	5.246E+01	1.448E+03	<2.20E-16
Area Harvested * Water Used	1	7.270E+00	7.270E+00	2.005E+02	<2.20E-16
Year * GI Region	243	4.546E+02	1.870E+00	5.162E+01	<2.20E-16

Table .15: Comparison of Model Residuals

	Df	Sum Sq	Mean Sq
Model 1	5231	4.913E+02	1.000E-01
Model 2	4868	1.085E+03	2.200E-01
Model 3	2818	7.111E+01	3.000E-02
Model 4	2629	9.528E+01	4.000E-02

Table .16: Comparison of Model performance.

	RSE	R2	Adjusted R2	F-statistic	P-Value
Model 1	3.065E-01	9.072E-01	9.061E-01	7.753E+02	<2.2e-16
Model 2	4.722E-01	7.951E-01	7.770E-01	4.403E+01	<2.2e-16
Model 3	1.589E-01	9.753E-01	9.748E-01	1.885E+03	<2.2e-16
Model 4	1.904E-01	9.669E-01	9.638E-01	3.095E+02	<2.2e-16