Modelling the Fermentation Process in Winemaking using Temperature and Specific Gravity

Ivana Kovačević, Mihaela Orić, Ivana Hartmann Tolić, Emmanuel Karlo Nyarko
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2b, 31000 Osijek, Croatia
ivana.kovacevic@ferit.hr, mihaela.oric@student.ferit.hr, ivana.hartmann@ferit.hr, karlo.nyarko@ferit.hr

***Abstract**. This article examines the fermentation process in winemaking. The authors analyse fermentation data from a local winery in Croatia and propose two preliminary models to determine the optimality of ongoing fermentation processes using measurements of specific gravity and temperature. The data include measurements of temperature and specific gravity taken during successful fermentations of three different white wines. The models developed are a multivariate linear regression model and an artificial neural network model. The performance of each model is evaluated using metrics such as R-squared, Mean Squared Error, Root Mean Squared Error, and Maximum Absolute Error. The results show that both models accurately represent the fermentation process, with the artificial neural network model slightly outperforming the multivariate linear regression model.*

Keywords. Fermentation, Temperature control, Specific gravity, Modelling

1 Introduction

Traditional fermented foods are usually produced by spontaneous fermentation, which is dependent on climatic conditions. The winemaking process, like most of the other food production processes, has several distinct stages of production, namely: harvesting, crushing and pressing, fermentation, aging and maturation, clarification and stabilization, bottling and, finally, distribution and consumption. These stages in wine production involve a careful combination of art and science, with each step contributing to the final quality, characteristics, and enjoyment of the wine. Fermentation is a key step in winemaking, where yeast converts sugars in the must into alcohol, carbon dioxide, and other compounds. Temperature and specific gravity are vital parameters for controlling and monitoring fermentation. Temperature influences yeast metabolism, while specific gravity provides insights into sugar concentration and potential alcohol production. Maintaining an optimal fermentation temperature is crucial for yeast activity and the overall success of the fermentation process. Temperature influences yeast growth rate, sugar utilization, and the production of aroma and flavour compounds. High temperatures can lead to excessive yeast metabolism, resulting in the production of undesirable off-flavours and off-odours. Conversely, low temperatures can slow fermentation or cause yeast dormancy. Winemakers employ various techniques, such as cooling jackets, temperature-controlled tanks, or refrigeration systems, to regulate and maintain the desired fermentation temperature range.

Specific gravity is a valuable parameter for assessing sugar content and potential alcohol levels in the must. It is typically measured using hydrometers or refractometers. Specific gravity readings are taken before and during fermentation, providing insights into the progress and completeness of fermentation. Initially, the specific gravity is high due to the sugar content in the must. As yeast consumes sugar and produces alcohol, the specific gravity gradually decreases. Monitoring specific gravity allows winemakers to estimate the alcohol potential of the resulting wine and determine when fermentation has finished.

The authors of this article are involved in a project to develop an expert system to help winemakers in the production of wine. As mentioned earlier, one of the steps in wine production is fermentation. In this paper, the authors analyse the data recorded by a local winery in Croatia during the successful fermentation of three different white wines. Based on this data, the authors propose two preliminary models that could be used to determine whether the ongoing fermentation process is optimal or not.

2 Methodology in related research

In recent years, machine learning (ML) techniques have gained significant popularity and have been applied in various fields to detect and classify patterns in complex data structures. ML techniques are applied in various fields such as computer vision, natural language processing, healthcare, finance, recommender systems and many others. The choice of ML method to be used depends on the type of data, availability of labelled data, and the problem or task at hand.

Wine is a fascinating and sophisticated product with special characteristics that distinguish it from other products. Accordingly, the analysis, modelling and control of each stage in wine production has been researched exhaustively in an effort to simplify the control of each stage.

In (Tardaguila et al., 2021) an overview of smart applications and digital technologies that are currently applied in viticulture is provided. The test methods used to assess wine quality are demanding and varied. The advice of a wine expert is valuable, but also expensive and subjective. For this reason, various machine learning-based methods for predicting wine quality have been proposed (Chiu et al., 2021, Lee et al., 2021).

In (Román et al., 2011), the authors propose ANNs for effectively recognizing nonlinear patterns in wine fermentation processes and accurately predicting fermentation problems. They explored different sub-cases by varying the predictor variables and the time of fermentation. The predicted variables were total sugar, alcohol, glycerol, density, organic acids, and nitrogen compounds, and the time of fermentation was examined at three different durations. The results highlight the potential of ANNs as a valuable tool for monitoring and optimizing wine production processes. In (Urtubia et al., 2021) two cases were studied for the prediction of problematic wine fermentations. In the first case, the SVM algorithm was used to analyse different groups of chemical variables, including amino acids, saturated fatty acids, unsaturated fatty acids, organic acids, and fermentation control variables. In the second case, individual chemical variables, namely density, YAN (yeast assimilable nitrogen), Brix (sugar content), and acidity, were studied as predictors of problematic wine fermentation using SVM. The integration of artificial intelligence (AI) and machine learning (ML) techniques in fermentation bioprocesses is proposed in (Florea et al., 2022). They combine the fields of computer engineering and food engineering to demonstrate the importance of digitalization in modern industrial processes. The research involves the development and implementation of a software application that models a fully configurable 3-layer feed-forward multilayer perceptron neural network. This neural network, trained with experimental data, is able to predict the evolution of quantities characteristic of the alcoholic fermentation process in white wines. The study of the volatile organic compounds (VOCs) profile of pomelo wine and the development of an index to predict the degree of fermentation led to the improvement of the quality of pomelo wine in industrial production (Wei et al., 2022). Gas chromatography-mass spectrometry was used to analyse the changes in VOC content during the different fermentation periods. Principal component analysis, cluster analysis and partial least squares (PLS) regression were used to evaluate the variations of VOC content in the different fermentation phases. The PLS models identified the α-phellandrene/geraniol ratio as a potential index for determining the degree of fermentation of pomelo wine. Multivariate analysis techniques were also used to create an index to predict the degree of fermentation of pomelo wine under controlled laboratory conditions. The proposed index provides winemakers and wine chemists with valuable insights into the VOC composition and profile of pomelo wine throughout the fermentation process.

Modelling cold fermentations with non-conventional Saccharomyces species for establishing an experimental-modelling pipeline is proposed in (Henriques et al., 2018). The experimental pipeline includes conducting small-scale micro-vinifications and wine fermentations under controlled conditions. During these fermentations, the authors monitored growth rate and critical extracellular metabolites such as glucose, fructose, ethanol, glycerol, and acetic acid. They compared several candidate models that account for different biomass growth, transport, and inhibition mechanisms described in the literature. Developing a minimal model with the fewest parameters necessary for structural identifiability ensures that the model can be fit uniquely to the data while improving practical identifiability. For the most successful models, a combined modelling strategy is used to increase their robustness and minimize predictive uncertainty. The results are presented quantitatively and they use an ensemble of models to make robust predictions for processing conditions, such as initial inoculation and temperature, in order to optimize alcohol and glycerol production.

In (Nelson et al., 2022), the authors combine in *situ* measurements of density based on differential pressure measurements with a wine kinetic model and parameter estimation routine to predict the progression of a 1500-L wine fermentation.

3 Wine Fermentation Data

The data analysed are fermentation data recorded by a local winery in Croatia over a period of eleven years: 2010-2020. The winery uses measurements of temperature and specific gravity to implement control strategies during fermentation. Temperature control involves monitoring and adjusting the fermentation temperature to create optimal conditions for yeast activity. This can be achieved by cooling or heating, depending on the desired wine style and yeast strain used. The specific gravity measurement is used to help decide when to add nutrients, adjust sugar levels or decide when to stop fermentation. The data provided are recorded series of fermentations where each data sample included two measurements of the wine must, namely the specific gravity (γ) in the unit Oechsle [°Oe] and the temperature T [°C]. The Oechsle unit [°Oe] is a unit of measurement commonly used in winemaking to estimate the potential alcohol content of grapes or grape must. It is based on the specific gravity of the grape juice or must and measures the soluble solids, mainly sugars, in the juice, which directly correlate with the potential alcohol that can be produced during fermentation. The Oechsle scale usually ranges from 50°Oe to 200°Oe, with higher values indicating a higher sugar concentration and potential alcohol content. It is important to note that the Oechsle unit is mainly used in German winemaking, especially for classifying German wines based on maturity and sugar content. Other scales and units may be used in other winegrowing regions, such as Brix in the United States or Baumé in France. All data provided refer to successful fermentations (series) of three different white wines, i.e. white grape varieties: Graševina, Chardonnay and Pinot Gris. An overview of the distribution of the measured data can be found in Fig. 1 and Table 1.

Table 1. Total number of data samples per grape variety

	Graševina	Chardonnay	Pinot Gris
Total no. of fermentation processes	20	22	22
Total no. of data samples	397	393	329

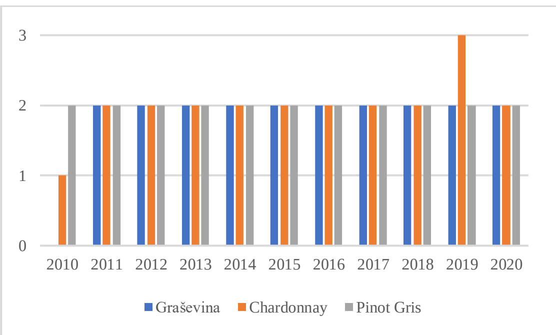

Figure 1. Number of fermentation processes recorded per year and per grape variety

A total of 64 different fermentation processes were recorded, comprising 1119 data samples. The minimum and maximum readings of the data samples are shown in Table 2.

It can be seen in Table 2 that the measured values for specific gravity ranged from 26 °Oe to 104 °Oe, while the temperature values ranged from 12.4 °C to 24 °C. It can also be seen that the interval between the measurements of the data samples is between 1 and 5 days. Of the 1119 data samples, 1073 measurements were taken every 24 hours, 35 after 48 hours, 10 after 72 hours and 1 after 5 days (Table 3).

Table 2. Minimum and maximum values of the data measured during fermentations.

Parameter Measured	Min. value	Max. value
Specific Gravity, γ [°Oe]	26	104
Temperature T [°C]	12.5	24
Interval between measurements [days]	1	5
No. of data samples per fermentation process	7	34

Table 3. Distribution of measurement interval

Interval between measurements [days]	No. of data samples
1	1073
2	35
3	10
5	1

The distribution of the temperature recorded is given in Fig. 2.

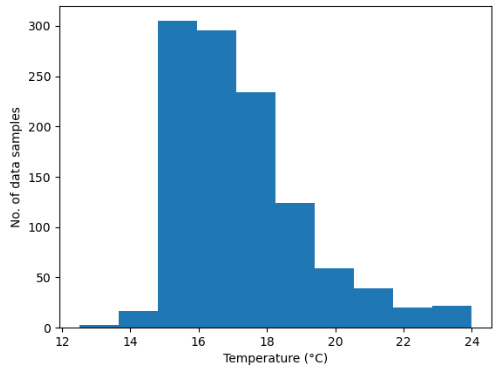

Figure 2. Histogram of temperature measured

Fig.2 shows that most of the measured data samples have a temperature between 15 °C and 18 °C, indicating that the winery controls the fermentation process mainly within this temperature range. An example of a fermentation process is shown in Fig. 3. It can be seen that during the fermentation process, the specific gravity value decreases with time, indicating a decrease in the sugar content, and an increase in the alcohol content.

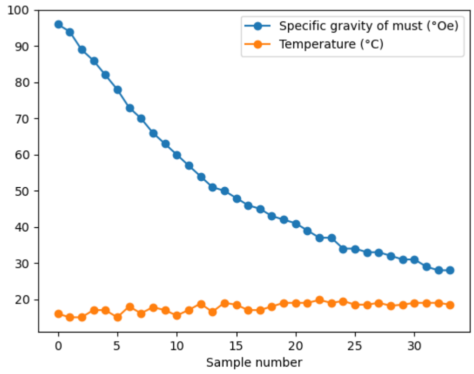

Figure 3. Example of a fermentation process showing the specific gravity and corresponding temperature.

Based on the recorded data, the authors attempt to model the fermentation process. For this purpose, the data is first divided into two subsets: the training set and the test set. The training set is used to determine the parameters of the model, while the fit of the model is determined using the test set. It should be noted that the data was split without taking the grape variety into account. However, it is imperative that a complete fermentation process or series is included in either the training set or the test set. 22% of the complete fermentation processes were randomly selected as the test set and the rest were used as the training set. The division of the measured data into the training set and the test set is given in Table 4.

Table 4. Division of the measured data into train set and test set

	Total	Train set	Test set
Total no. of fermentation processes	64	50	14
Total no. of data samples	1119	868	251

4. Model evaluation

All models considered in this paper are evaluated using the following evaluation metrics: R-squared (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Maximum Absolute Error (MAE).
R^2 represents the proportion of the variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1, with higher values indicating better fit. It is defined by eq. 1:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - y_M)^2} \quad (1)$$

where:
y_i – represents the measured or actual value,
f_i – represents the predicted value,
y_M – represents the mean of the measured value.

MSE (eq. 2) measures the mean squared difference between predicted and actual values. It provides a general measure of the accuracy of the model, with lower values indicating better performance.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - f_i)^2 \quad (2)$$

RMSE (eq. 3) is the square root of MSE and provides a more interpretable metric in the same units as the dependent variable.

$$RMSE = \sqrt{MSE} \quad (3)$$

MAE (eq. 4) is the maximum absolute difference between predicted and actual values. It gives an indication of the largest individual prediction error of the model.

$$MAE = max(|y_i - f_i|) \quad (4)$$

5 Fermentation Modelling

Based on the specific gravity measured at time *k*, and the temperature at time *k*+1, the model should provide an estimate of the specific gravity at time *k*+1 for a given fermentation process. Such a model would help a wine producer to control the fermentation process by adjusting the temperature based on the current specific gravity value and the desired specific gravity value after one day. The model to be determined, *F*, is defined by the following equation:

$$\gamma(k+1) = F(\gamma(k), T(k+1), k) \quad (5)$$

where:
k – sample number,
$\gamma(k)$ – specific gravity measured at time k,
$T(k+1)$ – temperature at time $k+1$.
The model to be determined, *F*, thus has 3 input variables (features) and one output variable.

As this is a preliminary study, two models are analysed and compared: a multivariate linear regression model and an artificial neural network model. The reason for selecting these models at this stage was because of their key differences: Multivariate linear regression is simpler and suitable for linear relationships, while artificial neural networks are more complex and versatile and can capture both linear and non-linear patterns in the data. All modelling was done in Python using the machine learning library scikit-learn (https://scikit-learn.org/stable/index.html). Before training the models, all data were preprocessed by performing data standardization with normal distribution, i.e. all input parameters were transformed to have a mean of 0 and a standard deviation of 1.

5.1 Multivariate linear regression model

Based on the training data, the specific gravity at time *k*+1 for a given fermentation process is given by the model:

$$\gamma(k+1) = 50.138 + 19.880\gamma(k+1) - 0.520T(k+1) + 1.504k \quad (6)$$

The performance of the linear regression model on the test dataset is shown in Table 5 and Fig. 4. Analysing Table 5, it can be concluded that the multivariate linear regression model has a high R^2 value of 0.9854 and a low RMSE value of 2.3902.

Table 5. Performance metrics of the multivariate linear regression model on the test dataset

R^2	MSE	RMSE	MAE
0.9854	5.7129	2.3902	7.2288

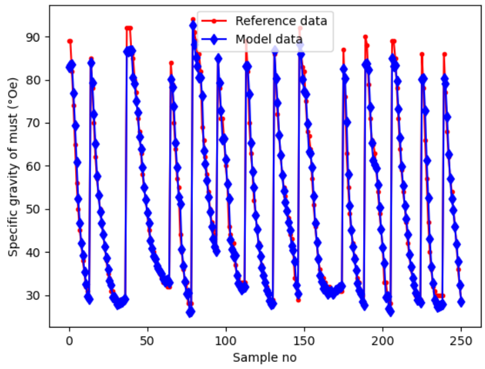

Figure 4. Comparison of the output values of the multivariate linear regression model with the actual values

Looking at Fig. 4, it can be concluded that the multivariate linear regression model models the fermentation process relatively well. However, it should be noted that it does not estimate the initial fermentation values very well.

5.2 Artificial Neural Network Model

A simple feed-forward neural network model with a single hidden layer was used. The structure of the neural network model was determined using the Grid Search Cross-Validation "GridSearchCV" function in the scikit-learn library. It was assumed that the neural network model has only one hidden layer. The number of neurons in the hidden layer was varied from 2 to 15, and three activation functions of this layer were also considered: logistic, tanh and relu. The linear activation function was used in all cases for the neuron in the output layer. The best artificial neural network model had three neurons in the hidden layer with the logistic-sigmoid activation function. The performance of the selected artificial neural network model on the test dataset is shown in Table 6 and Fig. 5.

Table 6. Performance metrics of the artificial neural network model on the test dataset

R^2	MSE	RMSE	MAE
0.9885	4.502	2.122	6.6514

Table 6 shows that the artificial neural network model has a high R^2 value of 0.9885 and a low RMSE value of 2.122.

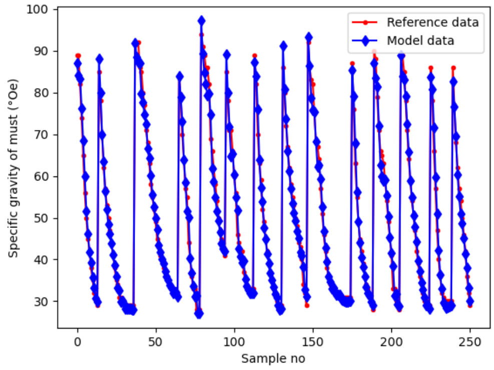

Figure 5. Comparison of the output values of the artificial neural network model with the actual values

Looking at Fig. 5, it can be concluded that the artificial neural network model represents the fermentation process quite well.

6 Discussion

The comparison of the two models shows that the artificial neural network slightly outperforms the multivariate linear regression model. The R^2 value increases by 0,3 % and the values of MAE and RMSE decrease by 8 % and 11 % respectively. The slightly better performance of the artificial network model could be due to the fact that the relationship between the data in the proposed model has some non-linearity that cannot be captured by the multivariate linear regression model.

7. Conclusions and Future Work

The aim of this work was to develop models for assessing the fermentation process of white wines based on specific gravity and temperature measurements. The models demonstrate good performance, with the artificial neural network model having a slight advantage over the multivariate linear regression model in terms of accuracy. The multivariate linear regression model showed high accuracy in estimating the specific gravity at the next time step, as indicated by the high R^2 value of 0.9854 and low RMSE value of 2.3902. However, the model had difficulty in accurately estimating the values for the initial fermentation. In contrast, the artificial neural network model with a single hidden layer and three neurons provided a good representation of the fermentation process. These results suggest that the models have potential for predicting fermentation outcomes in winemaking and provide winemakers with a valuable tool for optimizing the process and ensuring high-quality wine production. Further research could explore the generalizability of the models to other wine varieties and the investigation of additional factors that influence the fermentation process.

Acknowledgments

We thank Krauthaker Vineyards and Winery for providing the fermentation data used in this research. This work results from implementing research activities on the project *Development of an expert system for food production and processing management (ESIA-Expert System for Intelligent Agriculture KK.01.1.1.07.0036)* funded by the European Regional Development Fund.

References

Chiu, T. H.-Y., Wu, C.-W., & Chen C.-H., (2021). A Hybrid Wine Classification Model for Quality Prediction. In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, Proceedings, Part IV. Springer-Verlag, Berlin, Heidelberg, 430–438. https://doi.org/10.1007/978-3-030-68799-1_31

Florea, A., Sipos, A., & Stoisor, M. C. (2022). Applying AI Tools for Modeling, Predicting and Managing the White Wine Fermentation Process. Fermentation, Vol. 8, Page 137, 8(4), 137. https://doi.org/10.3390/FERMENTATION8040137

Henriques, D., Alonso-del-Real, J., Querol, A., & Balsa-Canto, E. (2018). Saccharomyces cerevisiae and S. kudriavzevii synthetic wine fermentation performance dissected by predictive modeling. Frontiers in Microbiology, 9(FEB), 317309. https://doi.org/10.3389/FMICB.2018.00088/BIBTEX

Lee, C. K., Law, K. M., & Ip, A. W. (2021). A Rule-Based Quality Analytics System for the Global Wine Industry. Journal of Global Information Management (JGIM), 29(3), 256-273. http://doi.org/10.4018/JGIM.20210501.oa1

Nelson, J., Boulton, R., & Knoesen, A. (2022) Automated Density Measurement With Real-Time Predictive Modeling of Wine Fermentations," IEEE Transactions on Instrumentation and Measurement, 71(1-7), 3509607. https://doi.org/10.1109/TIM.2022.3162289.

Román, R. C., Hernández, O. G., & Urtubia, U. A. (2011). Prediction of problematic wine fermentations using artificial neural networks. Bioprocess and Biosystems Engineering, 34(9), 1057–1065. https://doi.org/10.1007/S00449-011-0557-4/METRICS

Tardaguila J., Stoll M., Gutiérrez S., Proffitt T., & Diago M. P. (2021). Smart applications and digital technologies in viticulture: A review, Smart Agricultural Technology, 1,100005. https://doi.org/10.1016/j.atech.2021.100005.

Urtubia, A., León, R., & Vargas, M. (2021). Identification of chemical markers to detect abnormal wine fermentation using support vector machines. Computers & Chemical Engineering, 145, 107158. https://doi.org/10.1016/J.COMPCHEMENG.2020.107158

Wei, Q., Liu, G., Zhang, C., Sun, J., & Zhang, Y. (2022). Identification of characteristic volatile compounds and prediction of fermentation degree of pomelo wine using partial least squares regression. LWT, 154, 112830. https://doi.org/10.1016/J.LWT.2021.112830