Management tool for oenological decision-making: Modeling and optimization of a hybrid model for fermentative maceration of Cabernet Sauvignon

Ricardo Luna[1], Cristóbal Torrealba[1,2], J. Ricardo Pérez-Correa[2], and Álvaro González[1]

[1]Centro de Investigación e Innovación de Viña Concha y Toro, Ruta K-650 km 10, Pencahue, Chile
[2]Departamento de Ingeniería Química y Bioprocesos, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

Abstract. This work presents a hybrid model for *Cabernet Sauvignon* (CS) red wine-making that combines mechanistic and data-driven approaches to optimize the fermentation process and improve the quality of red wine. The model incorporates two sub-units representing the interaction between alcoholic fermentation and phenolic extraction, considering factors such as temperature, products addition, draining time, and must composition. To develop and validate the model, a database of 270 industrial CS fermentation from 2017-2021 harvest seasons was collected. The models were calibrated using experimental data, achieving an average R^2 of 0.94 for fermentation kinetics model and 45% and 80.9% test accuracy for tannins and anthocyanins predictors, respectively. A multi-objective dynamic optimization problem was formulated and solved to find fermentation operation conditions that optimize simultaneously phenolic quality, process costs and productivity. A similar distribution of the Pareto fronts were obtained for varietal and premium wines. Finally, these tools were packed in a digital platform for practical use in industrial cellars. The models generate the predictions and recipes prescription for each fermentation tank when the pre fermentative juice is analyzed. As a result, it is obtained useful information for wine decision-making like maceration length and wine phenolic composition at least five days in advance.

1 Introduction

By incorporating Industry 4.0 technologies into winemaking, the sector is revolutionized by integrating physical production and operations with cutting-edge digital tools such as the Internet of Things (IoT), cloud computing, data analytics, and artificial intelligence (AI). This fusion gives rise to intelligent factories outfitted with state-of-the-art sensors, comprehensive software, and robotics that gather and interpret data for better decision-making [1,2]. As a result, wineries can streamline production processes, boost efficiency, and increase wine quality, fostering a more sustainable, efficient, and lucrative industry [2].

In this study, we report the application of Industry 4.0 technologies to the industrial-scale production of red wine. The red wine fermentation process implicates an interplay of alcoholic fermentation and skin maceration, which may involve seeds and occasionally stems. Maceration is a physicochemical process that promotes the extraction of compounds from grape skins [3,4], while alcoholic fermentation, driven by yeast, is a biological process that converts sugars into ethanol and carbon dioxide. This transformation enriches the wine's complexity through the generation of secondary metabolites [4,5].

Numerous traditional techniques have been employed to improve the extraction of grape constituents and fermentation in winemaking, emphasizing red wine's color and phenolic compounds. Pre-fermentation approaches include cold soaking, carbonic maceration, enzyme addition, thermovinification, flash détente, and accentuated cut edges (ACE) [4,6]. During fermentation, managing the cap is critical for promoting the extraction of grape components, particularly polyphenols, from the solid cap into the liquid must [4,7]. This is accomplished by regularly contacting the cap with the liquid, typically using pump-over or punch-down methods. Maceration duration and temperature also significantly influence the extraction process. Anthocyanins and tannins are extracted from the skins in the initial stages, whereas seed tannins extraction prevails in the latter stages as ethanol levels rise [3,8,9]. Maintaining temperature control is crucial for achieving consistent fermentation and extracting polyphenols [7,10-12]. Micro-oxygenation can enhance wine quality during fermentation by intensifying

the red hue, stabilizing the wine aroma, and simulating barrel aging [4,9,13].

Mathematical models for process simulation have become increasingly necessary in enhancing the quality of red wine [4,9,14,15]. These models can be categorized into two primary types: mechanistic and non-mechanistic. Mechanistic models depict processes grounded in physical, chemical, and biological principles, while non-mechanistic models rely on data-driven approaches [16]. Furthermore, hybrid models, which integrate the benefits of both mechanistic and non-mechanistic models, demonstrate potential for utilization in process control and optimization [16].

Recent advancements in mechanistic modeling have yielded promising tools for simulating red wine production. Examples of contemporary alternatives include using Genome-Scale Metabolic Models (GEMMs) or expanding traditional dynamic models with secondary aroma metabolism [17,18]. Additionally, models that combine fermentation and extraction kinetics of phenolic compounds during maceration have exhibited promise in large-scale red wine fermentation operations [10,19,20].

By incorporating IoT infrastructure, wineries can now gather and utilize data to make informed decisions regarding wine phenolic composition and maceration length several days ahead [5,17,21]. Combining predictive models with in-line sensing and real-time data acquisition enables the development of automated process control systems for optimized fermentation and extraction [5,17,21]. Effective control strategies take into account temperature, mechanical operations, and aeration to enhance phenolic compound extraction and fermentation kinetics. This technological integration contributes to winemaking efficiency, facilitating the production of high-quality red wine with greater consistency, predictability, and cost-effectiveness [2,3,9,19,21].

Nonetheless, optimizing the control of alcoholic fermentation in winemaking presents a formidable challenge. The control parameters of both processes, fermentation, and maceration, differ, necessitating the identification of compatible conditions for managing them simultaneously [9,20]. Commercial winemaking aims to optimize factors such as sugar exhaustion, fermentation duration, and the energy needed to regulate fermentation temperature, which can be difficult to quantify [4,9,17,21,22]. Consequently, future advancements in this field will continue to enhance the precision and accessibility of Industry 4.0 tools to refine winemaking processes further.

This article illustrates how mathematical models were applied to optimize industrial winemaking processes. We developed a hybrid model to optimize Cabernet Sauvignon red winemaking recipes by integrating first-principles modeling and machine learning techniques, balancing time, cost, and phenolic composition. The model comprises two interconnected sub-units representing the interplay between alcoholic fermentation and phenolic extraction.

We trained and validated the models using a database of 270 industrial Cabernet Sauvignon fermentations, subsequently incorporating them into a digital platform for practical implementation in industrial wineries. The models generate tailored predictions and recipes for each fermentation tank by analyzing the pre-fermentative juice. This optimization process ultimately streamlines the winemaking timeline, reduces costs, and refines the phenolic composition of the wine.

2 Materials and methods

2.1 Machine learning model construction

Data-driven models were developed to predict draining point anthocyanin and tannin concentration. For this, we employed KDD as the processing pipeline considering the nature of data used for model construction [23,24].

2.1.1 Industrial process database

A database was developed from 270 Cabernet Sauvignon industrial winemaking processes during the 2017-2021 harvest seasons. Information in this dataset included chemical and phenolic composition of grapes and wines, oenological product applications, monitoring data including temperature and density, total weight of processed grapes, fermentation tank volume, and maceration times. The database included seven wineries and Cabernet Sauvignon grapes from 5 valleys in Chile. Samples were collected in two points during the maceration process; pre-fermentative must and wine at the draining point. The chemical analysis for pre-fermentative must included density, Brix, pH, total acidity, free SO_2, total SO_2, and YAN. For must and wine, total anthocyanins and tannins were measured by UV/VIS spectrophotometry (Cary 60, Agilent Technologies). The anthocyanins measurement methodology consisted of mixing a 10 ml sample with 1 ml of a 0.1% w/v HCl in 95% ethanol solution and 20 ml of 2% w/v HCl distilled water solution. The mixed solution is then split into two test tubes with 10 ml each, adding 4 ml of a 15% w/v $Na_2S_2O_5$ solution to the first tube, and 4 ml of distilled water to the second tube [25]. Then, the solutions were left for 20 minutes before reading their absorbance at 520 nm against a water blank. The concentration of total tannins was measured using a precipitation method by a methylcellulose reaction and reading the absorbance at 280 nm in a UV/VIS spectrophotometer [26]. A curve with eight calibration points was used for sample analysis.

2.1.2 Transformation and Data Mining

The engineering variable process was used to transform, select, and adapt variables from the database to be used in training machine learning (ML) algorithms. For example, must volume was estimated using grape destemming industrial yield, and then oenological product data such as yeast, diammonium phosphate (DAP), among others, were transformed from amount into concentration using

the must volume. Subsequently, relevant patterns, relationships, and insights hidden in the converted data were determined using Python's Scikit-Learn package. A standard scaler was applied for numerical normalization, principal component analysis (PCA) was applied for reducing data dimension and feature extraction. A Random Forest algorithm was applied for identifying relevant features and variables to ensure model interpretability, considering this classifier's exceptional performance and robustness against overfitting [27,28].

2.1.3 Benchmark and Model selection

Several machine learning algorithms included in Python's Scikit-Learn package were trained for ML model development. Algorithms included in this study were Kernel Ridge, Support Vector Machine, Random Forest Regression, Gradient Boosting, Stochastic Gradient Descent, K-Nearest Neighbors, and Gaussian Processes (varying in kernels).

Hyperparameters of ML models were optimized using nested cross-validation, achieving accurate predictions of draining-point total anthocyanins and tannins concentrations while overcoming bias in performance evaluation during model selection [29]. Model performance was assessed using standard metrics such as mean-squared error and R-squared, and the data was split 80% for training and 20% for validation. The best-performing model, with the highest accuracy and reliability, was selected based on the validation results.

2.2 Mechanistic model of alcoholic fermentation

A model for simulating alcoholic fermentation was developed using a procedure for robust model structure selection and parameter estimation combined with fermentation kinetics data generated through experimental fermentation processes [30].

2.2.1 Experimental design for fermentation kinetics

Six red wine fermentations were performed using Cabernet Sauvignon grapes from various producers in Chile's Maule region. The calibration of the wine fermentation model involved both laboratory-scale and pilot-scale experiments. Following a standardized commercial winery protocol, the grapes were crushed to obtain juice and grape solids. The pre-fermentative juice underwent adjustments for sugar content, pH, and yeast assimilable nitrogen (YAN) through dilution and adding specific compounds. Inoculation was done using the commercial Saccharomyces cerevisiae yeast strain (Maurivin PDM), and DAP was added during fermentation.

The laboratory-scale experiments were performed in a 5 L reactor with temperature control set at 26 °C using a heating/cooling jacket. Pumping-over applied thrice daily for two minutes enhanced pomace extraction and nutrient distribution. The pilot-scale experiments were carried out in a 1000 L cubic reactor with temperature control set at 26 °C using a heating/cooling coil, employing a similar pumping-over system and parameters from the laboratory-scale experiments, scaled up accordingly. Sampling was based on density reduction, with samples collected at 10-15 g/L intervals.

The samples were analyzed using an automatic spectrophotometric analyzer (Y15, BioSystems) specifically designed for commercial wineries. The analyzer measured key oenological metabolites, including glucose, fructose, and yeast assimilable nitrogen (YAN) concentrations. Sugar consumption (°Brix and density) was measured using a portable densimeter (DMA35, Anton Paar). Additionally, must and cap temperatures were recorded each minute using PT1000 sensors.

2.2.2 Modeling and parameter estimation

A systematic model reparameterization procedure was applied to generate reliable, robust, and flexible model structures from lab-scale data, which can then be transferred to large-scale systems [30]. The model structures were developed using data from a 5 L laboratory-scale bioreactor, and a 1000 L pilot-scale fermenter was used to validate the derived model structures.

A priori and a posteriori regression diagnostics were used to assess each model structure's parameter identifiability, significance, sensitivity, and fitting performance [31]. Multi-criteria decision-making methods were then utilized to select a reduced set of models with desirable characteristics. Statistical indices were used to further analyze the selected model structures with additional calibration and validation data, resulting in a single, fully calibrated, and robust model for simulation that can fit data from different experiments of the same or similar system.

The obtained model structure displayed a good predictive capacity, which includes free parameters that are influential, uncorrelated, and significantly different from zero [30]. The proposed procedure was applied to two models taken from the literature [32,33]. The resulting model was calibrated and validated using data gathered from laboratory-scale and pilot-scale systems. For each case, calibration and validation sets were defined using a random set of 4 training and two validation datasets.

2.3 Hybrid model

2.3.1 Hybrid model structure

A hybrid model scheme for optimizing red wine alcoholic fermentations is shown in Fig. 1. The model combines the previously described mechanistic and data-driven models to comprehensively understand fermentation kinetics and tannin/anthocyanin extraction during red wine fermentation. The data-driven model takes as input the operating parameters of fermentation and chemical analysis of the red grape must and predicts the final concentration of tannins and anthocyanins achieved through this process.
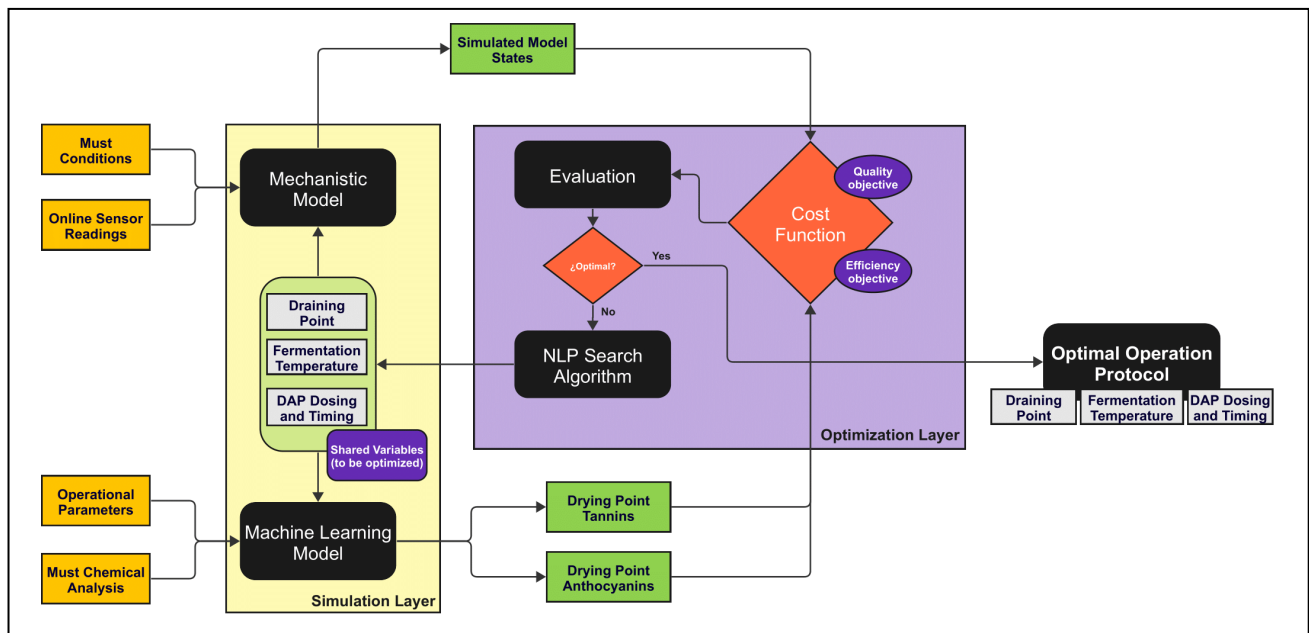
Figure 1. Modular view of the hybrid model for wine alcoholic fermentation and maceration proposed in this article.

On the other hand, the mechanistic model describes the synthesis and consumption of the main metabolites involved in alcoholic fermentation, considering the effects of operational parameters such as temperature and the addition of nitrogen, the limiting substrate.

The hybrid model was designed to optimize the fermentation process by considering variables shared between both models, such as temperature, nutrient dosing, and total fermentation time. Shared variables are used to interconnect predictions between models, allowing for a coordinated approach to the optimization process.

The combination of the mechanistic and data-driven models provided a comprehensive and accurate understanding of the fermentation process, allowing the optimization of key parameters to achieve the desired wine quality.

2.3.2 Modeling and parameter estimation

The multi-objective cost function considers three objectives: quality maximization, process costs minimization and productivity maximization. Quality was quantified as the phenolic composition at the draining point, including tannins and anthocyanins. The total phenolic composition is important for the sensory attributes and health benefits of red wine, contributing to its color, taste, and antioxidant properties [4].

Process cost was quantified as the sum of dynamic diammonium phosphate (DAP) additions throughout fermentation. DAP addition is justified as it is a key source nutrient for yeast growth and fermentation performance, ensuring optimal fermentation outcomes [20].

Productivity is given by the time before reaching the draining point of maceration. The time until draining point is important since it influences the fermentation length and the overall wine quality, achieving a balance

between flavor development and avoiding negative sensory outcomes [19,21].

The multi-objective optimization problem was formulated as follows:

$$\text{minimize } J_1 = -(A(\mathbf{X}) + T(\mathbf{X})) \tag{1a}$$

$$\text{minimize } J_2 = \sum \text{DAP}_{add,i} \tag{1b}$$

$$\text{minimize } J_3 = t_f \tag{1c}$$

subject to:

$$\dot{x} = f(\mathbf{x_i}, \mathbf{y_i}, \mathbf{u_i}, \boldsymbol{\theta}, t_f) \tag{1d}$$

$$A = f(t_f, \mathbf{X}) \tag{1e}$$

$$T = f(t_f, \mathbf{X}) \tag{1f}$$

$$g(\mathbf{x_i}, \mathbf{y_i}, \mathbf{u_i}, \boldsymbol{\theta}) = 0 \tag{1g}$$

where $\mathbf{x_i}$, $\mathbf{y_i}$, $\mathbf{u_i}$, $\boldsymbol{\theta}$, and t_f correspond to state variables, algebraic variables, control variables, parameters, and the process time for the mechanistic model, respectively. *A* and *T* correspond to the data-driven models for anthocyanins and tannins, respectively, and t_f is the process time for maceration that matches with the fermentation time. The dynamic inputs ($\mathbf{u_i}$) correspond to the temperature profile and DAP additions.

X corresponds to the optimized variables included in each recipe and used as input for both mechanistic and/or machine learning models (f_A and f_T for anthocyanin and tannin prediction models, respectively). These variables include Free-K, Oak powder, Tartaric acid, SO2 addition dosing, draining time, and fermentation kinetic model inputs ($\mathbf{u_i}$) that are shared between the models (dynamic temperature and DAP addition).

A single-shooting method with a total of $i = 1...n = 50$ finite elements was used for dynamic optimization [34]. Operational and regulatory constraints ($\mathbf{g_i}$) involved in

industrial winery processing were considered. Specifically, total DAP additions were restricted to a maximum of 40 g/hL, and temperature shifting between simulation steps was restricted to a maximum value of 4 °C.

3 Results and Discussion

3.1 Machine learning models

Machine learning algorithms were trained to predict the concentration of total anthocyanins and total tannins at the draining point of wine alcoholic fermentation. Table 1 shows the results of the training and validation for each machine learning algorithm using the mean absolute error (MAE) as the scoring metric.

For the total anthocyanins concentration predictor, the Gradient Boosting algorithm achieved the best accuracy for both the training and test datasets, with accuracies of 99.99% and 80.86%, respectively.

Regarding the total tannins predictor, most of the algorithms exhibited overfitting, meaning they performed well on the training dataset but had poor accuracy on the test dataset. However, the Support Vector Machine algorithm showed a better balance between predictions for the test dataset (45% accuracy) and lower overfitting (a smaller difference between training and test yields). Thus, the Support Vector Machine algorithm was selected for the tannins predictor.

To further enhance the tannins predictor, it is recommended to include new relevant features related to the extraction process and consider the addition of new observation data. These improvements can help improve the accuracy and generalization ability of the model.

3.2 Mechanistic model for alcoholic fermentation

The multi-criteria approach presented in [30] was used to select, reparametrize, and calibrate the most suitable mechanistic model for simulating wine alcoholic fermentation from a list of two models [32,33]. A bootstrapping cross-calibration scheme was employed using a fermentation kinetics training dataset, where subsets of different combinations of experiments were formed. The squared-error function was defined based on simulated states and measured values for each experiment in a subset. The global error function for parameter calibration was then defined as the sum of errors calculated for all subsets. In our example, bootstrap parameter estimates were obtained using enhanced Scatter Search (eSS) for global search with *fmincon* as the local solver [36].

The model selection process involved evaluating various criteria to ensure robustness and performance in different dimensions. The evaluated criteria included goodness-of-fit, identifiability, and sensitivity. For the goodness-of-fit criterion, the corrected Akaike Information Criterion (AICc) was utilized to assess the model's parsimony and goodness-of-fit to the calibration set. Lower AICc values indicated better model performance. The identifiability criterion measured the overall uncertainty in a model structure and was evaluated using the Mean of the Normalized Confidence Intervals (MNCI). Smaller MNCI values indicated better overall significance of the model structure. Finally, the sensitivity criterion focused on the model's parametric sensitivity, and the global parametric sensitivity score (GSS) represented the cumulative sensitivity across all state variables and parameters. Different multi-criteria decision-making (MCDM) scenarios were explored by assigning different weights to each criterion, and various MCDM methods were applied to select the best model structure for each scenario, balancing among these criteria according to [30].

The alcoholic fermentation models underwent evaluation using independent experimental validation data from two experimental scales. The evaluation process involved the computation of performance indices, including a global performance index (GPI) while analyzing residual normality. The GPI measured the overall performance of each model structure based on adjusted determination coefficients (R_{adj}) for measured model states. Based on the performance indices, the best-overall model structure (BOMS) was determined for each evaluated model. Robustness indicators, as well as GPI

Table 1. Accuracy and average error obtained for each modeling approach in the construction of total tannin and anthocyanin machine learning models.

Model	Average Error / % Accuracy			
	Training tannins	**Test tannins**	**Training anthocyanins**	**Test anthocyanins**
Kernel ridge	178.6 / 37.9%	277.6 / 46.6%	65.3 / 86.1%	129.8 / 76.2%
Support vector machine	87.1 / 78.8%	277.6 / 45%	24 / 95.7%	135.5 / 75.5%
Random forest	75.4 / 74.2	288.9 / 41.5%	35.5 / 92.1%	109.2 / 79.3%
Gradient boosting	0 / 99.9%	290.3 / 44.3%	0 / 99.9%	109.1 / 80.9%
Stochastic gradient descent	217.4 / 24.1%	276.5 / 42.6%	116.8 / 74%	116.6 / 76.6%
K-nearest neighbors	0 / 100%	288 / 42.3%	0 / 100%	115.1 / 77.6%
Gaussian processes (RBF)	0 / 99.9%	435.8 / 35.3%	0 / 99.9%	211.8 / 62%
Gaussian processes (Matern)	0 / 99.9%	335.6 / 45.7%	0 / 99.9%	148.9 / 73.3%
Gaussian processes (ExpSineSquared)	0 / 99.9%	404.6 / 36.4%	0 / 99.9%	197.2 / 65%
Gaussian processes (Rational quadratic)	2.84 / 99.9%	280 / 43.07%	1.49 / 99.9%	115.9 / 77.4%

for original and reparametrized BOMS, are displayed in Table 2. The BOMS exhibited significant sensitivity to calibration, good predictive capacity, and minimal bias. It demonstrated the highest averaged GPI and the highest number of normally distributed uncorrelated residuals among the evaluated model structures.

Table 2. Robustness and goodness-of-fit results achieved through the reparametrized and calibrated mechanistic models.

Description		Regression diagnostics			
		AICc	MNCI	GSS	Av. GPI
Coleman	Original (Lab.)	-220.1	0.53	47.4	0.97
	BOMS (Lab.)	-274.2	0.15	74.5	0.97
	Original (Pilot)	-208.1	0.69	31.5	0.90
	BOMS (Pilot)	-250.7	0.22	37.7	0.87
Zenteno	Original (Lab.)	-185.6	0.35	12.7	0.94
	BOMS (Lab.)	-241.1	0.03	23.3	0.96
	Original (Pilot)	-182.1	0.14	8.6	0.81
	BOMS (Pilot)	-202.7	0.12	16.5	0.85

Overall, the selected and calibrated mechanistic model provided a robust framework for simulating wine alcoholic fermentation, capturing the complexities of the process and yielding accurate predictions. Figure 2 presents a graphical demonstration of this, displaying a comparison between mechanistic modeling approaches used to predict fermentation kinetics over one of the pilot-scale testing experiments and overall displaying a high predictive capacity (averaged GPI of above 0.95 and 0.85 for laboratory and pilot-scale BOMS, respectively).

3.3 Mechanistic model for alcoholic fermentation

The aim to develop and applied the hybrid model (Fig. 1) is to establish a knowledge-based framework for

designing optimal red wine fermentation and maceration recipes. This considers the achievement of objectives related to maceration and fermentation process considering operational and wine quality constraints as well the process productivity.

To explore the above, we simulated maceration and fermentation conditions corresponding to a varietal and premium wines using data obtained from Viña Concha y Toro. These wines differ significantly in grape chemical composition and the fermentation operation management. Therefore, it is expected that the optimal operation conditions for these wines be different. Table 3 shows the initial conditions for varietal and premium wines. The initial phenolic composition of must (DO280, color index, total anthocyanins and tannins) at the filling point was much higher for premium wines than varietal wines. The fermentation initial conditions were the same for both scenarios.

The results from multi-objective dynamic optimization are shown in Fig. 3, obtaining a set of Pareto efficient solutions which trade-off the quality, process cost and productivity objectives. Each point represents a specific fermentation operation condition, considering specific values for the decision variables according to the production objective functions. Analyzing the distribution fermentation operation conditions among the three production objectives, a strong trade-off between total phenolics and draining point time is identified, where, as suggested in literature, shorter macerations lead to lower final polyphenol concentration [19]. Additionally, a smaller effect of DAP additions is identified, with higher additions being related to a slight increase in final phenolics concentration while also displaying an insignificant effect on draining-point time. Another interesting observation we observed from Fig. 3 is the shared behavior among the premium and varietal wines simulated in this study, both displaying the same patterns discussed previously.
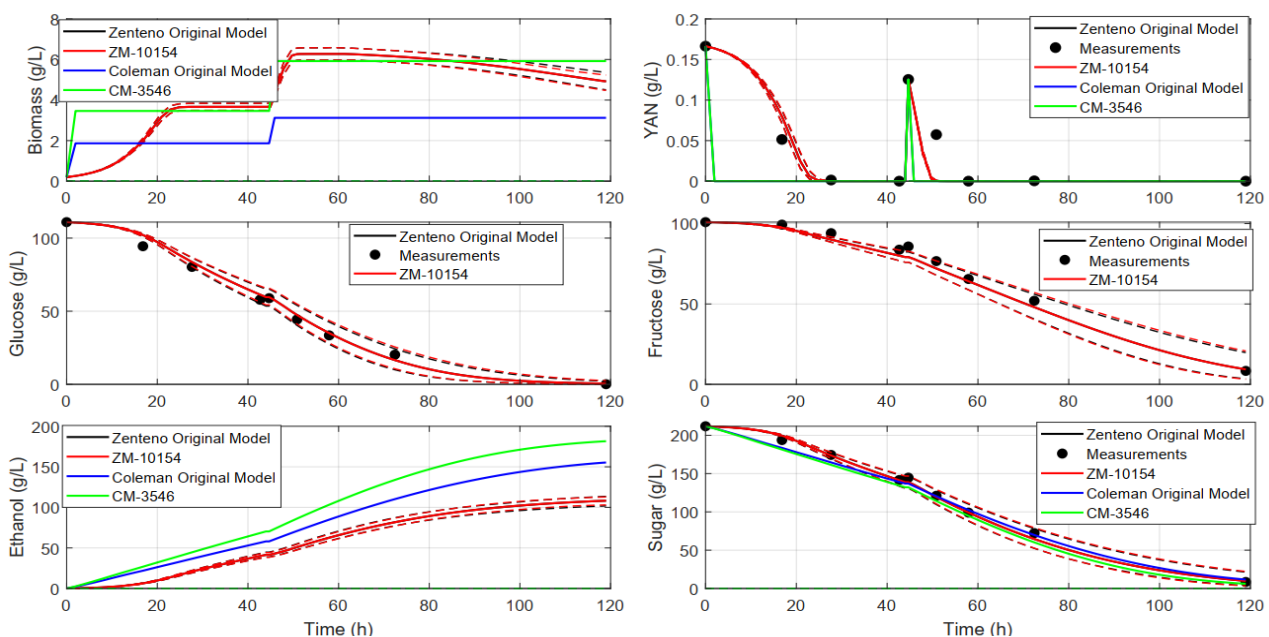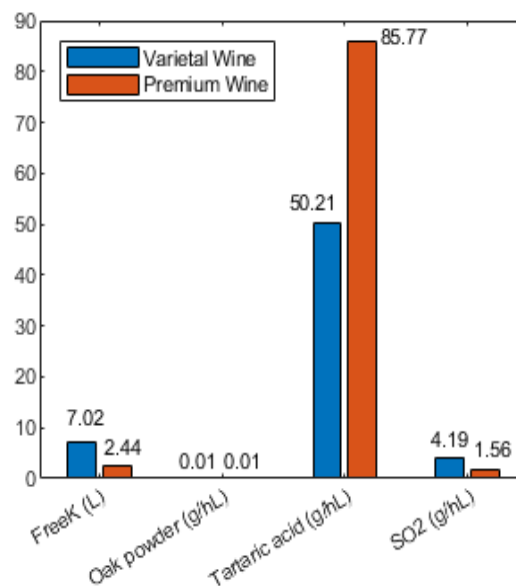

Figure 2. Simulation results obtained through the reparametrized and calibrated mechanistic models in pilot-scale test experiment 1.
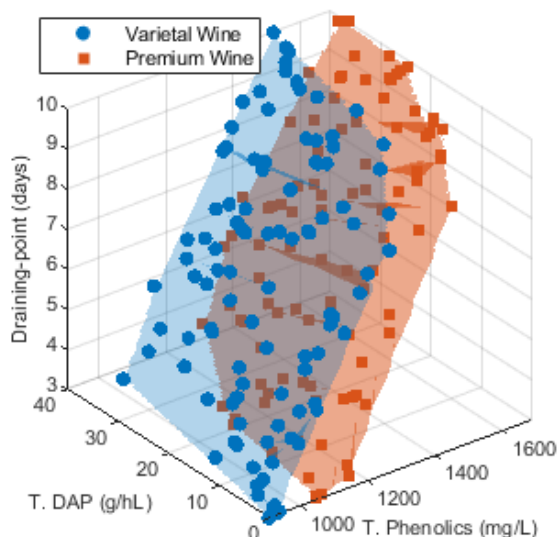
Figure 3. Pareto Front obtained through the hybrid model in the different winery scenarios.

Figure 4. Bar plot of product addition in selected optimal wine recipes predicted through the hybrid model.

For further analysis, a solution of the Pareto-set was selected for both the varietal and premium wines. Figure 4 and Table 3 show the optimal fermentation conditions and evaluated objective function values. Significant differences are observed among varietal and premium wine recipes and their outputs in each process. For example, a lower use of tartaric acid and higher use Free-K are predicted for varietal wine processing, which is adequate in this type of processes given that these products are substitutes among themselves with Free-K being a lower cost option. However, abnormal quantities are observed in Free-K additions, which typically move among the thousand liters in most operations. This displays an opportunity to enhance machine-learning models and optimization, as pH affecting wine acidity should be considered in the quality maximization objective to effectively use this variable.

Objective function values observed for each optimal solution display smaller maceration time with a consequently lower phenolic extraction and higher DAP additions for the varietal wine. Conversely, premium wine obtained longer maceration with higher phenolic extraction and lower DAP additions. This matches industrial protocols observed in varietal wineries, as grapes processed in these typically arrive with a high concentration of tannins related with negative sensorial attributes. Therefore, high fermentation rates and short macerations are seeked by winemakers in these scenario. Again, this is the reverse situation of premium wineries, where extended macerations with controlled extraction procedures (e.g., pumping-over and air-mixing protocols) are searched. This relates to the desire for extracting valuable phenolic compounds present in higher concentration in premium grapes pomace, at a reduced

Table 3. Optimal recipes and objective function values from selected example recipes for each winery scenario.

Maceration conditions	*D280*	*Color Index*	*Filling-point anthocyanins*	*Filling-point tannins*	*Volume*
Varietal	9.29	1.47	34.13 mg/L	40 mg/L	40124 L
Premium	18.71	4.55	108.37 mg/L	240 mg/L	32200 L
Fermentation conditions	*Biomass*	*YAN*	*Sugar*	*Density*	
Varietal/Premium	0.23 g/L	120 mg/L	216.5 g/L	1098.4 g/L	
Pareto optimal recipe	*Free-K*	*Oak Powder*	*Tartaric acid*	*SO2*	
Varietal	7.02 L	0.01 g/hL	50.21 g/hL	4.19 g/hL	
Premium	2.44 L	0.01 g/hL	85.77 g/hL	1.56 g/hL	
Objective function values	*Total Phenolics*	*Total DAP addition*	*Draining point*		
Varietal	1098.51 mg/L	28.81 g/hL	4.56 days		
Premium	1582.85 mg/L	12.46 g/hL	9.81 days		

risk of spoiling wine because of the lower concentration of off-flavor compounds.

Figure 5 displays fermentation kinetics for each selected wine recipe, which are observed to vary significantly in function of the temperature and DAP addition protocols. Similar timing is observed among the processes analyzed for each winery, where most DAP is added during the start of fermentation and near reaching density 1050 g/L. Moreover, DAP dosing varies significantly between the varietal and premium winery (28.81 vs 12.46 g/hL, respectively). This coincides with the argument that varietal wines are seeked to be fermented faster to minimize extraction of off-flavor related phenolics during maceration. In this instance, optimization is not coupled with drying requirements of the fermented wines. However, winemakers tend to seek to match fermentation drying-point with maceration the draining-point in order to simplify further processing.
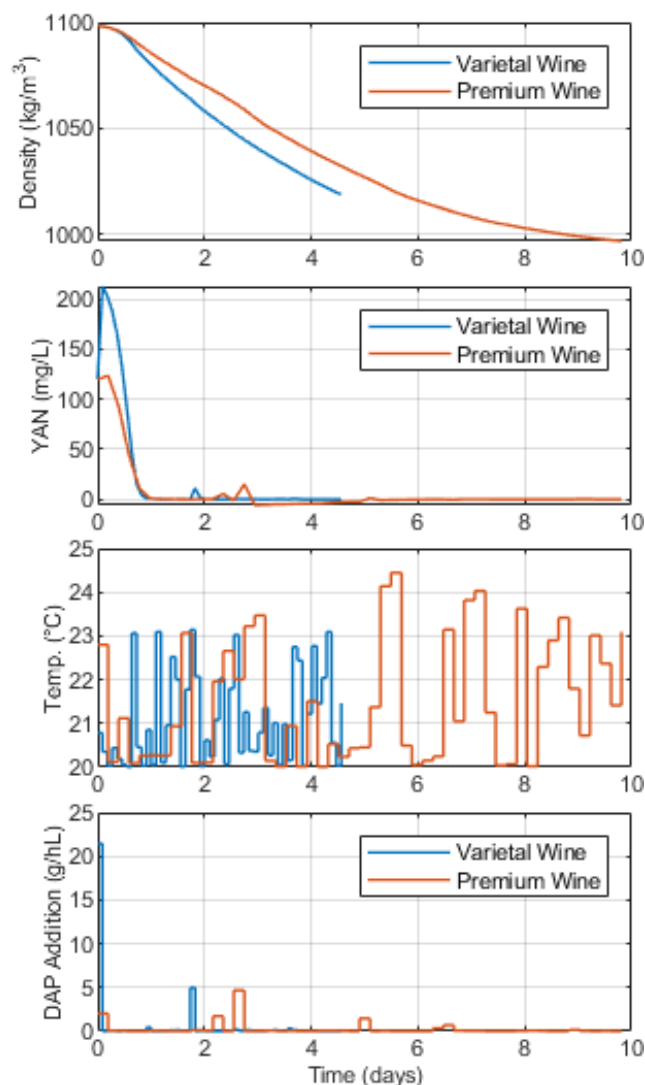
Figure 5. Fermentation kinetics and optimal inputs related to selected optimal recipes for each winery scenario.

Overall, the hybrid model and its optimal wine recipes have significant practical implications for winemakers in both varietal and premium wineries. They present

potential to enhance wine quality by considering multiple objectives, such as total phenolic composition, DAP addition, and fermentation duration. This enables winemakers to achieve desired wine characteristics while efficiently managing resources. This modeling approach serves as a valuable tool for decision-making processes, allowing winemakers to set multiple objectives and optimize the wine production process accordingly. By visualizing trade-offs through the generated Pareto front, winemakers can make informed decisions based on data-driven insights and select the recipe that best aligns with their priorities during the harvest season.

Future research directions can focus on incorporating additional process variables, such as yeast strains and grape maturity parameters, to enhance the comprehensive understanding of the winemaking process. Refining the machine learning models by improving accuracy and robustness would further enhance the reliability of the optimal wine recipes. Additionally, validating the optimized recipes through experimental trials would provide practical confirmation of their performance in terms of sensory attributes, chemical composition, and consumer acceptance. By addressing these future research directions, the hybrid model can continue to evolve and offer valuable insights, advancements, and practical applications in optimizing red wine fermentation and maceration for the wine industry and winemakers.

4 Conclusion

In this study, we developed a hybrid model for red winemaking that integrates and combines mechanistic and data-driven approaches to optimize fermentation recipes and improve the phenolic composition of red wine.

A comprehensive database of multiple-scale *Cabernet Sauvignon* fermentation from 2017-2021 harvest season was collected. Robust fermentation kinetic model was generated with high performance and predictors for anthocyanins and tannins with 45% and 80.9% accuracy, respectively.

By incorporating common variables for fermentation and macerations such as temperature, time, and mixture volume, the hybrid model enabled to optimize fermentation operation conditions that simultaneously improved wine quality, process cost and productivity. The models predictions and recipe prescriptions present the potential to be generated in advance for each fermentation tank, providing valuable information for wine decision-making in industrial cellars.

The integration of Industry 4.0 technologies, such as IoT, analytics, AI, and machine learning, has revolutionized the winemaking industry, enabling the development of smart factories and improving production processes, efficiency, and wine quality. Our hybrid model exemplifies the power of these technologies in enhancing red wine production and decision-making.

Future advancements in this field will focus on further improving the accuracy and accessibility of these tools, making them more widely available to wineries of all scales. The combination of mechanistic and data-driven models provides a promising approach for optimizing red

wine fermentation and extraction processes, leading to consistent, predictable, and cost-effective production of high-quality red wine. Overall, the hybrid model contributes to the ongoing digital transformation of the winemaking industry and paves the way for a more sustainable, efficient, and profitable future for red wine production.

References

1. F. Yang, S. Gu, Complex Intell. Syst. **7**, 1311 (2021)
2. V. Giovenzana, S. Baroffio, R. Beghi, A. Casson, A. Pampuri, A. Tugnolo, D.D. Filippi, R. Guidetti, LII (2021)
3. L.F. Casassa, J.F. Harbertson, Annu. Rev. Food Sci. Technol. **5**, 83 (2014)
4. J. Unterkofler, R.A. Muhlack, D.W. Jeffery, Appl. Microbiol. Biotechnol. **10411** 104, 4737 (2020)
5. C.A. Aceves Lara, V. Athès, P. Buche, G. Della Valle, V. Farines, F. Fonseca, V. Guillard, K. Kansou, M. Kristiawan, V. Monclus et al., Innov. Food Sci. Emerg. Technol. **46**, 54 (2018)
6. K.L. Sacchi, L.F. Bisson, D.O. Adams, Am. J. Enol. Vitic. **48**, 197 (2005)
7. L.A. Lerno, S. Panprivech, R. Ponangi, L. Hearne, T. Blair, A. Oberholster, D.E. Block, Am. J. Enol. Vitic. **69**, 295 (2018)
8. P.C. Setford, D.W. Jeffery, P.R. Grbin, R.A. Muhlack, Molecules **23**, 1 (2018)
9. R.A. Tindal, D.W. Jeffery, R.A. Muhlack, Aust. J. Grape Wine Res. **27**, 219 (2021)
10. P.C. Setford, D.W. Jeffery, P.R. Grbin, R.A. Muhlack, Trends Food Sci. Technol. **69**, 106 (2017)
11. L. Lerno, M. Reichwager, R. Ponangi, L. Hearne, D.E. Block, A. Oberholster, Am. J. Enol. Vitic. **66**, 444 (2015)
12. K.V. Miller, A. Oberholster, D.E. Block, Biotechnol. Bioeng. **116**, 781 (2019)
13. P.A. Smith, J.M. Mcrae, K.A. Bindon, Aust. J. Grape Wine Res. **21**, 601 (2015)
14. E. Henrichs, T. Noack, C. Krupitzer, A. María, P. Piedrahita, M.A. Salem, J. Stolz (2022)
15. I.A. Udugama, W. Kelton, C. Bayer, Digit. Chem. Eng. **7**, 100087 (2023)
A. Moser, C. Appl, S. Brüning, V.C. Hass, Digital Twins: Tools and Concepts for Smart Biomanufacturing 133-180 (2021)
16. J.R. Mouret, E. Aguera, M. Perez, V. Farines, J.M. Sablayrolles, Fermentation **7**, 1 (2021)
17. D. Henriques, R. Minebois, S.N. Mendoza, L.G. Macías, R. Pérez-Torrado, E. Barrio, B. Teusink, A. Querol, E. Balsa-Canto, mSystems **6** (2021)
18. P.C. Setford, D.W. Jeffery, P.R. Grbin, R.A. Muhlack, Food Bioprod. Process. **131**, 217 (2022)
19. K.V. Miller, D.E. Block, J. Food Eng. **273**, 109783 (2020)
20. R. Luna, B.M. Lima, J. Cuevas-valenzuela, J.E. Normey-rico, J.R. Pérez-correa, 1-14 (2021)
21. C. Schenk, V. Schulz, A. Rosch, C. von Wallbrunn, Food and Bioproducts Processing **103**, 131 (2017)
22. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, Communications of the ACM **39**, 27 (1996)
23. W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus, AI magazine **13**, 57 (1992)
24. E.B. Schwarze, J.A.S. Barriga-Bravo, Análisis químico del vino (Ediciones Universidad Católica de Chile, 2000)
25. AWRI, AWRI Standard Methods. Determination of tannins in grapes and red wine using the MCP (methyl cellulose precipitable) tannin assay (AWRI Standard Methods, 2000), www.crcv.com.au
26. J.M. Fontana, M. Farooq, E. Sazonov, Estimation of feature importance for food intake detection based on Random Forests classification, in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE, 2013), 6756-6759
27. R. Caruana, A. Niculescu-Mizil, An empirical comparison of supervised learning algorithms, in Proceedings of the 23rd international conference on Machine learning (2006), 161-168
28. G.C. Cawley, N.L. Talbot, The Journal of Machine Learning Research 11, 2079 (2010)
29. C. Torrealba, Ph.D. thesis, Pontificia Universidad Catolica de Chile (Chile) (2021)
30. B.J. Sánchez, D.C. Soto, H. Jorquera, C.A. Gelmi, J.R. Pérez-Correa, Ind. Eng. Chem. Res. (2014)
31. M.I. Zenteno, J.R. Pérez-Correa, C.A. Gelmi, E. Agosin, J. Food Eng. (2010)
32. M.C. Coleman, R. Fish, D.E. Block, Appl. Environ. Microbiol. (2007)
33. Dynamic Optimization Methods with Embedded DAE Solvers, chap. 9, 251–286
34. J. Blank, K. Deb, IEEE Access **8**, 89497 (2020)
35. J.A. Egea, M. Rodríguez-Fernández, J.R. Banga, R. Marti, Journal of Global Optimization **37**, 481 (2007)