Imperial College London
Department of Earth Science and Engineering
MSc in Applied Computational Science and Engineering

Independent Research Project
Final Report

Predictive Modelling of Shelf Life for FMCG Snack Products

by
Jiaxuan Cheng

Email: jc2024@ic.ac.uk
GitHub username: esemsc-jc2024
Repository: https://github.com/ese-ada-lovelace-2024/irp-jc2024

Supervisors:
Dr. Jorge Avalos-Patino
Dr. Catalina Aguilar-Rivera

Aug 2025

Table of Contents

Abstract

Shelf-life prediction for snack products is a critical challenge in the highly competitive FMCG sector, where product quality is influenced by multiple degradation mechanisms. Moisture gain and lipid oxidation are two key pathways affecting product stability, yet they are often studied independently. Moisture behaviour is often described using the Guggenheim-Anderson-de Boer (GAB) model together with moisture transfer equations, while lipid oxidation follows temperature-dependent kinetic models based on Arrhenius principles. The goal of this project is to employ a deterministic approach and develop a combined computational model that integrates both moisture and lipid oxidation effects to estimate shelf life more accurately, providing valuable insights for product development and packaging design.

1. Introduction

The snack food segment of the Fast-Moving Consumer Goods (FMCG) industry is highly competitive; thus, maintaining product quality is not only expected but necessary. Predicting shelf life is crucial in different aspects, from consumer satisfaction to inventory management. For snack products, which often undergo quality degradation due to moisture gain and chemical deterioration, particularly lipid oxidation, having a reliable method to estimate shelf life is not just helpful but essential to avoid spoilage, reduce waste, and support timely distribution.

Over the years, a number of mechanistic models have been developed to address the spoilage pathways. Research on predicting shelf life limited by moisture gain is well established, with Moisture Sorption Isotherm (MSI) models forming the foundation of most approaches. These models describe how a product's moisture content relates to its water activity (a_w), a key factor in determining textural stability [1; 2]. The Guggenheim-Anderson-de Boer (GAB) model is one of the most widely used due to its flexibility and strong fit across a broad water activity range [3; 4]. Its parameters also offer physical interpretability, such as the monolayer moisture content (m_0), which is often considered a critical point for stability. As a result, the GAB model is frequently used in shelf-life predictions for packaged low-moisture foods. Simpler linear MSI models have also proven effective and can offer comparable accuracy in the low water activity ranges typically observed in crispy snacks [5]. These models are often used to estimate the time required to reach a critical water activity or moisture content level, which is closely tied to sensory attributes like crispness.

In parallel, shelf-life predictive modelling for lipid oxidation has received substantial attention, particularly for products containing fats and oils. Lipid oxidation leads to off-flavours and rancidity and is a major driver of quality loss [6; 7]. Predictive models are typically based on chemical kinetics and rely on Accelerated Shelf-Life Testing (ASLT) in this area, using temperature to accelerate deterioration [7]. Progress is monitored by tracking primary oxidation products (such as hydroperoxides) and secondary products (such as hexanal) [7; 8]. Kinetic behaviour is often modelled using pseudo-zero or first-order reaction assumptions, with temperature effects described by the Arrhenius equation to derive parameters like the rate constant (k), activation energy (E_a), and Q_{10} value [7; 9]. Some studies extend this with Eyring-Polanyi equations to incorporate thermodynamic parameters or apply multi-phase models to capture more complex reaction stages, such as initiation, propagation, and secondary acceleration phases [8; 9; 10].

While both moisture-driven and lipid oxidation-driven shelf-life predictive models are individually well developed, they have typically been applied separately. However, both mechanisms may simultaneously occur and not be entirely independent in some FMCG snack products [7]. Although some studies monitor water activity alongside oxidation during ASLT, few approaches couple the two processes within an integrated predictive model. This project seeks to address that gap by developing a combined shelf-life prediction model which will improve predictive accuracy and offer valuable insights for packaging design, product development, and shelf-life extension in the FMCG sector.

2. Methodology

2.1. Mathematical Models

This project models shelf life in snack products by focusing on two dominant degradation pathways: moisture gain and lipid oxidation.

2.1.1. Modelling Based on Moisture Gain

One aim of this project is to solve the equation that defines the shelf life of the product:

$$t_{s,\text{moisture}} = \int_{a_{w0}}^{a_{wc}} \frac{W_s \cdot L \cdot S(a_{w,\text{food}})}{P_w \cdot A \cdot p_0 \cdot (a_{w,\text{env}} - a_{w,\text{food}})} \, da_{w,\text{food}} \tag{1}$$

This integral represents the time required for the food's water activity to increase from its initial value a_{w0} to a critical threshold a_{wc}, beyond which product quality is compromised. The function $S(a_{w,\text{food}})$ represents the slope of the sorption isotherm and is typically derived from the GAB model:

$$w = \frac{W_m \cdot C \cdot K \cdot a_w}{(1 - K a_w)(1 - K a_w + C K a_w)} \tag{2}$$

Regarding the reasons for selecting the GAB model, several other sorption isotherm models have also been proposed in the literature to describe the relationship between equilibrium moisture content and water activity, such as the Moisture-Henderson Equation (MHE), the Modified Chung-Pfost (MCP) model, the BET equation, and so on. While the MHE and MCP models often provide accurate fits for particular food categories such as grains and seeds, the GAB model has been widely recognised as the most robust and generalisable formulation. It offers reliable performance across a wide range of product categories and water activity ranges, and is therefore adopted in this project for the estimation of sorption behaviour.

Moisture transfer through packaging is modelled via a one-dimensional form of Fick's law, incorporating packaging parameters like water vapour permeability (P_w), surface area (A), and film thickness (L). The rate of a_w change within the product is expressed as a function of the environmental $a_{w,\text{env}}$, food product $a_{w,\text{food}}$, and $S(a_{w,\text{food}})$ at the current state:

$$\frac{da_{w,\text{food}}}{dt} = \frac{P_w \cdot A \cdot p_0}{W_s \cdot L \cdot S(a_{w,\text{food}})} \cdot \left(a_{w,\text{env}} - a_{w,\text{food}}\right) \tag{3}$$

For practical shelf-life prediction, it is often advantageous to substitute the permeability-related terms ($P_w \cdot p_0 / L$) with the Water Vapour Transmission Rate (WVTR), a parameter readily obtained from packaging specifications. Incorporating

WVTR into Equation (3) leads to a simplified predictive formulation of moisture-driven shelf life:

$$t_{s,\text{moisture}} = \frac{W_s}{A \cdot \text{WVTR}} \cdot (M_c - M_0) \tag{4}$$

where M_c and M_0 respectively represent the equilibrium moisture contents at the critical and initial water activity levels. Unlike Equation (1), this expression eliminates the need for explicit integration, thereby enabling direct application in engineering design and comparative packaging studies.

2.1.2. Modelling Based on Lipid Oxidation

Lipid oxidation is modelled as a temperature-dependent reaction, following pseudo-zero or first-order kinetics. The Arrhenius equation is used to describe how the reaction rate varies with temperature, with key parameters including the activation energy (E_a) and pre-exponential factor (Z, also named k_0):

$$k_{\text{ox}} = Z \cdot e^{-E_a/(R \cdot T)} \tag{5}$$

where k_{ox} is the rate constant for the oxidation reaction, R is the universal gas constant, and T is the absolute temperature.

The induction period (IP), which represents the time passed before a rapid increase in oxidation products, is inversely related to the oxidation rate constant:

$$IP \propto \frac{1}{k_{ox}} \tag{6}$$

As lipid oxidation progresses, product quality declines sharply after the induction period, making it a reliable substitute for shelf life in oxidation-dominated systems. Thus, rather than tracking the entire reaction curve, the prediction of shelf life may be simplified to the estimation of IP.

To simplify parameter estimation and regression analysis, the relationship between IP and temperature can be linearised. By taking natural logarithms on both sides of

Equation (5) and combining it with the Arrhenius formulation, another equation obtains:

$$\ln\left(\frac{1}{IP}\right) = \ln Z - \frac{E_a}{R \cdot T} \tag{7}$$

This transformation allows the use of $\ln(1/\text{IP})$ as the dependent variable in a linear model with respect to reciprocal temperature $(1/\text{T})$, thereby enabling straightforward estimation of Arrhenius parameters from experimental data.

2.1.3. Combined Predictive Model

The overall shelf life is defined as the minimum time at which either the critical a_w or a critical oxidation marker (IP) is reached:

$$t_{s,\text{overall}} = \min\left(t_{s,\text{moisture}},\ t_{s,\text{oxidation}}\right) \tag{8}$$

2.2. Descriptive Statistical Analysis

This project obtained a total of 904 moisture experimental data points and 710 lipid oxidation experimental data points from Grupo Nutresa S.A.S. The food categories covered include crackers, sweet biscuits, cream filling, baked snacks, snacks (fat), snacks (oil), sandwich biscuits, and sugar wafers. For convenience, a code was assigned to each category as shown in Table 1.

Table 1: Category code and name

Code	Name
C	Crackers
D	Sweet biscuits
F	Cream filling
G	Snacks (fat)
O	Snacks (oil)
S	Sandwich biscuits
P	Baked snacks
W	Sugar wafers

Regarding moisture experimental data, the statistical characteristics of the two indicators—water activity and moisture content—are as follows: the mean water activity is 0.597, with a standard deviation of 0.185, ranging from 0.191 to 0.902; the mean moisture content is 0.091, with a standard deviation of 0.067, distributed between 0.018 and 0.427. The standard deviation of both indicators is relatively large compared to the mean, indicating significant variability among different samples. Further analysis of the scatter plot of moisture content versus water activity by category (Figure 1) shows a typical S-curve trend, consistent with the characteristics of general moisture adsorption isotherms.

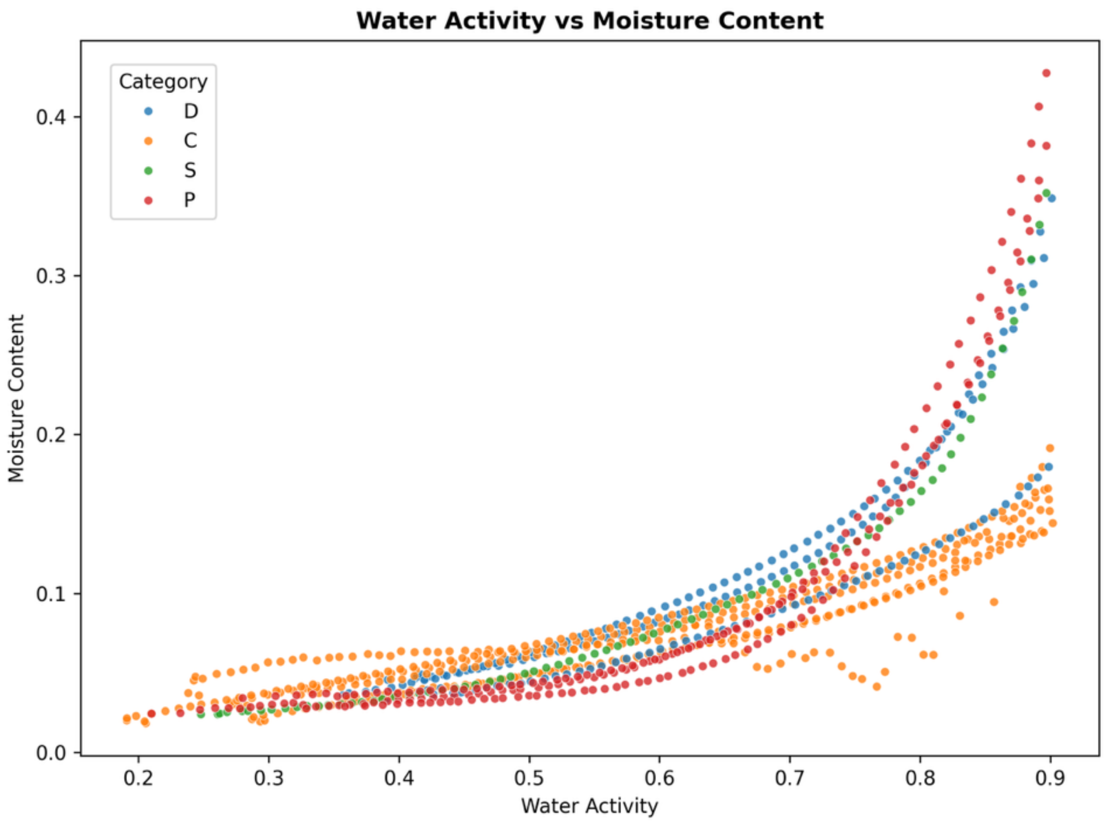

Figure 1: Scatter plot of moisture data

Regarding oxidation experimental data, the statistical results for the IP at high temperature (90–110°C) are as follows: the mean IP is 10.43 hours, with a standard deviation of 6.80, and a wide distribution range (0.017–23.967 hours), indicating a high degree of dispersion. Although the IP generally shortens with increasing temperature, the scatter plot by category (Figure 2) does not clearly show this trend, suggesting that the data may contain noise or systematic errors. Therefore, further validation and cleaning of the data are required to ensure they meet the basic assumptions of the oxidation kinetic model.

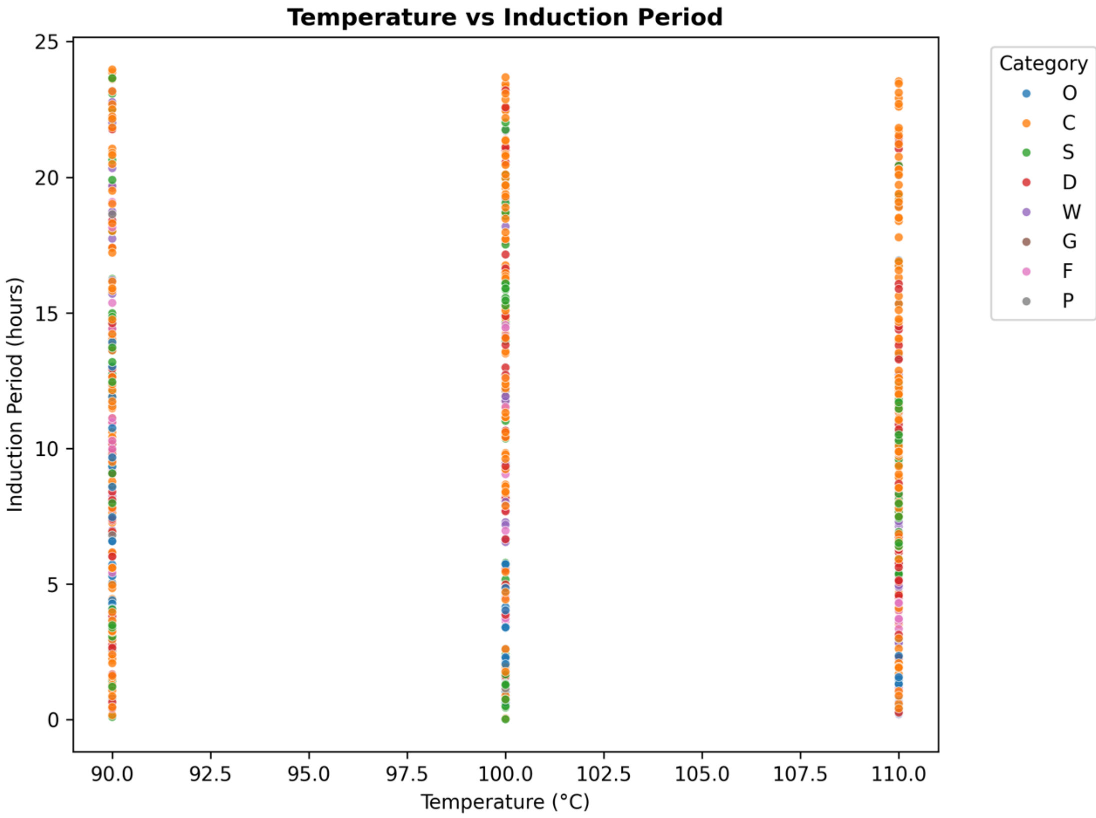

Figure 2: Scatter plot of oxidation data

2.3. Data Preprocessing

To ensure the reliability of the data, this project performed data preprocessing on the original dataset, primarily involving the removal of invalid data and the handling of outliers.

First, the validity of the lipid oxidation experimental data was verified. The basic principles of lipid oxidation kinetics indicate that the induction period should decrease with increasing temperature. However, a significant portion of the product data did not follow this fundamental pattern, suggesting potential abnormalities during the experimental process, recording, or transmission. As a result, these data were classified as invalid and excluded. Unfortunately, such abnormal data accounted for a high proportion, and after exclusion, the valid data volume for lipid oxidation experiments was reduced to only 125 entries. Although the sample size was significantly reduced, the reliability of the remaining data was improved.

Secondly, a data cleaning process was implemented using Python, covering all moisture and lipid oxidation experimental data. The processing steps included: handling missing values and outliers.

Missing value handling: Data containing missing values are directly excluded to avoid interference with subsequent model development.

Outlier handling: To reduce the potential influence of outliers on the overall analysis, an outlier identification method based on robust Z-scores was adopted. This method uses the median rather than the mean as the central location estimate and the median absolute deviation (MAD) rather than the standard deviation as the measure of dispersion, making it less sensitive to outliers and more robust. Robust Z-scores are especially calculated for variable IP in oxidation model, and data points with absolute values exceeding a predefined threshold are identified as outliers and removed. Simultaneously, it also ensures that outliers are not removed when the sample size is too small (size<6).

Through this process, a clean dataset was ultimately obtained, laying the foundation for subsequent shelf-life prediction modelling.

2.4. Model Development

This project utilised Python to develop a series of models for moisture adsorption isotherm fitting, lipid oxidation kinetic parameter estimation, and final shelf life prediction, and implemented a complete modelling and prediction process through custom scripts.

2.4.1. Modelling Based on Moisture Gain

To describe the equilibrium relationship between moisture content and water activity in food, this project employs the widely used GAB model in food science, whose equation is given in Equation (2).

The `moisture_fit.py` script was developed to perform model fitting. First, the GAB model function was defined based on the above equation. Next, the fitting function was defined, and the `curve_fit` function from SciPy was used for nonlinear least-squares fitting to solve for the model parameters W_m, C, and K. This algorithm iteratively optimises the parameters to minimise the sum of squared residuals between the model predictions and the experimental observations. Its advantage lies in its ability to fit complex nonlinear models well, without the need to provide a Jacobian matrix, making it highly adaptable.

When calling the function, the script first reads the experimental data from the raw data file and performs preprocessing (including removing missing values). After fitting, four key metrics are calculated using the NumPy and Scikit-learn libraries to assess the model's fitting quality: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and residual sum of squares (RSS). Among these, R^2 measures the extent to which the model explains data variability, with values closer to 1 indicating better fitting; RMSE and MAE quantify the average deviation between predicted and actual values from different perspectives; and RSS reflects the overall magnitude of residuals. These metrics collectively provide a comprehensive assessment of model performance.

Finally, the script plots the fitted adsorption isotherm curves by product category and exports the optimal GAB model parameters for each category to a CSV file, providing key inputs for the shelf-life prediction.

2.4.2. Modelling Based on Lipid Oxidation

The lipid oxidation reaction rate is described using a modified Arrhenius equation, where the relationship between the induction period (IP) and absolute temperature (T) can be expressed as a logarithmic linear form, as shown in Equation (7).

Similarly, the `oxidation_fit.py` script was developed for modelling. First, the valid data was pre-processed after removing outliers and missing values. Next, the Arrhenius equation function in the logarithmic form was defined, and the `curve_fit` function was also used for linear fitting to solve for the parameters k_0 and E_a. After fitting, metrics such as R^2, RMSE, MAE, and RSS were calculated to assess the reliability of the linear relationship. The script finally outputs Arrhenius fitting line graphs for each category and a corresponding kinetic parameter CSV file.

2.4.3. Combined Predictive Model

To comprehensively assess the combined effects of moisture gain and lipid oxidation on food shelf life, the `combined_model.py` script was developed to achieve a unified prediction framework.

The script first implemented single-mechanism shelf life predictive models based on the GAB model for moisture prediction and the Arrhenius model for oxidation prediction. Next, a combined prediction model was constructed: for given storage conditions (temperature, packaging characteristics, etc.), both the moisture-dominated shelf life and the lipid oxidation-dominated shelf life are calculated simultaneously. The final predicted shelf life is determined by the smaller of the two values, and the dominant mechanism at this point is recorded.

This script provides two prediction output modes. One is using the model parameters obtained from curve fitting, predicting the shelf life and dominant factor for all food categories under a set of standard default conditions, and saving the results to a CSV file. The other is interacting with the user via the command-line interface, allowing the user to customise parameters such as food category, storage temperature, and packaging properties. After executing, the predicted shelf life and dominant model for the specific scenario are directly output in the command line, enhancing the model's practicality and flexibility.

3. Results

3.1. Model Performance

This project conducted fitting and evaluation of the moisture adsorption isotherm model (GAB model) and the lipid oxidation kinetic model (Arrhenius model). By modelling different food categories, precise parameter estimates were obtained, and their predictive performance was assessed.

3.1.1. Performance of Moisture Model

The GAB model was used to fit the moisture adsorption behaviour of various food categories. The fitting parameters and evaluation metrics are shown in Table 2 (including categories C, D, P and S). Overall, the GAB model demonstrated excellent fitting quality for all food categories.

Table 2: GAB model fitting parameters and performance metrics

Category	W_m	C	K	R^2	RMSE	MAE	RSS
C	0.0384	9.8737	0.8455	0.8694	0.0132	0.0097	0.0659
D	0.0493	1.7842	0.9354	0.8751	0.0245	0.0167	0.1099
P	0.0563	0.5418	0.9893	0.9816	0.0128	0.0098	0.0353
S	0.0367	2.4512	1.0018	0.9982	0.0034	0.0029	0.0009

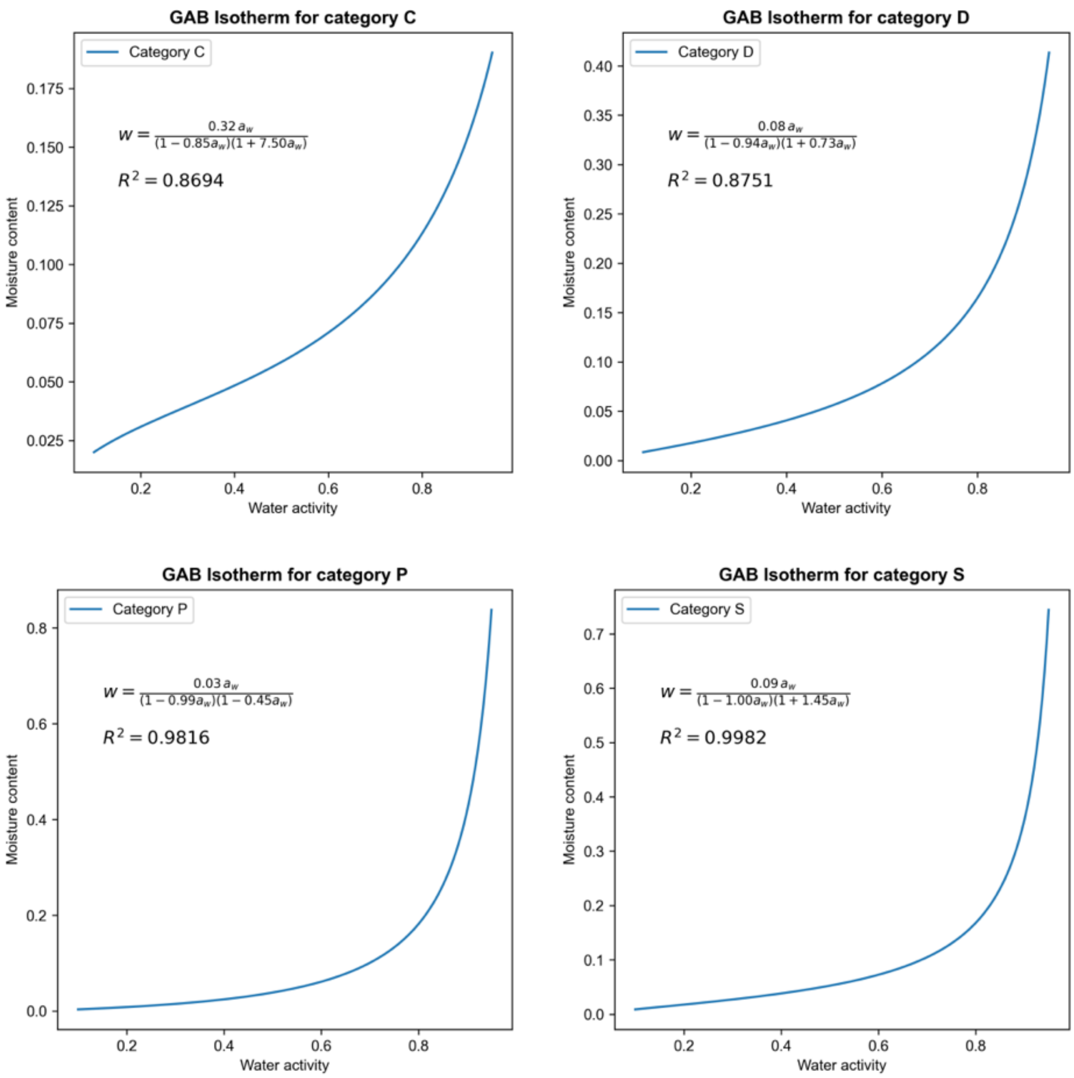

Figure 3a-d: Fitted GAB isotherm curves

Fitting Parameter Analysis: The parameter W_m (monolayer moisture content) varies across different categories. For example, the W_m value for category P (0.0563) is the highest, while the W_m value for category S (0.0367) is the lowest, reflecting the significant influence of product composition and structure on their monolayer water capacity. The constants C and K are also related to product characteristics. For example, category C has a higher C value (9.8737), suggesting stronger binding energy with water.

Fitting Quality Analysis: The coefficient of determination R^2 for all categories exceeds 0.86, with category S reaching 0.9982, indicating that the GAB model can explain the majority of experimental data variability. Category S has extremely low

RMSE (0.0034) and MAE (0.0029), and its RSS (0.0009) is significantly smaller than that of other categories, indicating the highest prediction accuracy for this product category. Although the R^2 value for category D is also high (0.8751), its RMSE (0.0245) and RSS (0.1099) are relatively large, indicating that the predicted values deviate to some extent. This may be due to the complexity of the moisture adsorption behaviour of this category or the high variability of the experimental data. Nevertheless, the fitting results for all categories generally meet the requirements for high-precision prediction.

3.1.2. Performance of Oxidation Model

The linearised Arrhenius model was used to fit the relationship between the induction period and temperature, yielding the activation energy Ea and pre-exponential factor k_0. Key parameters and fitting metrics are shown in Table 3.

Table 3: Arrhenius model fitting parameters and performance metrics

Category	E_a (kJ/mol)	k_0 (1/h)	R^2	RMSE	MAE	RSS
C	77.2084	6.4187×10^{9}	0.9999	0.0039	0.0036	0.0000
D	88.9015	2.8022×10^{11}	0.9955	0.0421	0.0397	0.0053
F	111.9194	9.1943×10^{14}	0.9833	0.1031	0.0972	0.0319
O	103.1274	7.8047×10^{13}	0.9949	0.0519	0.0489	0.0081
S	81.5554	2.4395×10^{10}	1.0000	0.0007	0.0007	0.0000
W	88.8285	3.2514×10^{11}	0.9970	0.0347	0.0327	0.0036

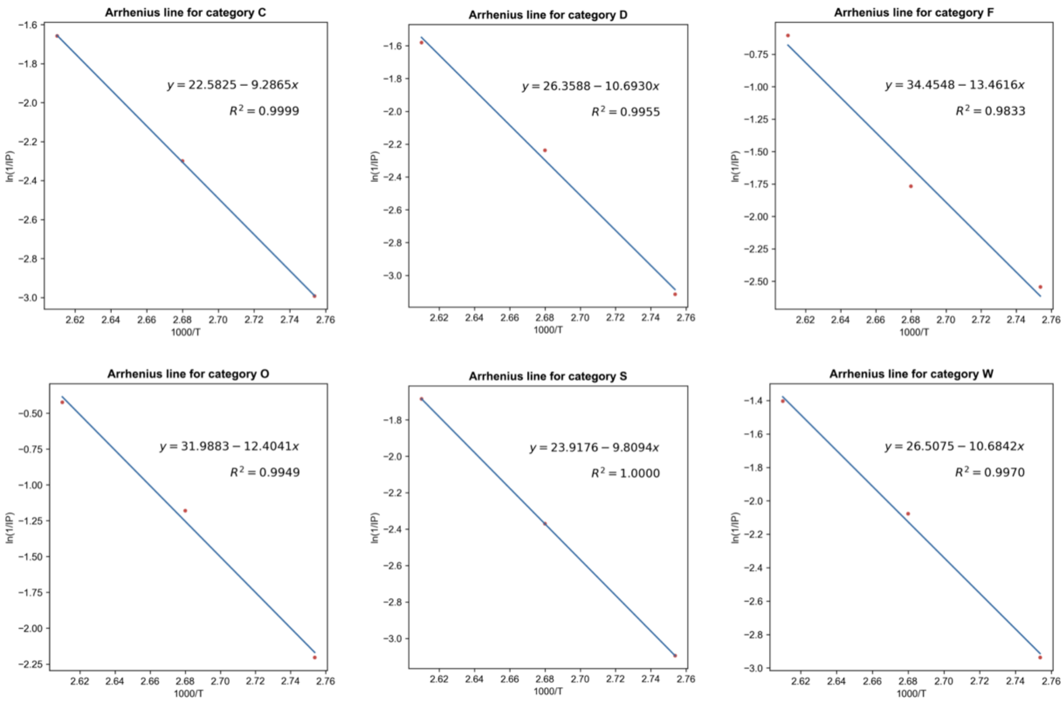

Figure 4a-f: Fitted Arrhenius lines (separate)

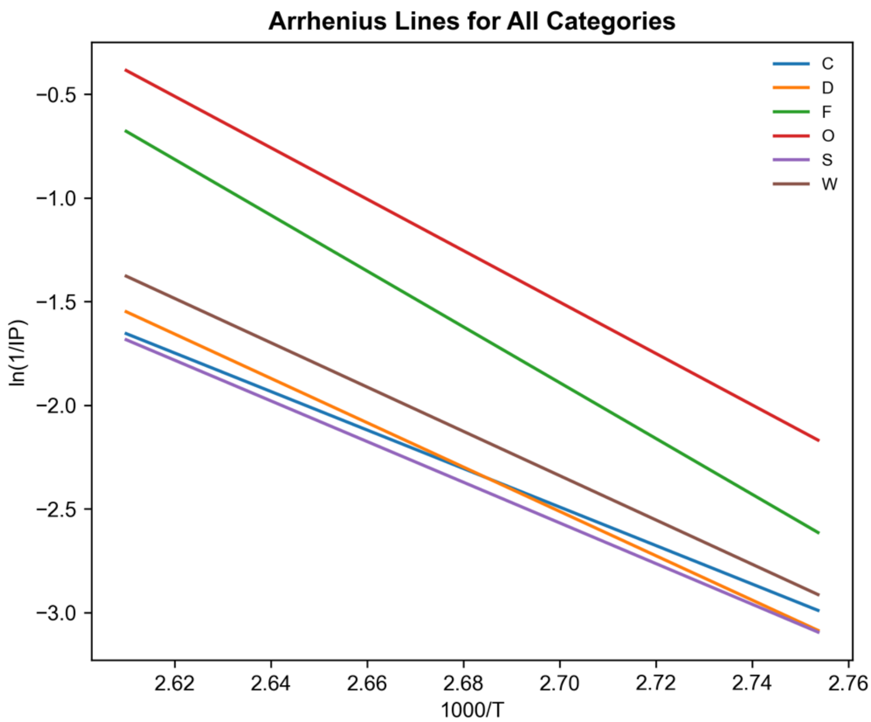

Figure 4g: Fitted Arrhenius lines (combined)

Fitting Parameter Analysis: The activation energy Ea varies significantly across different categories (77.21–111.92 kJ/mol), indicating that the sensitivity of lipid oxidation reactions to temperature differs among different food types. For example, category F has the highest Ea (111.92 kJ/mol), indicating that its oxidation reaction rate is most strongly influenced by temperature; whereas category C has the lowest Ea (77.21 kJ/mol), exhibiting relatively weaker sensitivity to temperature.

Fitting Quality Analysis: All categories of Arrhenius models exhibit high goodness-of-fit, with R^2 values exceeding 0.98. Categories C and S have R^2 values close to 1.000, and their RMSE, MAE, and RSS are all near zero, indicating that in these categories, there is nearly perfect linearity between $\ln(1/IP)$ and $1/T$, with minimal model prediction errors. The R^2 value for category F (0.9833) is slightly lower than that of other categories, and the error metrics (RMSE = 0.1031) are the highest. This may be due to inherent scatter in the experimental data points or minor variations in the reaction mechanism within the specific temperature range of this product category. However, overall, the fitting accuracy remains sufficient for reliable shelf-life prediction.

3.2. Deliverables

The core deliverables of this project include a set of model parameter datasets and a food shelf life prediction tool.

One of the deliverables is a series of CSV files containing the key kinetic parameters fitted for different food categories. The GAB model parameter file includes the optimal parameters for the GAB model of various food categories, namely the monolayer moisture content (W_m), Guggenheim constant (C), and correction parameter (K). The Arrhenius model parameter file contains the optimal parameters for the Arrhenius equation, namely the activation energy (E_a) and pre-exponential factor (k_0) for the oxidation reaction.

The other deliverable is a Python-based combined shelf-life prediction script (combined_model.py). This script forms the core application of this project and has the following functional features:

Automated batch prediction: The script can read the pre-calculated model parameters, automatically calculate the shelf life of all food categories under a set of predefined default storage conditions (e.g., standard room temperature and packaging characteristics), and identify the dominant spoilage mechanism (moisture or lipid oxidation). Results are output in a clear and structured format CSV file for easy comparison and analysis by users.

Table 4: Default prediction results ($T=25\ °C$, $W_s=200\ g$, $A=0.1\ m^2$, $WVTR=0.5\ g/(m^2\cdot day)$)

Category code	Category name	Shelf life (days)	Dominant model
C	Crackers	125.17	Moisture gain
D	Sweet biscuits	149.87	Moisture gain
F	Cream filling	1839.90	Lipid oxidation
O	Snacks (oil)	624.55	Lipid oxidation
P	Baked snacks	185.23	Moisture gain
S	Sandwich biscuits	137.51	Moisture gain
W	Sugar wafers	468.45	Lipid oxidation

Interactive customised prediction: To enhance the model's applicability, the script also provides a command-line interface. Users can dynamically input specific parameters such as food category, storage temperature, and packaging characteristics based on actual scenarios. The script then invokes the corresponding model for calculation and immediately outputs the predicted shelf life and dominant mechanism on the terminal interface, which is able to serve diverse practical decision-making scenarios such as product development, logistics optimisation, and retail environment assessment.

Table 5: Customised prediction results (Example)

Code	T (°C)	W_s (g)	A (m^2)	WVTR ($g/(m^2\cdot day)$)	Shelf life (day)	Dominant model
C	25.0	200.0	0.10	0.5	125.17	Moisture gain
C	30.0	200.0	0.10	0.5	125.17	Moisture gain
C	25.0	200.0	0.10	0.8	78.23	Moisture gain
C	25.0	200.0	0.05	0.5	218.48	Lipid oxidation
S	20.0	200.0	0.10	0.5	137.51	Moisture gain
S	25.0	200.0	0.10	0.5	137.51	Moisture gain
S	35.0	200.0	0.10	0.5	114.15	Lipid oxidation
S	20.0	100.0	0.10	0.5	68.75	Moisture gain

4. Discussion

This project successfully established a food shelf life prediction framework based on moisture and lipid oxidation theory and validated the excellent performance of its core models. However, this project also has several limitations, which point the way for future research and model optimisation.

4.1. Limitations of Moisture Model Assumptions

The GAB model used in this project is currently based on a key assumption: moisture adsorption isotherms are measured at a constant room temperature (25°C). However, in actual storage and transportation processes, temperature often changes, which can significantly affect moisture activity and adsorption behaviour, thereby influencing moisture-based spoilage rates. The current predictive model has not incorporated the dynamic effects of temperature on moisture properties, which may introduce prediction biases in actual scenarios.

4.2. Limitations of Diversity of Spoilage Mechanisms

The accuracy of the model highly depends on the completeness of the input data. This project observed that for certain categories (such as categories F and O), the model-predicted lipid oxidation-dominated shelf life was significantly longer than the reasonable range recognised in practice (Table 4). This deviation is likely due to the limitations of the current dataset: for these products, only data on the single mechanism of lipid oxidation was obtained. In reality, their spoilage may result from the combined effects of multiple mechanisms (such as microbial growth, enzymatic reactions, and changes in physical properties), and these mechanisms not included in the model may be the true dominant factor for the actual shelf life.

4.3. Limitations of Mechanism-Driven Models

The current model's predictive results primarily rely on physical-chemical mechanisms, with experimental data used to adjust the mechanism model parameters. However, the final predictions have not yet been adjusted using real, labelled shelf-life endpoint data. This results in the model tending to provide a theoretical ‘induction period’ or ‘threshold value’ rather than the strictly defined ‘shelf life’ used in the market.

5. Conclusion

This project successfully developed and validated a food shelf life prediction framework based on physical and chemical mechanisms. By integrating two core spoilage mechanisms, moisture adsorption (GAB model) and lipid oxidation (Arrhenius model), the framework constructs a combined prediction model capable of scientifically identifying the dominant factor influencing food shelf life under specific storage conditions and exporting the predictive values. Although further expansion is still needed, this project has laid a solid foundation for the future development of more generalised prediction models.

6. Future Work

Based on the limitations mentioned above, future work will focus on the following three aspects:

6.1. Temperature-Dependent Moisture Models

Future research should focus on obtaining moisture adsorption experimental data under different temperature conditions to establish temperature-dependent GAB model parameters or introduce more complex coupled models to enhance the model's universality under variable conditions.

6.2. Multi-Mechanism Model Expansion

Future efforts should expand the dimensions and scope of experimental data, systematically measuring more potential spoilage indicators to construct a multi-factor spoilage model. At the same time, it is necessary to consider the interactions between different spoilage mechanisms rather than simply taking the minimum value obtained from separate predictions as the shelf life. On the other hand, more food categories should be included to enhance the practical applicability and reliability of the prediction tool.

6.3. Data-Mechanism-Driven Combination

A direction for future research is to collect accurate time data on product spoilage under real storage conditions. Once such real shelf-life data is obtained, advanced methods such as Physics-Informed Machine Learning (PIML) can be used to optimise existing models. PIML can embed established physical models such as GAB and Arrhenius as constraints within machine learning algorithms, using real-world data to slightly adjust model outputs. This allows the models to maintain physical interpretability while ensuring that predicted values align more closely with actual observed shelf lives. This shift from purely mechanism-driven approaches to a data-mechanism-driven combination will significantly enhance prediction accuracy.

References

[1] Lee, D. S., & Robertson, G. L. (2022). Shelf-life estimation of packaged dried foods as affected by choice of moisture sorption isotherm models. Journal of Food Processing and Preservation, 46(3), e16335.

[2] Yusuf, H. L., Igwegbe, A. O., & Nassarawa, S. S. (2024). Moisture Sorption Studies using the GAB Model of a Beef Based Dambun Nama (A Ground Meat Product) Stored over a Period of Six Months. African Journal of Agriculture and Food Science, 7(2), 51-63.

[3] Kulchan, R., Boonsupthip, W., & Suppakul, P. (2010). Shelf life prediction of packaged cassava-flour-based baked product by using empirical models and activation energy for water vapor permeability of polyolefin films. Journal of Food Engineering, 100(3), 461-467.

[4] Staudt, P. B., Tessaro, I. C., Marczak, L. D. F., de P. Soares, R., & Cardozo, N. S. M. (2013). A new method for predicting sorption isotherms at different temperatures: Extension to the GAB model. Journal of Food Engineering, 118(2), 247-255.

[5] Robertson, G. L., & Lee, D. S. (2021). Comparison of linear and GAB isotherms for estimating the shelf life of low moisture foods packaged in plastic films. Journal of Food Engineering, 291, 110317.

[6] Wang, D., Xiao, H., Lyu, X., Chen, H., & Wei, F. (2023). Lipid oxidation in food science and nutritional health: A comprehensive review. Oil Crop Science, 8(1), 35-44.

[7] Bravi, E., Sileoni, V., Perretti, G., & Marconi, O. (2020). Accelerated shelf-life model of gluten-free rusks by using oxidation indices. Food Chemistry, 326, 126971.

[8] Farhoosh, R. (2022). New insights into the kinetic and thermodynamic evaluations of lipid peroxidation. Food Chemistry, 375, 131659.

[9] Jaimez-Ordaz, J., Pérez-Flores, J. G., Castañeda-Ovando, A., González-Olivares, L. G., Añorve-Morga, J., & Contreras-López, E. (2019). Kinetic parameters of lipid oxidation in third generation (3G) snacks and its influence on shelf-life. Food Science and Technology, 39(Suppl. 1), 136-140.

[10] Nguyen, K. A., Hennebelle, M., van Duynhoven, J. P. M., Dubbelboer, A., Boerkamp, V. J. P., & Wierenga, P. A. (2024). Mechanistic kinetic modelling of lipid oxidation in vegetable oils to estimate shelf-life. Food Chemistry, 433, 137266.

Appendices

Symbol	Description
w	Equilibrium moisture content
a_w	Water activity
W_m	Monolayer moisture content
C	Guggenheim constant (related to monolayer binding energy)
K	Multilayer constant (related to multilayer water properties)
$\frac{da_{w,food}}{dt}$	Rate of change of food water activity over time
P_w	Water vapour permeability of the packaging material
A	Surface area of the package
p_0	Saturation vapour pressure of pure water at storage temperature
W_s	Mass of dry solids in the product
L	Thickness of the packaging film
$a_{w,env}$	Water activity of the external environment
$a_{w,food}$	Water activity of the food product
a_{w0}	Initial water activity of the food product
a_{wc}	Critical water activity at which the product is considered unacceptable
$S(a_{w,food})$	Slope of the sorption isotherm at the current food water activity
$t_{s,moisture}$	Shelf life limited by moisture gain
k_{ox}	Rate constant of lipid oxidation
Z	Pre-exponential factor
E_a	Activation energy of the oxidation reaction
R	Universal gas constant
T	Absolute temperature

Table 6: Appendices