Data Driven Shelf Life Prediction*

M. Ceelen†

(Dated: July 12, 2019)

Artificial intelligence is used in this research to predict the shelf life of strawberries. The prediction of shelf life is based on temperature measurements from the moment a package of strawberries is harvested till the moment this same package is bought by a customer in a local PLUS supermarket. The strawberries are harvested in the south of Spain, near Huelva and distributed to local PLUS Supermarkets near Rotterdam. After the packages with strawberries, including temperature loggers, arrive at the supermarket shelf, the packages are moved to a shelf life room for visual inspection. During this daily inspection, the actual shelf life of the strawberries is determined by a classified inspector. The combination of the actual shelf life and the temperature profile through the supply chain is used to train, validate and test different machine learning algorithms. The most reliable shelf life prediction algorithm is the Exponential Gaussian Process Regression Algorithm, with the smallest confidence interval and an average deviation of 14.1 %. To conclude, the possible improvements in the supply chain based on shelf life prediction, like traceability, food date labeling and quality grading are evaluated.

I. INTRODUCTION

In the supply chain of perishable food products, like fruit and vegetables, large losses are incurred between the grower and the consumer. Given the ever growing population, and the limited land resources the food losses need to be reduced. Improving the efficiency of the food supply chain could help to bring down food waste [1, 2].

Improvement in global supply chain management has the potential to reduce food waste as well as the costs of the consumer, given that the cost of food waste is increasing towards the consumer end of the supply chain. A company participating in global supply chain management of fruit and vegetables is The Greenery. The Greenery distributes fruit and vegetables around the world from growers to supermarkets. A major costs item is the product returns of dissatisfied customers. A significant reason for these returns is the quality of the product. The shelf life period of the products is not as expected. The cost made by the Greenery for each product based on incorrect quality determination or deterioration in the supply chain.

The highest costs for incorrect quality determination and deterioration are made in the strawberry supply chain. The strawberry is a valuable product which degrading process highly depends on the surrounding conditions. Problems in the supply chain do result in quality changes or deterioration of the product.

Food waste in the strawberry supply chain can be reduced by improving traceability [3], food date labelling [4] and quality grading [5]. The major challenge of this research is to find out how reliable artificial intelligence can predict the shelf life of as strawberry and how it can be implemented to optimize traceability, food date labelling and quality grading.

* This work is supported by The Greenery
† Faculty of Mechanical, Maritime and Materials Engineering; m.m.ceelen@student.tudelft.nl

The outline of this paper is as follows. In the next section the research scope is evaluated. 'Shelf Life Prediction' explains the variables which influence the shelf life of a strawberry. The artificial intelligence algorithms used to predict the shelf life of strawberries and the selection criteria for these algorithms are evaluated in "Predicting Algorithms". The research mythology is explained in "Research Methodology". Finally, the results, conclusion and recommendations are given.

II. LITERATURE REVIEW

It is a big challenge to reduce the losses during transportation of perishable products. The combination of reasons for deterioration of perishable products are always unique. The limiting factor for predicting the expiring date is the available hardware to measure the conditional data [6]. The ability to measure a product through the whole supply chain and the processing models which are currently pre-programmed or rule-based [7].

Currently done research [8] integrated a continuous monitoring system through a food supply chain. The research compares different measurements systems with each other and recommends the active RFID tags. This active RFID tags send measured data via a gateway to an online database.

A temperature monitoring system has been developed by the project FRISBEE [9]. This web based platform can predict the remaining shelf life based on the conditional data in the previous supply chain by using various scenarios in Monte Carlo simulation.

This research differs from previous work, by the focus on the development of a real-time self learning model considering the dynamic data of measurements on product level through the whole supply chain. The aim of this research is to determine the accuracy of the prediction of shelf life of a self learning model and to evaluate the improvement on the current supply chain.

III. RESEARCH SCOPE

A. Strawberry Supply Chain

During the time period of this research, strawberries were imported from Spain. The grower which participated in this research is growing his strawberries in the south of Spain near Huelva. At this point, the temperature loggers are added to the strawberries. The strawberries are transported to The Netherlands by truck and distributed at The Greenery located in Breda. The strawberries are packed and shipped to the distribution centre of PLUS, in Barendrecht. From this location the strawberries are order picked and shipped to local supermarkets. At the local supermarket the trackers are collected and returned to the distribution centre of The Greenery. Where the measured data is analysed.

B. Assumptions

The growing process of the strawberry is not inside the measuring scope. The initial condition of the strawberries is assumed to be equal. To make sure the used strawberries are equal as possible, one grower and one variety of strawberries is selected. In this way the used strawberries have had the same treatment during the growing process.

IV. SHELF LIFE PREDICTION

The shelf-life of products is not a static parameter but is a highly dynamic variable. The product shelf life is affected by conditions in the environment surrounding the product. The environmental conditions are the inputs for the shelf life prediction process. The prediction of shelf life (output) depends on the measured parameters (input). The Artificial Intelligence transfers the incoming data into a shelf life prediction.

A. Input Parameters

The most important environmental factors described by Pol Tijskens [10] are temperature, relative humidity and the concentrations of oxygen, carbon dioxide and ethylene. Out of these environmental factors the temperature is selected to be measured as input parameter. The temperature is selected based on the fact that it is the most influencing parameter for shelf life and a limited research budget. To test the potential of shelf life prediction by Artificial Intelligence this parameter is considered sufficient enough. In the current supply chain the temperature is not measured. Which means, to measure this parameter on product level a measurement system needs to be selected.

TABLE I. AHRMA Temperature Sensor Specification

AHRMA Temperature Sensor Specifications	
Accuracy	+/- 0.5 degrees Celsius
Mearsurement Interval	10 minutes
Total Measurement Period	+/- 5 days
Signal Range	80 meter

B. Measurement System

To generate a lot of data samples on product level with a small budget, the selected data logger for this research is supplied by AHRMA. The loggers fit in the 500 gram strawberry package and transmit their data wire-less via a gateway to the online database. The data is easily accessible by a online dashboard.

C. Feature Selection

Good features are informative for a machine learning algorithm. In this case, the output of the machine learning algorithm is the predicted shelf life. An informative feature has a relation to the shelf life of the strawberry. Based on literature study [10] and interviews, temperature, humidity and the atmosphere are evaluated to be implemented as features. Due to scope of this research and budget limitations, the temperature is chosen to be measured on product level through the supply chain. The following relations between the shelf life and the temperature measurements are selected:

- **Fluctuating temperatures** are measured when the strawberries move through different stages of the supply chain [11]. The storage facility, the loading dock and the truck have different temperatures. These conditional changes are visual in the temperature profile.
- **Average temperature**, since the temperature is the main factor affecting all (bio)chemical processes through its effects on activation enthalpy and entropy of the under-laying reactions, a relation between the average temperature and the shelf life is assumed [12].
- **Reduction to storage temperature**, the strawberry is harvested and transported to a cool storage. In this cool storage the temperature is decreased to the preferred 3 degrees Celsius. A fast temperature reduction has a positive effect on the shelf life of a strawberry.
- **Time spend in a particular temperature zone**, the temperature zone around 3 degrees Celsius is preferred for strawberry distribution [13], according to the literature. When temperatures are

under or above this temperature it will have an effect on the shelf life.

V. PREDICTING ALGORITHMS

The prediction of shelf life is based on artificial intelligence. Artificial intelligence is a name for a wide range of algorithms which all have different purposes. Besides machine learning, the natural language processing, planning and reasoning models are evaluated for predicting shelf life of a strawberry. The ability to learn from and adapt a model based on the data rather than explicit programming is the reason for selecting machine learning for the prediction of shelf life of a strawberry. The learning process of machine learning is based on features and labels, the labels are the actual shelf life determined by an inspector in the shelf life room. The use of labels made this research based on supervised learning. Supervised learning can be divided classification, regression and neural network models. The inputs are continuous and the outputs are continuous. This means that classification is not suitable but the regression and neural network models are applicable for predicting shelf life of a strawberry.

VI. METHODOLOGY

This research methodology is divided in four categories, namely system analysis, studying, testing and use of shelf life prediction in the fruit and vegetable supply chain. These are the steps used to build a black box model which can predict self life of a strawberry.

A. System Analysis

For a period of two weeks, six temperature loggers entered the supply chain of strawberries. Each temperature logger had his own path through the supply chain but all ended up at the shelf life room. One logger went directly to the shelf life room at the distribution center of The Greenery. Another logger went to the distribution center of PLUS and returned to the shelf life room. The remaining four temperature loggers went through the whole supply chain from the moment the strawberries are harvested at the grower, along the distribution center of The Greenery, to a distribution centre of the PLUS which distribute the strawberries over 5 predefined local supermarkets and finally to the shelf life room of The Greenery. In the shelf life room the product is visual inspected on daily basis till the strawberry is expired. At that moment the temperature logger is removed from the package of strawberries. The measured temperature data by the temperature logger is send wireless to an online data base. Before data is used for this research it is processed. The process starts with data import from the temperature sensors. This time-temperature profile is analysed and noise is removed. The data is used to generated features.

B. Machine Learning Selection

To find the most accurate regression learning algorithm for predicting the shelf life, 19 different regression algorithms are trained by the Matlab regression learner application. These algorithms are evaluated and finally tested on predicting the shelf life based on the available data samples. To use these data samples in an optimal way, cross validation is applied. The next section will evaluated the method of cross-validation and the selection of learning algorithm.

Besides the regression learning algorithm, a neural network is evaluated to use for shelf life prediction of strawberries. Three different optimization algorithms, Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient are validated and tested by the Matlab Neural Network fitting application.

C. Testing

The selected machine learning algorithms are tested on new data. This data is not used for training or validation of the algorithm. By testing the selected algorithms on new data, the predicting accuracy of the system will be evaluated.

D. Use of Shelf Life Prediction

For the last stage of this research, potential implementations of self life prediction are evaluated. The implementation of shelf life prediction in objective quality control, customer assignment and dynamic pricing are described.

VII. RESULTS

Based on the measurements, the machine learning algorithms are trained, validated and tested. The results of the regression and neural network models will be evaluated and finally a selection of these models will be tested on new data samples which are not used for training or validation.

A. Measurements Of Temperature Loggers

In total 55 data samples are used in this research. The measured temperature profile is shown in figure 1. For all data samples features and labels are determined. With

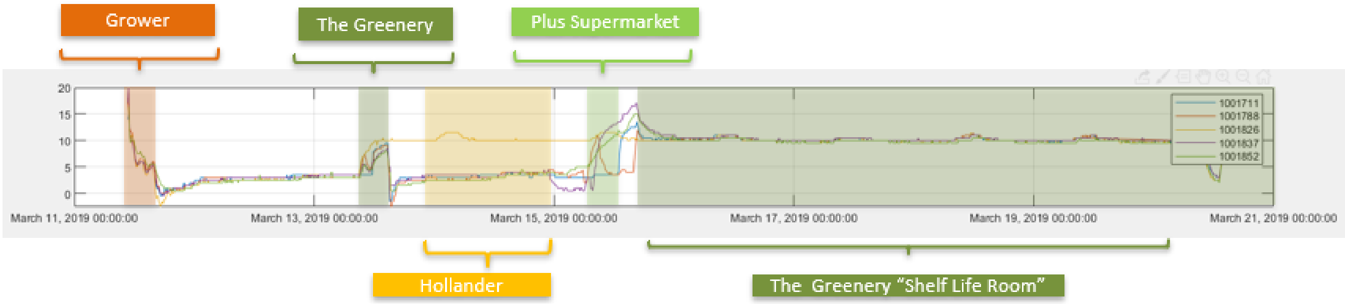

FIG. 1. Temperature profile overview, actors in the strawberry supply chain are selected.

the help of the Matlab regression learner and neural network fitting , the features and labels are used to predict the shelf life of strawberries based on the actual shelf life determined in the shelf life room.

B. Algorithm Selection

The 19 different regression algorithms are compared to each other based on the RMSE value. The algorithm with the lowest RMSE values is favourable above the other algorithms. The regression algorithms with the lowest RMSE values are Linear, Fine Gaussian, Medium Gaussian and Coarse Gaussian Support Vector Machine, and Gaussian Process Regression Exponential. These selected regression algorithms will be tested on new data samples.

Besides the regression learner algorithms, the accuracy of neural networks with different kind of optimisation algorithms is evaluated on accuracy. Like the regression models, the same data samples are used to train a neural network with different optimisation algorithms. The neural network with the three different optimisation algorithms is compared to the regression models based on new data samples in tabel II.

C. Test Results

The selected machine learning algorithms are tested on ten new data samples. These data samples are not used for training or validation of the machine learning algorithms. The shelf life prediction generated by each algorithm is given in table II.

The difference between the actual shelf life and the predicted shelf life is given by the deviation. The exponential Gaussian process regression algorithm has the least deviation percentage.

A way to compare the performance of the model on the training and test data is to use a type of average loss function. The average loss function used in this research is the MSE value. The MSE value of the model on the test data is lower than the MSE value obtained on the training data. This means that there is no overfitting in the prediction of shelf life of the strawberries.

Based on the deviations between predicted shelf life and the actual shelf life, the interval of the deviation with 95 percent confidence is determined. The interval boundaries are given on each side of the mean. The distance between the interval boundaries is given in the last column. The machine learning algorithm with the smallest boundary is the Levenberg-Marquardt optimised Neural Network.

When the algorithm is used in practise, the reliability of the system is most important. This means that a narrow 95% confidence interval is important and the data spreading should be small. This is the reason why the Gaussian Process Regression Exponential algorithm is favourable to be used in practise.

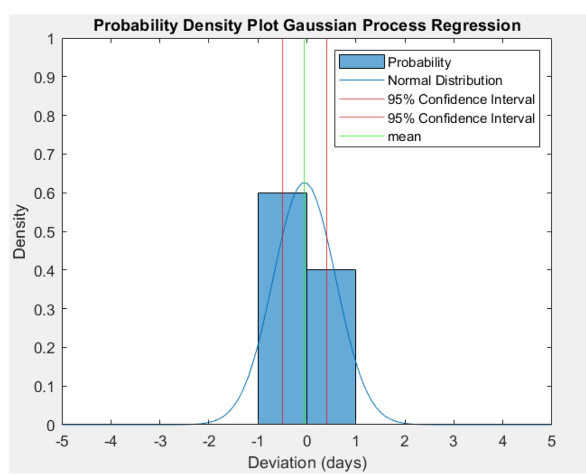

FIG. 2. Predicted Density Plot Gaussian Process Regression Exponential.

TABLE II. Machine learning algorithms compared to each other based on ten new data samples

labels	Linear SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM	Exponential SVM	-LM-Neural Network	-BR-Neural Network	-SCG-Neural Network	Actual Shelf life
1001781	4.8	4.5	5.0	4.4	4.9	4.8	4.3	4.7	4
1001848	4.9	4.3	4.7	4.5	4.3	5.1	4.3	4.7	5
1001768	4.9	4.3	5.0	4.5	4.5	4.8	4.3	4.7	4
1002034	4.3	3.7	4.1	4.2	3.5	3.8	4.3	4.2	4
1002003	5.1	4.3	4.6	4.5	4.3	4.4	4.4	3.8	5
1001811	3.4	4.3	3.8	3.9	4.3	4.1	4.2	4.0	5
1002002	4.1	3.8	4.1	4.2	3.6	3.7	4.2	3.9	4
1002033	3.6	3.8	3.1	4.0	3.6	3.5	4.1	3.5	3
1002038	3.0	4.3	2.8	3.8	3.9	2.9	4.0	2.9	4
1001770	3.2	4.1	2.9	3.9	3.6	3.3	4.1	3.3	3
Deviation (%)									
Max	31.8	36.3	31.9	31.9	21.9	27.2	37.8	27.4	
Min	2.5	5.8	4.2	4.2	1.9	1.6	0.9	1.4	
Average	14.1	14.5	12.7	14.0	14.1	13.2	12.8	13.2	
Mean Squared Error (days)									
Training	1.6	1.57	1.35	1.49	1.57	1.85	1.15	0.88	
Test	0.55	0.39	0.52	0.41	0.37	0.41	0.42	0.51	
95% confidence interval boundaries (days):									
Left	-0.53	-0.43	-0.63	-0,39	-0.51	-0.54	-0.36	-0.66	
Right	0.59	0.51	0.45	0.57	0.41	0.42	0.60	0.40	
Interval	1.12	0.94	1.08	0.95	0.91	0.97	0.96	1.06	

D. Shelf Life Prediction In Strawberry Supply Chain

The introduction of this research described three major problems in the fruit and vegetable supply chain. In this section the potential improvements in traceability, food date labelling and human quality grading will be evaluated. The improvements are illustrated in figure 3.

The measurement of temperature through the supply chain helps to monitor the conditions in the supply chain and the conditions of the strawberries to reduce deterioration in the supply chain. Besides deterioration, improved traceability makes it easier to find the problems in the supply chain which will reduce the costs for insurance of products.

The shelf life, is one of the quality indications assigned by the inspector. Based on the results of this research the selected algorithm is able to predict the shelf life of a strawberry (in the scope of this research) by half a day, with an accuracy of 95 percent. Based on this accuracy, strawberries could be delivered to customers with a margin of one day. A customer who is ordering products with a shelf life of four days, gets products with a predicted shelf life of five days. To make sure the shelf life of the product is sufficient.

The ability to assign accurate shelf life predictions to products can help finding the right retailer or trade company for the right product. Knowing the resulting shelf life gives the retailer more insight into the product they have to sell to their customers. Food date labelling based on shelf life prediction can help to improve customer assignment and dynamic pricing.

VIII. CONCLUSION

The results of this research show the potential of artificial intelligence for a data driven shelf life prediction. This research is based on one influencing parameter for shelf life, the temperature. The features based on the measured temperature data are used to train, validate and test regularly used algorithms. The most reliable algorithms is the algorithm with the smallest confidence interval, the exponential Gaussian Process Regression algorithm with an average deviation of 14.1 %.

The measurement of temperature through the supply chain to predict shelf life will improve the traceability of products or in this case strawberries. The traceability makes it possible to find and remove problem sources in the supply chain. The costs for incorrect quality determination can be reduced by objective quality assignment by supporting the quality inspector with shelf life predictions based on measured data on product level. The thrid potential improvement is food date labelling based on product level measurements. The ability to assign accurate shelf life predictions to products can help to find the right retailer or trade company. These retailers and trade companies have a better idea of the shelf life of the

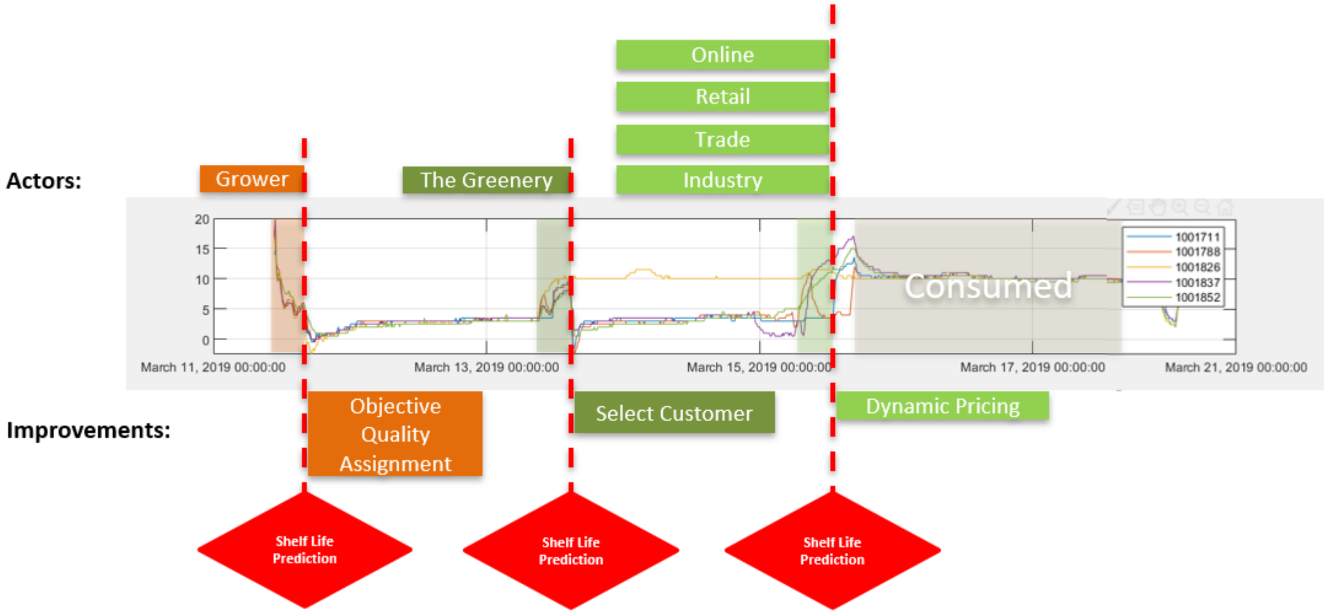

FIG. 3. Temperature profile overview, actors in the strawberry supply chain are selected.

product and can participate on this, this participation will result in a reduction of food waste in the whole fruit and vegetable supply chain.

IX. RECOMMENDATIONS

Before this systems is able to predict the shelf life of a variety of strawberries from different growers and different areas more research is needed. The most important recommendations for further research are given:

- **Measure more parameters**, parameters like humidity and atmosphere measurements are not included in this research. More input parameters will help the system to get more informative features.

- **Increase accuracy with more samples**, by executing measurements on a bigger scale, the correlation plots will be more clear. Besides the correlation plots, the influence of data samples on the accuracy of shelf lifer prediction is analysed. For the regression learning algorithms, more samples did not result in a more accurate prediction. In case of the neural networks, it was clear that the accuracy is very diverse based on the data samples available in this research. The relative small number of data can be the reason for this, more data samples will reduce the weights of noise and out of bound data samples.

- **Start condition of strawberries**, at the moment the logger is implemented into the supply chain is considered mutual for all strawberries. Unless the circumstances are made as mutual as possible, there is not a clear correlation found between the features and the shelf life. The initial conditions can be the reason for this error, for example weather conditions or soil characteristics can make big differences for the final the shelf life of a strawberry, these conditions are not in the scope of this research.

- **Feature Selection**, the current feature selection is based on literature and expert knowledge. A next step would be to find more feature, but also find the optimal combination of features to train, validate and test the machine learning algorithms. An optimization of the features has potential to improve the accuracy of the shelf life prediction.

[1] U. M. Carthy, I. Uysal, R. Badia-melis, S. Mercier, C. O. Donnell, and A. Ktenioudaki, Trends in Food Science & Technology **77**, 11 (2018).
[2] I. Congress, *FAO Food waste report* (2011).
[3] P. Olsen and M. Borit, Trends in Food Science and Technology, Elsevier Ltd **29**, 142 (2013).
[4] C. Priefer, J. Jörissen, and K. R. Bräutigam, Resources, Conservation and Recycling **109**, 155 (2016).
[5] S. Mahajan, A. Das, and H. K. Sardana, Image acquisition techniques for assessment of legume quality, Trends

in Food Science and Technology **42**, 116 (2015).

[6] R. Badia-Melis, U. Mc Carthy, and I. Uysal, Data estimation methods for predicting temperatures of fruit in refrigerated containers, Biosystems Engineering **151**, 261 (2016).

[7] L. Wang, S. K. Kwok, and W. H. Ip, A radio frequency identification and sensor-based system for the transportation of food, Journal of Food Engineering **101**, 120 (2010).

[8] G. Alfian, J. Rhee, H. Ahn, J. Lee, U. Farooq, M. F. Ijaz, and M. A. Syaekhoni, Integration of RFID, wireless sensor networks, and data mining in an e-pedigree food traceability system, Journal of Food Engineering **212**, 65 (2017).

[9] S. G. Gwanpua, P. Verboven, D. Leducq, T. Brown, B. E. Verlinden, E. Bekele, W. Aregawi, J. Evans, A. Foster, S. Duret, H. M. Hoang, S. Van Der Sluis, E. Wissink, L. J. Hendriksen, P. Taoukis, E. Gogou, V. Stahl, M. El Jabri, J. F. Le Page, I. Claussen, E. Indergård, B. M. Nicolai, G. Alvarez, and A. H. Geeraerd, The FRISBEE tool, a software for optimising the trade-off between food quality, energy use, and global warming impact of cold chains, Journal of Food Engineering **148**, 2 (2015).

[10] P. Tijskens, *Discovering the Future : Modelling Quality Matters* (2004) p. 233.

[11] M. C. Nunes, J. P. Emond, and J. K. Brecht, Quality of strawberries as affected by temperature abuse during ground, in-flight and retail handling operations, Acta Horticulturae **604**, 239 (2003).

[12] M. M. Aung and Y. S. Chang, Temperature management for the quality assurance of a perishable food supply chain, Food Control **40**, 198 (2014).

[13] M. L. A. T. M. Hertog, H. A. M. Boerrigter, G. J. P. M. Van Den Boogaard, L. M. M. Tijskens, and A. C. R. Van Schaik, Predicting keeping quality of strawberries (cv. 'Elsanta') packed under modified atmospheres: An integrated model approach, Postharvest Biology and Technology **15**, 1 (1999).

[14] N. Ndraha, H. I. Hsiao, J. Vlajic, M. F. Yang, and H. T. V. Lin, Time-temperature abuse in the food cold chain: Review of issues, challenges, and recommendations, Food Control **89**, 12 (2018).