

¹ ,
² Graphical Abstract

³ **An exploratory analysis of the influence of resource use on the yield**
⁴ **verse quality trade-off in Australian vineyards**

⁵ Bryce Polley



⁶ Highlights

⁷ **An exploratory analysis of the influence of resource use on the yield**
⁸ **verse quality trade-off in Australian vineyards**

⁹ Bryce Polley

¹⁰ • Research highlight 1

¹¹ • Research highlight 2

12 An exploratory analysis of the influence of resource use
13 on the yield verse quality trade-off in Australian
14 vineyards

15 Bryce Polley^{a,b,c}

^a*QUT, , , QLD,*

^b*AWRI, , , SA,*

^c*Food Agility CRC, , , Vic,*

16 **Abstract**

17 *Keywords:* Keyword one, keyword two

18 *PACS:* 0000, 1111

19 *2000 MSC:* 0000, 1111

20 **1. Introduction**

21 The global focus on sustainability in agronomic industries has changed the
22 way in which these enterprises do business. When strategies for a sustainable
23 winegrowing industry are assessed, there is a trade-off between balancing the
24 amount of resources invested and the resultant yield verses quality produced.
25 This dilemma exists across agriculture through shared fundamental consider-
26 ations such as water use and nitrogen levels (Hemming et al., 2020; Kawasaki
27 and Uchida, 2016; ZHU et al., 2017). Quality in viticulture (the cultivation
28 of grapes for wine production) is driven through its integration within the
29 wine industry; with a wine’s potential quality being initially defined through
30 the chemical makeup of the grapes used in its production. The consideration
31 of sustainability within viticulture is further complicated by environmental

32 and socio-demographic pressures. In the Australian context, these include:
33 biosecurity, climate and international market demands.

34 In this analysis we observe relationships between yield and quality through
35 the use of linear models. An extensive amount of research into a variety
36 of factors' effect on grape quality and yield exists; but due to the lack of
37 long-term and in-depth data, individual effects are often studied in isolation
38 (Abbal et al., 2016). The lack of consolidated datasets also restricts the
39 ability to gain statistical insights at large scales and across multiple regions
40 (Keith Jones, 2002; Knight et al., 2019). The dataset used for this analysis
41 includes data collected for the past 10 years from a multitude of vineyards
42 located over a diverse range of Australian winegrowing regions.

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

51 **2. Methods**

52 We created four linear models to explore relationships between resource-
53 use and vineyard outputs (see Table1). The data was sourced from Sustain-
54 able Winegrowing Australia and Wine Australia. Variables used included:
55 yield, average sale price, region, water use, emissions, area harvested and

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 \times 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

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 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 relation-

ships 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

92 national wine industry sustainability program, which aims to facilitate grape-
93 growers and winemakers in demonstrating and improving their sustainability
94 (SWA, 2022). Wine Australia is an Australian Government statutory author-
95 ity governed by the Wine Australia Act 2013 (Win, 2019).

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

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

117 price of sale per tonne.

118 The first subset of data was used for Model 1 and Model 2 (see Table1).

119 This subset contained 5298 samples spanning the period from 2012 to 2022,

120 covering 55 GI Regions and 1261 separate vineyards.

121 The second subset of data, was limited to vineyards that recorded a value

122 for their average sale price of grapes per tonne. This subset was used for

123 Model 3 and Model 4 (see Table1); and contained 2878 samples spanning

124 the period from 2015 to 2022, covering 51 GI Regions and 944 separate vine-

125 yards. 1842 of the values for average price of sale per tonne were extracted

126 from Wine Australia surveys with the remaining 1036 being from Sustainable

127 Winegrowing Australia’s dataset.

128 Additional variables were considered for analysis but were excluded due to

129 being either underreported or had insignificant contributions to model accu-

130 racies. Variables explored but not used due to low reporting values included:

131 fertiliser, and scope 2 emissions. Variables considered but ultimately removed

132 due to a lack of significant contributions to models, included: the use of re-

133 newable energy, contractor use, and pressures such as frost, fire and disease.

134 Data preprocessing was conducted prior to analysis using the Python pro-

135 gramming language (G. van Rossum, 1995). Preprocessing included logarith-

136 mic transformations, centring and scaling by standard deviation. Variables

137 such as scope 1 emissions, which required prior calculations were also com-

138 puted using Python.

139 *2.4. Total Emissions*

140 The equation given from the Australian National Greenhouse Accounts

141 Factors, shown as

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

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, tCO_2e (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.

166 2.6. Model Validation

167 Models were validated using K-fold cross validation calculated through
168 the R Caret Package (Kuhn, 2008). K-fold cross validation works by remov-
169 ing a subset of data from the sample used to train models and then predicts
170 those variables to determine how sensitive the model is to changes in the sam-
171 ple data. For this analysis each model was validated using 10 folds, repeated
172 100 times.

173 3. Results

174 3.1. Data

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

180 3.2. Exploratory Analysis

181 Linear relationships between variables were explored using Pearson Cor-
182 relation Coefficients. Values for these coefficients reflect the linear relation
183 between two variables, on a scale between -1 and 1; the magnitude and sign
184 of a coefficient indicates the strength of the relation, and whether the rela-
185 tion is positive or negative respectively. This was undertaken for data on the
186 original scale and for data as a logarithmic transform. The logarithmic trans-
187 formed data showed the strongest correlations, likely due to a skew caused
188 by a greater number of smaller vineyards within the dataset (see Table 3).

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
$\frac{\text{Yield}}{\text{Area}}$	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
$\frac{\text{Average Sale Price}}{\text{Area Harvested}}$	1.347E+02	5.711E+02	1.753E-01	2.979E+04

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

189 Transforming data prior to calculating the coefficients changes several things:
 190 The logarithmic transform of the data alters the interpretation of the coef-
 191 ficients to percentage change - a coefficient will be indicative of the change
 192 in percentage of one variable compared to the other; scaling by standard de-
 193 viation also changes this interpretation to be a percentage of that variables
 194 standard deviation. Scaling by standard deviation also makes the Pearson
 195 Correlation Coefficient equal to the covariance of the two variables. With all
 196 this in mind, when considering the logarithmically transformed variables, a
 197 coefficient of 1 would indicate that: given the change of one variable by one
 198 percentage of its standard deviation, the other variable would change by one
 199 percent of its own standard deviation. The importance of this is the dimen-
 200 sionless nature of these relationships and that it can be translated directly
 201 to any vineyard's case that has a well known distribution.
 202 To determine if a coefficient was indicative of a strong relationship, confidence
 203 intervals were used. P-values reflected the significance of a given correlation
 204 coefficient when considering its relation to sample size via its incorporation as
 205 an element of standard error. Strong relationships were found to be present
 206 as all P-values, except for the non-transformed values for water used, were
 207 considered significant ($P < 2.200E-16$).

208 *3.3. General Linear Models*

209 General Linear Models were used to describe how response variables re-
 210 lated to predictors' values. Log transformed variables were used as inputs to
 211 these models as they resulted in higher R^2 values and described the relation-
 212 ships proportionally; reflecting coefficient values as percentages of a variable's
 213 standard deviation. Each model showed a strong relationship between the

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

214 predictors and the response (see Table 4). Model accuracy was measured in
 215 R^2 , as this allowed an easy comparison between their performances and their
 216 validation.

217 3.3.1. *F-tests*

218 To determine if predictors significantly related to a Model's response vari-
 219 able, F-tests were conducted. Aside from 3 variables, all F-tests across each
 220 model indicated a significant contribution at 95% confidence. The three ex-
 221 ceptions were: scope 1 emissions in Model 3 (P=2.221E-01) and Model 4
 222 (P=3.621E-01), and Model 2's interaction between area harvested and water
 223 used (P=2.192E-01).

224 Scope 1 emissions was included in all models to directly compare the response
 225 variables as ratios of vineyard size to raw values. Even though not significant
 226 within models 3 and 4, when using the Pearson Correlation Coefficients scope
 227 1 emissions was strongly correlated to every Model's response variable; this
 228 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

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

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 for each category, whilst continuous variables have only one. The coefficient for categorical variables is summarised in Figure 3.3.3; 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 Verse 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.

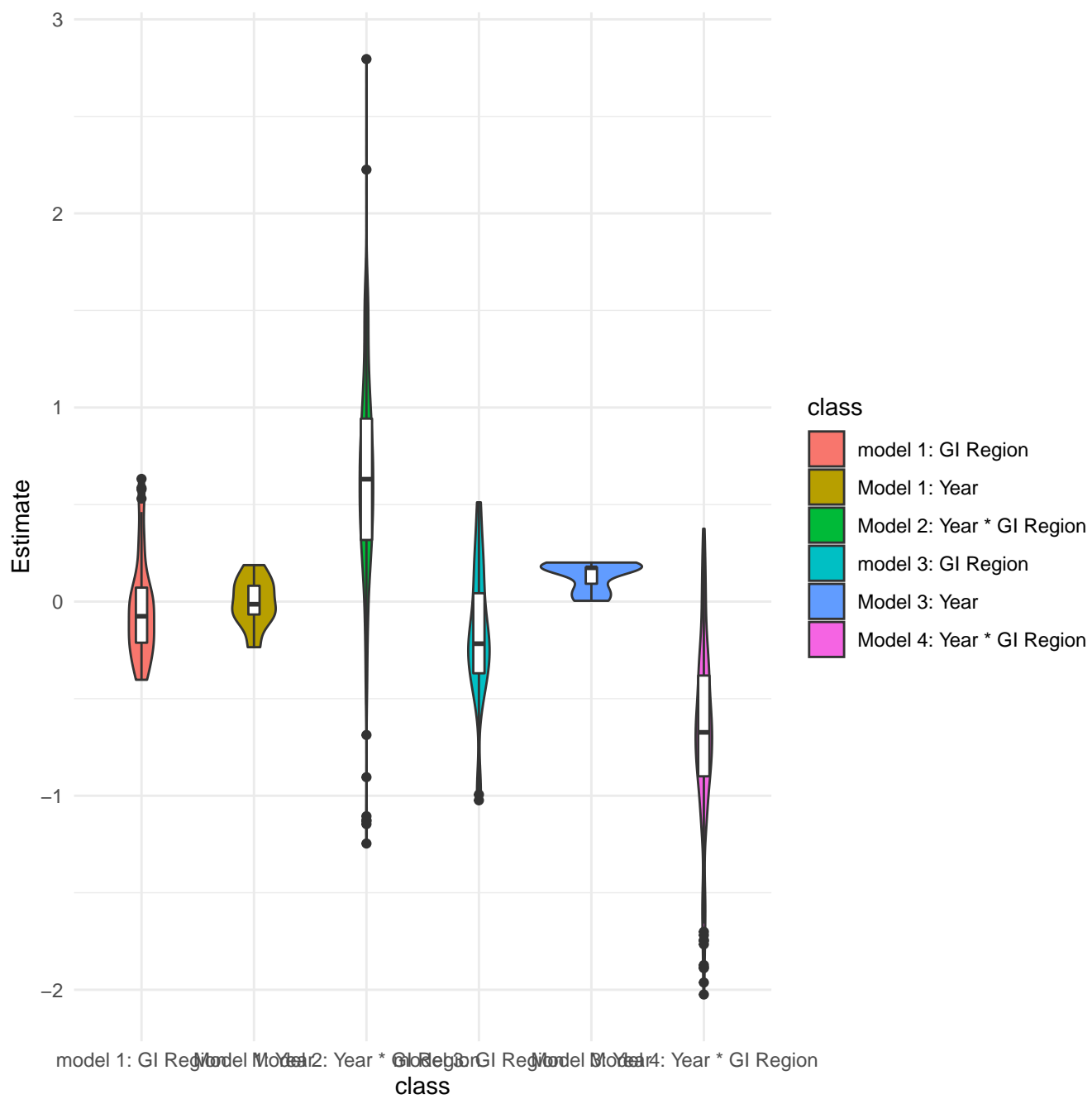


Figure 1: Violin plots of GI Region and Year coefficients for each model.

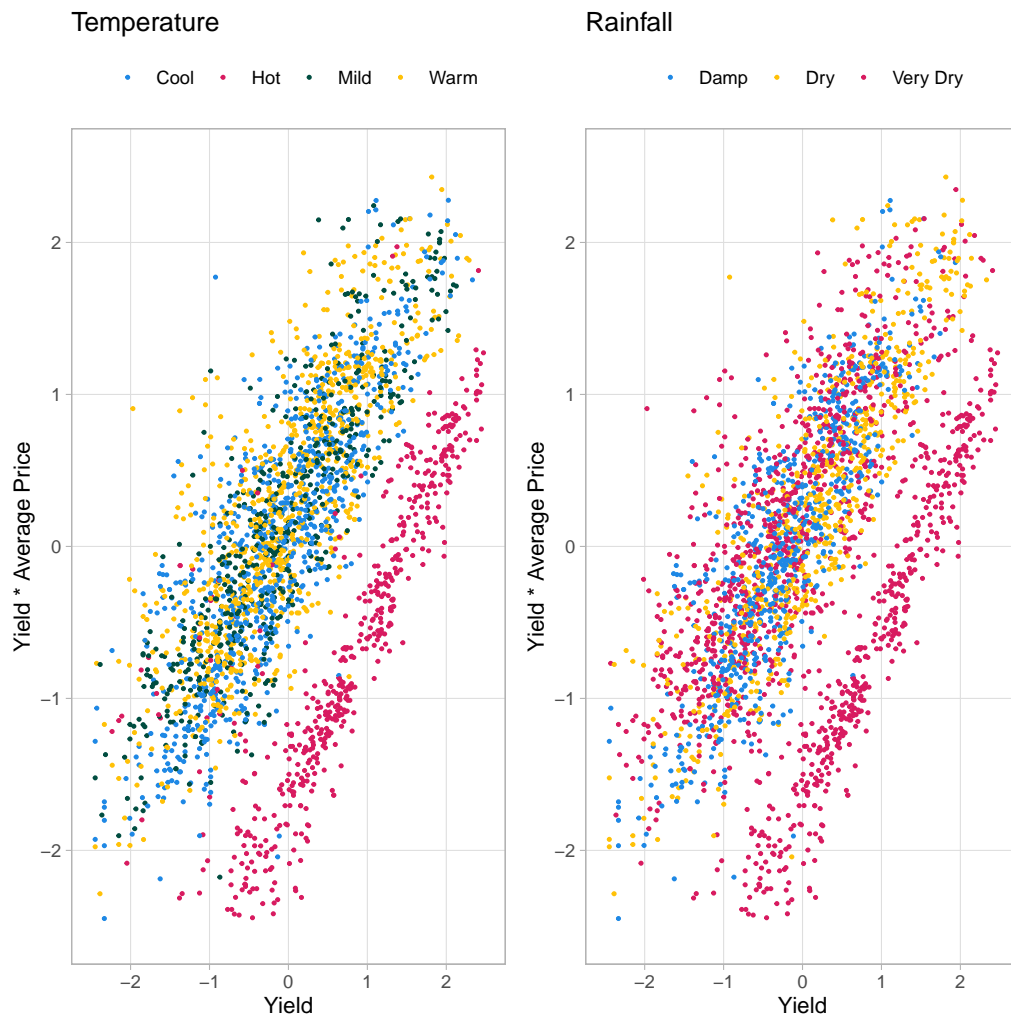


Figure 2: Scatter plot of vineyard yield against the product of yield and average price per tonne. The axes are in standard deviations with points coloured by climate.

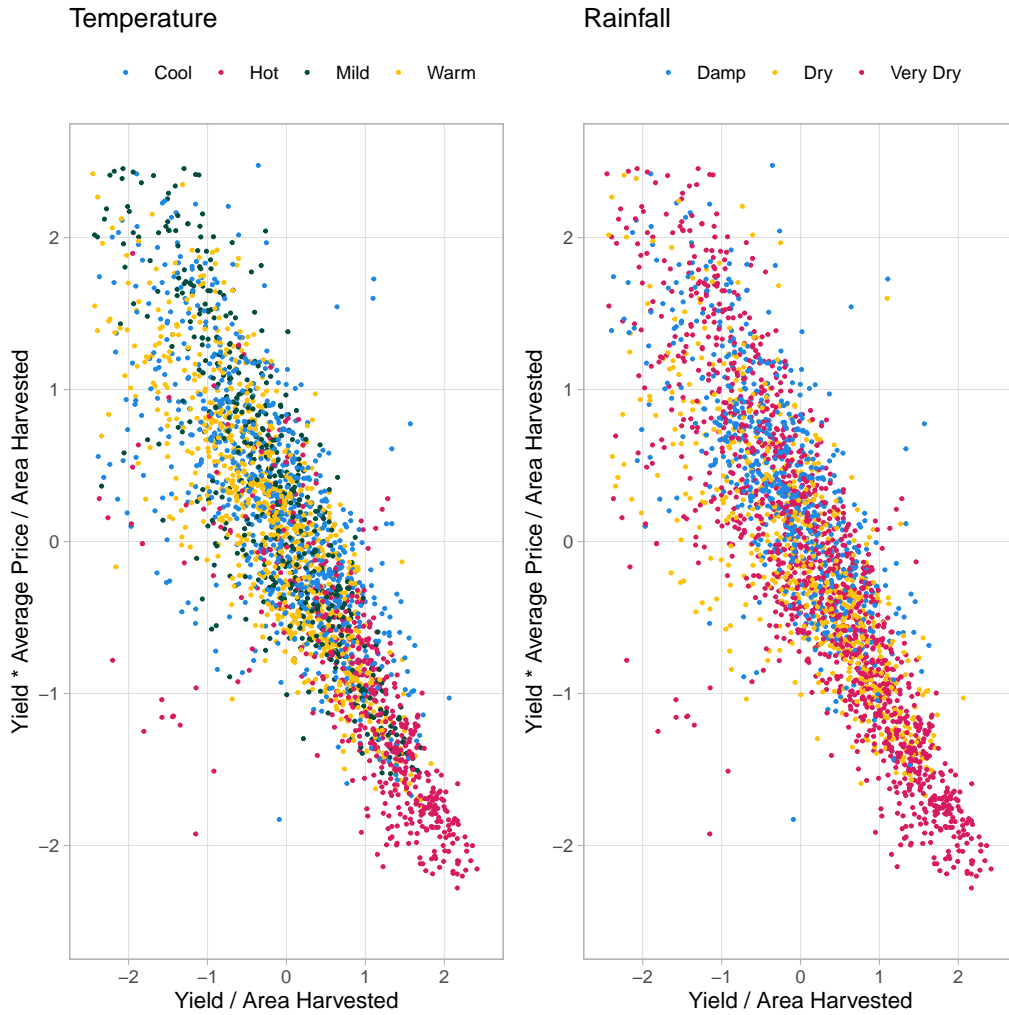


Figure 3: Scatter plot of vineyard yield against the product of yield and average price per tonne as ratios to area harvested. The axes are in standard deviations with points coloured by climate.

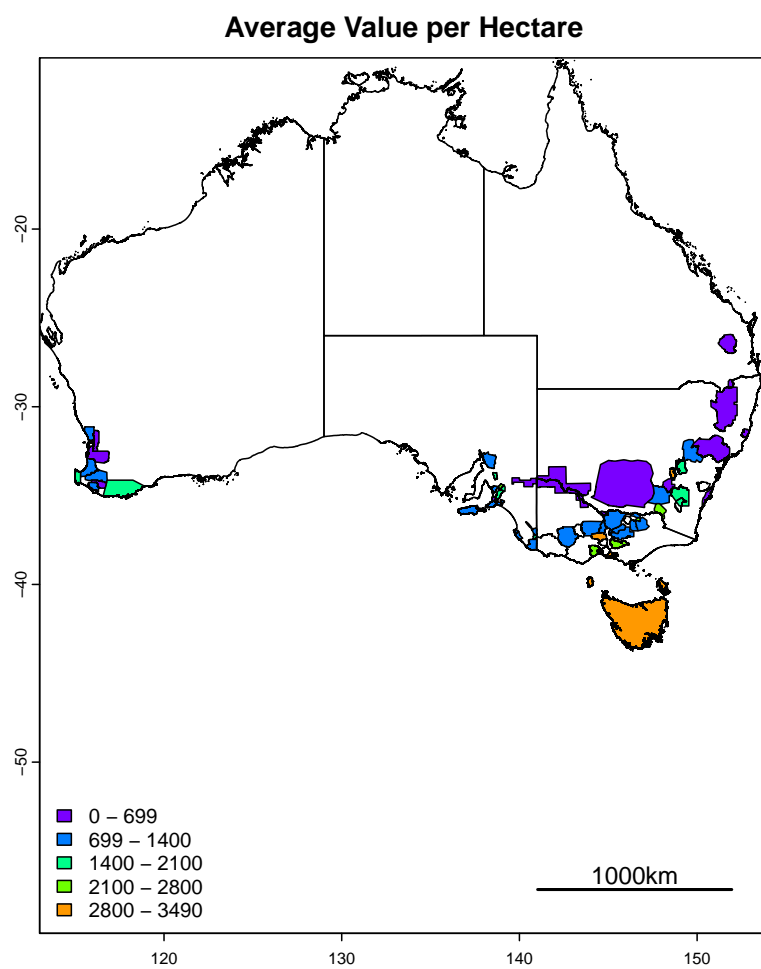


Figure 4: Map of regional average yield and value per hectare.

294 Observing Figure 3.3.4 shows an almost discrete difference between vine-
295 yards in 'Hot' areas than other regions. Comparing Figure 3.3.4 to Figure
296 3.3.4 shows almost opposing trends; a not so obvious difference between the
297 Figures, is that the difference is mostly a rotation (being 90° clockwise).
298 However, with area coming into play, many data points are scaled differ-
299 ently; specifically the vineyards from 'Hot' regions are then found to be on
300 the tail end, producing large quantity of lower value grapes. This is more
301 visible when comparing both graphs to the map of regional averages for re-
302 sponse variables, see Figure 3.3.4. There is a notable change between regional
303 averages when looking at yield verse value. Through the coefficients we can
304 deduce that: this difference is also a difference between more resources used
305 for the raw response variables; and a difference between overall resource use
306 and the size of the vineyard. Where resource use and area harvested have
307 a combined relationship through their interaction and separate relationships
308 as individual variables (see Table 5). A notable occurrence in Figure 3.3.4, is
309 that the 'Very Dry' vineyards which produce lower yields and higher quality
310 grapes are predominantly found in the Barossa Valley (a wine region known
311 for its high quality Shiraz). This note is important as it shows climate is not
312 exclusively the consideration, soil and other geographical phenomenon have
313 considerable impacts on vineyard outcomes.

314 3.4. Model Validation

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

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

318 4. Discussion

319 In alternative attempts at models it was found that without the incor-
320 poration of GI Region or year the predictions greatly under performed. The
321 possible reason behind this effect was that different strategies are likely em-
322 ployed between different regions, where some regions target the mass produc-
323 tion of cheaper grapes over quality. Reviewing the data to uncover reasons
324 for this included the use of binary variables such as the utilisation of re-
325 newable energy, contractors, and the occurrence of disease, fire and frost;
326 however none of these variables were able to explain why some vineyards
327 produced less, or why other vineyards sold at higher prices than predicted.
328 A wide variety of these influences were likely already explained within the
329 use of year and GI Region, or the interaction of both variables. The change
330 between some regions was dramatic, with particularly warmer and drier re-
331 gions producing much higher volumes of grapes at lower prices (See Figures
332 5 and 6). The use of other variables and methods, specifically splines, were

333 able to create a more normally distributed set of residuals but at a drastically
334 reduced accuracy when comparing R2 and RSE. The introduction of known
335 average prices per tonne also helped increase R2 values a small amount; it is
336 important to not that it is common practice for wineries to purchase grapes
337 at a regional average rate, likely resulting in much less variance within a
338 region.

339 different strategies are likely employed between different regions, where
340 some regions target the mass production of cheaper grapes over quality. This
341 is most notable when grouping regions by climate, especially when considering
342 GI Regions in the 'Hot Very Dry' climate (see Figure 7). The effect of
343 climate in the models was not more significant than the more granular use
344 of GI regions. The interaction between year and GI Region likely accounted
345 for localised events such as bushfires, which would be impactful, but only at
346 a local level in both time and space.

347 *4.1. Limitations*

348 Limitations included overestimating yield for models 1 and 2, (see Figures
349 1 and 2) and underestimating crop value in models 3 and 4 (see Figures
350 3 and 4). This study investigated the general relationships between input
351 resources of a vineyard, including fuel and water, and the outputs including
352 yield and value. Some regions appeared to produce many low quality grapes
353 at scale compared to attempting to produce fewer higher quality grapes. This
354 behaviour can be observed when reviewing Wine Australia's annual reports,
355 where it is apparent that warm inland regions such as the Riverland are
356 known to only produce large amounts of lower graded grapes Wine Australia
357 (2022); Winemakers' Federation of Australia (2017). Comparatively, regions

358 such as Tasmania only produce A grade grapes but in much smaller quantities
359 than the Riverland. Knowing that the difference in pricing per tonne can
360 exceed a magnitude of 10 between grades E and A, the operations in regions
361 that target different grades would have varied priorities. However, some
362 regions such as the Yarra Valley produce a Variety of different grades of
363 grapes, from C to A, highlighting that vineyard priorities, although may be
364 somewhat present within regional classifications, are not necessarily aligned
365 within a given region.

366 The opportunity to target different grades of grapes may not always be
367 available, with some regions being more renowned than others, and likely to
368 be sought after regardless (Halliday, 2009). The Barossa is an example of
369 this, known for its quality could also lend itself to a bias in purchasers not
370 considering other regions that may be capable of similar quality. This effect
371 could stifle the potential for market opportunities within these lesser known
372 regions. A further possibility is that there may be regional upper limits with
373 the relationship between resource input and the value gained becoming no
374 longer proportional due to diminishing returns. Climate was considered to be
375 a large determinant of the ability to grow a larger quantity of grapes, as well
376 as a determinant in grape quality (Agosta et al., 2012); however there were
377 vineyards in similar regions that were able to produce exceptionally better
378 results than others (See Figure 7).

379 The issue of model 1 and 2 over predicting yield, may have been due to
380 preventative measures brought on by regional pressures such as fire, frost and
381 disease. Where, more resources were required to prevent these issues from
382 spreading within a region, thus disproportionately effecting some vineyards

383 compared to others locally. This type of maintenance is not well captured
384 especially when considering that some regions, those in warmer areas are
385 not as prone to disease as cooler climates and could potentially have lower
386 operating costs per hectare. This could create a discrepancy in vineyards that
387 utilise preventative measures in wetter regions, as opposed to those who do
388 not, and thus expend less fuel and energy but risk disease. When reviewing
389 the differences between regions it is important to consider that vineyards
390 in Hot Very Dry areas can be hundreds of times the size of those in other
391 regions. It is interesting that while area, although significantly correlated to
392 the ratio of yield to area, was still lower than water and about the same as
393 emissions. This points to economies of scale playing a role but still being
394 only one consideration alongside the potential resources that can be used.
395 The negative trend between size and average sales price could also be a side
396 effect of mass supply verse demand, especially when looking at the level of
397 difference in production of some vineyards (see Table 4). The relationships
398 between yield, value and area are not simply about efficiently producing the
399 most grapes; sales price and by association grape quality, are integral to the
400 profitability, and this is strongly linked to resource-use and thus the longevity
401 and sustainability of a vineyard.

402 Literature shows that there are many on-the-ground decisions that in-
403 fluence both quality and yield. Where these decisions are governed by com-
404 plex physical and social forces such as international market demands, disease
405 pressures and natural disasters (Abad et al., 2021; Cortez et al., 2009; Hall
406 et al., 2011; I. Goodwin, et al., 2009; Kasimati et al., 2022; Oliver et al.,
407 2013; Srivastava and Sadistap, 2018). Many of these occurrences being high-

408 lighted throughout the past decades vintage reports (Wine Australia, 2019,
409 2021, 2022; Winemakers' Federation of Australia, 2013, 2014, 2015, 2016,
410 2017, 2018). It is also important to consider that these reports show that
411 the warm inland regions have seen a decline in profit during this period, as
412 they were often compared to other regions that focused more on quality than
413 quantity. This is an important consideration, as the size of some of these
414 vineyards when considering their ratio of value to area would only require a
415 marginal increase to out compete other regions. There are also differences
416 when comparing winegrowers to other agricultural industries as they are ver-
417 tically integrated within the wine industry, tying them to secondary and
418 tertiary industries, such as wine production, packaging, transport and sales.
419 This results in unique issues and considerations for each vineyard, where these
420 on-the-ground decisions may be influenced by other wine industry's choices,
421 such as the use of sustainable practices in vineyards as a requirement for sale
422 in overseas markets; notably these interactions are further complicated by
423 some winegrowers being totally integrated into wine companies, while others
424 are not (Knight et al., 2019). Incorporating such decisions into the model
425 could help describe the contributing factors to regional differences beyond
426 resource consumption and regional differences.

427 Having more data for each region would also be an improvement, allowing
428 greater comparison between regions. More variables may also help to discern
429 vineyards that can produce larger volumes of grapes at higher prices. The use
430 of semi transparent tools such as random forests and decision trees alongside
431 more variables and data may help to uncover the reasons for values that
432 were under or over estimated. These differences could be caused by the use

of alternative sustainable practices in the field. 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).

The relationship between scope one emissions and the response variables that included average sales price

It is possible that the relationships between scope one emissions and the response variables were closely tied to a vineyards area. This possibility could be explained through the emissions

Noting that irrigation systems use fuel and that the application of water was a significant variable in each model scope one emissions' lack of significance and contribution given its F-statistics (See Tables 7 and 8), indicated that it is possible other vineyard activities requiring fuel are not as determining factors for a vineyards grape quality.

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

456 Based on a Bayesian Network. *Journal of agricultural, biological, and*
 457 *environmental statistics* 21, 131–151. doi:10.1007/s13253-015-0233-2.

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

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

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

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

469 Fraga, H., Costa, R., Santos, J.A., 2017. Multivariate clustering of viticul-
 470 tural terroirs in the Douro winemaking region. *Ciência Téc. Vitiv.* 32,
 471 142–153.

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

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

- 477 Halliday, J.C.J.C., 2009. Australian Wine Encyclopedia. Hardie Grant
478 Books, VIC.
- 479 Hemming, S., de Zwart, F., Elings, A., Petropoulou, A., Righini, I., 2020.
480 Cherry tomato production in intelligent greenhouses-sensors and ai for con-
481 trol of climate, irrigation, crop yield, and quality. *Sensors* (Basel, Switzer-
482 land) 20, 1–30. doi:10.3390/s20226430.
- 483 I. Goodwin,, L. McClymont,, D. Lanyon, A. Zerihun, J. Hornbuckle, M.
484 Gibberd, D. Mowat, D. Smith, M. Barnes, R. Correll, 2009. Managing soil
485 and water to target quality and reduce environmental impact.
- 486 Kasimati, A., Espejo-García, B., Darra, N., Fountas, S., 2022. Predicting
487 Grape Sugar Content under Quality Attributes Using Normalized Differ-
488 ence Vegetation Index Data and Automated Machine Learning. *Sensors*
489 22. doi:10.3390/s22093249.
- 490 Kawasaki, K., Uchida, S., 2016. Quality Matters More Than Quan-
491 tity: Asymmetric Temperature Effects on Crop Yield and Quality
492 Grade. *American journal of agricultural economics* 98, 1195–1209.
493 doi:10.1093/ajae/aaw036.
- 494 Keith Jones, 2002. Australian Wine Industry Environment Strategy.
- 495 Knight, H., Megicks, P., Agarwal, S., Leenders, M., 2019. Firm resources and
496 the development of environmental sustainability among small and medium-
497 sized enterprises: Evidence from the Australian wine industry. *Business*
498 *Strategy and the Environment* 28, 25–39. doi:10.1002/bse.2178.

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

502 Mariani, A., Vastola, A., 2015. Sustainable winegrowing: Current perspec-
503 tives. International Journal of Wine Research 7, 37–48.

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

507 Oliver, D., Bramley, R., Riches, D., Porter, I., Edwards, J., 2013. Review:
508 Soil physical and chemical properties as indicators of soil quality in Aus-
509 tralian viticulture. Australian Journal of Grape and Wine Research 19,
510 129–139. doi:10.1111/ajgw.12016.

511 R Core Team, 2021. R: A Language and Environment for Statistical Com-
512 puting. R Foundation for Statistical Computing.

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

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

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

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

523 Wine Australia, 2019. National Vintage Report 2019 .

524 Wine Australia, 2021. National Vintage Report 2021 .

525 Wine Australia, 2022. National Vintage Report 2022 .

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

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

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

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

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

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

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

535 ZHU, D.w., ZHANG, H.c., GUO, B.w., XU, K., DAI, Q.g., WEI, H.y., GAO,
536 H., HU, Y.j., CUI, P.y., HUO, Z.y., 2017. Effects of nitrogen level on yield
537 and quality of japonica soft super rice. Journal of Integrative Agriculture
538 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	-	7.290E-01	3.500E-01	-2.262E-01	-1.644E-01
Area	7.440E+00	1.000E+00	5.331E-03	8.921E-01	7.854E-02	-1.178E-01	-2.042E-01
Water Used	-	-	1.000E+00	1.929E-03	-	-3.562E-02	-2.669E-02
Scope One Emissions	7.290E-01	8.921E-01	-	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	-	-	-	-9.422E-02	-	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	-	4.732E-01	1.000E+00

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

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{Yield}{Area}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353E-01
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742E-01
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292E-01
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086E-01
$\frac{Yield}{Area}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000E+00
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625E-01
$\frac{Average Price per tonne}{Area}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.070E-01

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
$\frac{Yield}{Area}$	6.836E-01
Average Price Per Tonne	5.600E-02
$\frac{Average Price per tonne}{Area}$	1.522E-01

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

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{Yield}{Area}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353E-01
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742E-01
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292E-01
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086E-01
$\frac{Yield}{Area}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000E+00
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625E-01
$\frac{Average Price per tonne}{Area}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.070E-01

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