Logistic Regression Modeling of the Growth Limit of Alicyclobacillus Acidoterrestris CRA7152 in Apple Juice

Loftsson, Valentin Oliver — valentin.loftsson@epfl.ch

Introduction

Microbiological contamination in processed acid fruit juices has been the subject of microbiological interest and research in the past few decades. Controlling such contamination is an important factor in the preservation of packed fruit juice products. To minimize contamination risk, it is essential to consider changes in the microbiology of the product over the entire production cycle, from cultivation and harvesting to preparation, distribution, and storage.

Peña *et al.* (2011) studied the contamination in apple juice caused by an acidophilic sporeformer bacterium, *Alicyclobacillus acidoterrestris* CRA7152, and developed a model to predict its growth probability. The bacterium causes "off-flavor" in fruit juices from certain chemical compounds. Several studies have shown that *Alicyclobacillus acidoterrestris* grows at temperatures between 25 to 60°C (Previdi and Vicini 1995) with pH varying from 3.0 to 5.5 (Eguchi 1997). The use of nisin, a chemical that is used as a food preservative, has proven effective as a minimum inhibitory concentration method to control the germination of the bacterium's spores. However, the effectiveness of nisin depends also on the temperature, pH, and soluble solids concentration.

Evaluating the bacterium's growth/nongrowth interface becomes cumbersome when multiple factors are involved. Logistic regression is a powerful tool for probabilistic microbial modeling, given that sufficient information is available about the product characteristics and conditions. It has been used frequently in similar studies. With logistic regression, it is possible to describe the growth probability and establish critical limits for the predictive variables that determine the condition of the apple juice.

In this paper, we reproduce the statistical analyses performed by Peña *et al.* on the original dataset. The main purpose of the original study is to "evaluate the growth responses of *Alicyclobacillus acidoterrestris* in apple juice with different pH values, concentrations of soluble solids and nisin subjected to different incubation temperatures". We begin with an exploratory data analysis and then perform a logistic regression analysis to determine and evaluate an optimal predictive model.

Exploratory Data Analysis

The dataset includes 74 samples of apple juice assays including a binary growth response (0=No Growth, 1=Growth) evaluated after 16 days of incubation. The dataset consists of two collections of 37 assays, tested in duplicate. The assays were carefully created by the researchers to ensure a balanced dataset with a varied combination of factors that determine the juice condition.

The effect of four factors was studied as summarized in Tables 1-4. The frequency table of the growth response reveals that we have a semi-balanced dataset, with several more negative examples and growth occurring in 26 out of 74 assays ($\sim$ 35%). For each factor, four discrete values were selected for testing after careful experimentation and analysis, and considering accumulated knowledge from previous studies. The tables reveal some interesting characteristics about individual factors independent of interaction effects: Growth was completely inhibited at pH 3.5 (Table 1) which is the minimum pH of the juice. It was also fully inhibited at the maximum nisin concentration 70 IU/ml (Table 2) and at the maximum soluble solids concentration of 19 °Brix (Table 4), which marks the concentration level above which no growth of *A. acidoterrestris* occurs (Splittstoesser, Churey, and Lee 1994).

Table 1: pH

	No Growth	Growth	Sum
3.5	18	0	18
4	12	8	20
5	8	10	18
5.5	10	8	18
Sum	**48**	**26**	**74**

Table 2: Nisin concentration (IU/ml)

	No Growth	Growth	Sum
0	8	14	22
30	12	6	18
50	10	6	16
70	18	0	18
Sum	**48**	**26**	**74**

Table 3: Temperature (°C)

	No Growth	Growth	Sum
25	16	2	18
35	10	8	18
43	10	10	20
50	12	6	18
Sum	**48**	**26**	**74**

Table 4: Soluble solids conc. (°Brix)

	No Growth	Growth	Sum
11	16	8	24
13	8	8	16
15	6	10	16
19	18	0	18
Sum	**48**	**26**	**74**

Model Fitting and Selection

Logistic regression is used as a method for model fitting as it is suitable for a binary response. Here, we use the logit function as the link transformation for the generalized model:

$$logit(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (1)$$

Essentially, this means that the linear transformation outputs the logit transformed probability of a positive response ($Y = Growth$). To obtain the probability, p, we must apply the inverse logit function:

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} \quad (2)$$

The parameters β_k are estimated by maximum likelihood.

To capture important characteristics of relationships between the explanatory variables, we start with a model that includes all main and interaction effects. Note that ^2 denotes the inclusion of all interaction effects:

```
growth ~ (ph + nisin + temperature + brix)^2
```

A likelihood-ratio test (LRT) was used to perform backward variable selection. The LRT compares two models where one is a sub-model of the other, by computing the LR test statistic, denoted here as λ:

$$\lambda = -2 \ln \frac{L(m_1)}{L(m_2)} = 2(\ln L(m_2) - \ln L(m_1)) \quad (3)$$

where m_1 corresponds to the more restrictive model and m_2 is the less restrictive model. $L(*)$ denotes the likelihood of the data samples given a model. The statistic is Chi-squared distributed.

The LRT essentially tests the following null and alternative hypotheses:

- H: $\beta_k = 0$ for all variables k that differ between the models
- A: $\beta_k \neq 0$

Table 5: LRT test on interaction effects

	LRT	P value
ph:nisin	0.00	0.9999
ph:temperature	0.00	0.9997
ph:brix	34.81	0.0000
nisin:temperature	0.00	0.9998
nisin:brix	38.78	0.0000
temperature:brix	32.12	0.0000

For each factor, we compare the full model described above against the sub-model with that factor removed. Table 5 summarizes the results. Clearly, the terms `ph:nisin`, `ph:temperature`, and `nisin:temperature` are not significant as their P-values indicate ($P > 0.05$). As a result, we do not reject the null hypothesis that the coefficients for these terms are zero and they are thus dropped from the model.

Final Model

The remaining terms of the final model are all the main effects and interactions of `brix` with each variable. The coefficients and corresponding P-values are presented in Table 6. Below is the model in mathematical terms:

$$\begin{aligned} logit(p(Growth)) = &-260.7453 \\ &+ 43.0019 \cdot \boldsymbol{pH} - 1.9031 \cdot \boldsymbol{nisin} + 3.1852 \cdot \boldsymbol{temp} + 13.4122 \cdot \boldsymbol{brix} \\ &- 2.2149 \cdot \boldsymbol{pH} \cdot \boldsymbol{brix} + 0.1002 \cdot \boldsymbol{nisin} \cdot \boldsymbol{brix} - 0.1688 \cdot \boldsymbol{temp} \cdot \boldsymbol{brix} \end{aligned} \tag{4}$$

Table 6: Coefficients of the final model for A. Acidoterrestris growth in apple juice

	Coefficient	P value
(Intercept)	-260.7453	0.0446
ph	43.0019	0.0497
nisin	-1.9031	0.0460
temperature	3.1852	0.0382
brix	13.4122	0.0492
ph:brix	-2.2149	0.0524
nisin:brix	0.1002	0.0463
temperature:brix	-0.1688	0.0425

Model Evaluation

The model concordance or classification accuracy was used to assess the model's performance. The resulting confusion matrix is presented in Table 7. The overall accuracy is 97.3%, meaning only 2.7% or 2 samples out of 74 were incorrectly classified. The final model and accuracy results match the findings of Peña *et al.* (2011).

Table 7: Confusion matrix summarizing prediction results from applying the final model on the data samples

Real response	Predicted response	
	No Growth	**Growth**
No Growth	48	0
Growth	2	24

Peña *et al.* (2011) confirmed the validity of the model through experiments on new samples. These experiments testified to the usefulness of nisin in preventing the growth of the bacterium and thereby increasing the shelf life of apple juice, especially under conditions of high temperature.

We visualize the behavior of the final model on surface graphs (Figures 1-2). The graphs show the predicted growth probability of *A. acidoterrestris* in apple juice as a function of two main effects. The graphs can be explored interactively online by clicking the links above the figure captions.

Figures 1.a and 1.b show the probability as a function of soluble solids concentration and temperature, given 0 IU/ml of nisin and pH 3.7 and 4.5, respectively. For pH 4.5 (Figure 1.b), growth was inhibited only in high concentrations of soluble solids (Brix). Interestingly, when pH drops down to 3.7 (Figure 1.a) no growth occurs for any value of Brix up to a temperature of 30°C. Above 30°C, the probability is close to zero only above 18°Brix.

Figures 2.a and 2.b show the probability as a function of soluble solids concentration and nisin concentration, given pH 4.0 and temperature 45°C and 30°C, respectively. For 45°C (Figure 2.a) growth can be inhibited by adding 40-70 IU/ml of nisin. When the temperature is decreased to 30°C (Figure 2.b) we observe that the model predicts zero or small probability of growth for any Brix level and low nisin levels. At zero nisin, the probability decreases as the Brix level increases.

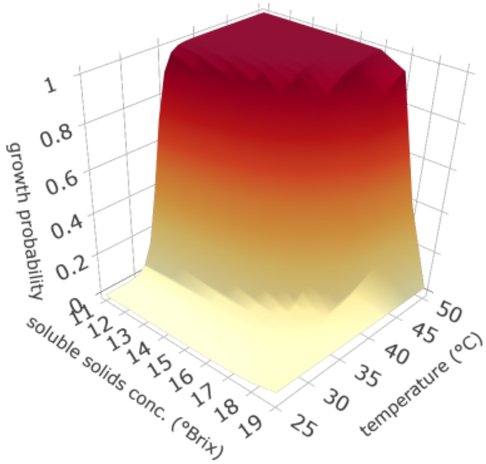
The graphs show the effect and interactions of the variables and reveal critical values that are important to establish minimum inhibitory limits. Such limits are necessary for minimizing the use of inhibitory chemicals in the product and depend on the pH and temperature levels. Peña *et al.* compute and discuss critical boundaries in more detail.

Conclusion

In this paper, we set out to reproduce the work of Peña *et al.* (2011) to model the growth probability of *Alicyclobacillus acidoterrestris* in apple juice depending on pH, incubation temperature, soluble solids, and nisin concentration. We successfully fit the data using logistic regression and found an optimal model using stepwise model selection. The results match the findings of the original study. Since the final model includes both main and interaction terms, the model coefficients are difficult to interpret. Nevertheless, the model's prediction accuracy has been verified through accuracy testing, visualization, and experimentation. The model provides a beneficial tool for apple juice production since it can be used to minimize the risk of bacterial spoilage in apple juice.

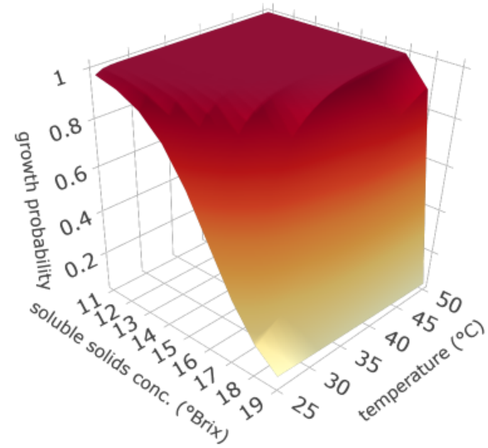
View interactive plot online

Figure 1.a: Growth probability of CRA7152 in apple juice with **nisin 0 IU/ml and pH = 3.7**



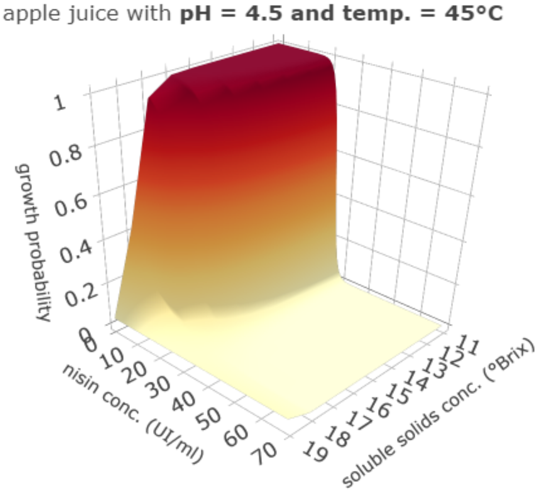
View interactive plot online

Figure 1.b: Growth probability of CRA7152 in apple juice with **nisin 0 IU/ml and pH = 4.5**



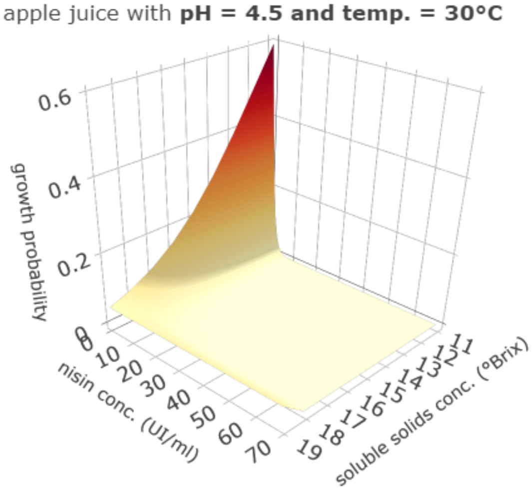
View interactive plot online

Figure 2.a: Growth probability of CRA7152 in apple juice with **pH = 4.5 and temp. = 45°C**



View interactive plot online

Figure 2.b: Growth probability of CRA7152 in apple juice with **pH = 4.5 and temp. = 30°C**



Software

The reproduction was performed in the R programming language (R Core Team 2019). Tables were created with kableExtra (Zhu 2019). Interactive figures were created with Plotly (Sievert 2020).

References

Eguchi, S. Y. et al. 1997. "Detection of Acidothermophilic Bacilli in Industrialized Fruit Juices, Fruit Processing." *FRUIT PROCESSING*, FRUIT processing, 7 (9): 350–53. https://www.tib.eu/de/suchen/id/BLSE%3ARN032043094.

Peña, Wilmer, Pilar Massaguer, A. D.G. ZUÑIGA, and Sergio Saraiva. 2011. "Modeling the Growth Limit of Alicyclobacillus Acidoterrestris Cra7152 in Apple Juice: Effect of pH, Brix, Temperature and Nisin Concentration." *Journal of Food Processing and Preservation* 35 (August). https://doi.org/10.1111/j.1745-4549.2010.00496.x.

Previdi, Colla, M. P., and E. Vicini. 1995. "Characterization of Alicyclobacillus, a Spore-Forming Thermophilic Acidophilic Bacterium." *The Indian Journal of Soil Conservation*, no. 70: 128–32.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman; Hall/CRC. https://plotly-r.com.

Splittstoesser, DF, JJ Churey, and CY Lee. 1994. "Growth Characteristics of Aciduric Sporeforming Bacilli Isolated from Fruit Juices." *Journal of Food Protection* 57 (12): 1080—1083. https://doi.org/10.4315/0362-028x-57.12.1080.

Zhu, Hao. 2019. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.