Modeling and Optimum Experimental Design of *Salmonella* Inactivation in Inoculated Wheat Flour

Kasey Nelson[1], Ian Klug[1], Yawei Lin[2], Dangkamol Wongthanaroj[3], Yunwei Chen[4], Kirk Dolan[1,2], Teresa Bergholz[2], Ian Hildebrandt[1], Michael James[1], Bradley Marks[1]

[1]Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing MI, USA [2]Department of Food Science and Human Nutrition, Michigan State University, East Lansing, MI, USA
[3]School of Packaging, Michigan State University, East Lansing, MI, USA [4]Department of Statistics & Probability, Michigan State University, East Lansing, MI, USA

Abstract

Introduction: Previous research involving consumer-based thermal heat treatments of raw flour is limited in scope. Generation of an inactivation model and optimum experimental design parameters will aid further investigations of the proposed consumer-based thermal treatments.
Purpose: The objective was to estimate modeling parameters for the inactivation of *Salmonella* Enteritidis PT30 using data gathered from pilot-scale experiments.
Methods: Multiple parameter estimation methods were conducted for comparison. Data were collected from pilot-scale thermal inactivation study of *Salmonella* in wheat flour. The primary log-linear model was combined with a modified Bigelow secondary model to estimate inactivation within non-isothermal treatment datasets at 121 and 177°C. The target parameters were $D_{90°C}$ (min), z_T(°C), log(N_0) (log CFU/g), and Ra_w (unitless). MATLAB was used to estimate parameter values through scaled sensitivity coefficients (SSC), ordinary least squares estimation (OLS), sequential estimation, and bootstrapping methods. Plots of residuals, SSC, confidence bands, and prediction lines were created at relevant estimation step.
Results: The resulting parameter values for both OLS were $D_{90°C}$=6.24 min, z_T =2.11E+11°C, log(N_0)=8.19 log CFU/g, and Ra_w = 0.42. Sequential estimation values were $D_{90°C}$=-128 min, z_T=328 °C, log(N_0)=7.38 log CFU/g, and Ra_w=33.3 . Bootstrap parameter values were $D_{90°C}$=10.7 min, z_T = 2.11E+11 °C, log(N_0)=7.24 log CFU/g, and Ra_w=0.89. Parameter values were inconsistent and did not converge after the implementation of each estimation method. SSC plots of the parameters indicated an ability to most accurately predict log(N_0) and to least accurately estimate z_T, with corresponding levels of lower and higher relative errors.
Significance: An iterative, multiple-estimation method approach to modeling *Salmonella* inactivation mechanisms within flour can optimize parameter estimates, provide insight to the nature of the model parameters and the errors they may contain, and inform the course of further experimentation. The generated model may have applications with other consumer-based thermal treatments in low-moisture foods.

Introduction

A *Salmonella* recall linked to all-purpose flour announced on 30 March 2023 caused reported illnesses in 14 individuals in the United States (as of 12 July 2023) *(1)*.
Prior research *(2)* determined that home-scale thermal treatments were not effective in reducing *Salmonella* in flour to acceptable levels (avg. 5 log reduction), but that research did not generate an inactivation model for consumer-based thermal treatments or use an additional target temperature based on consumer suggestions. Model generation and parameter estimation of *Salmonella* inactivation in all-purpose flour may provide insight into the efficacy of consumer-based thermal treatments.

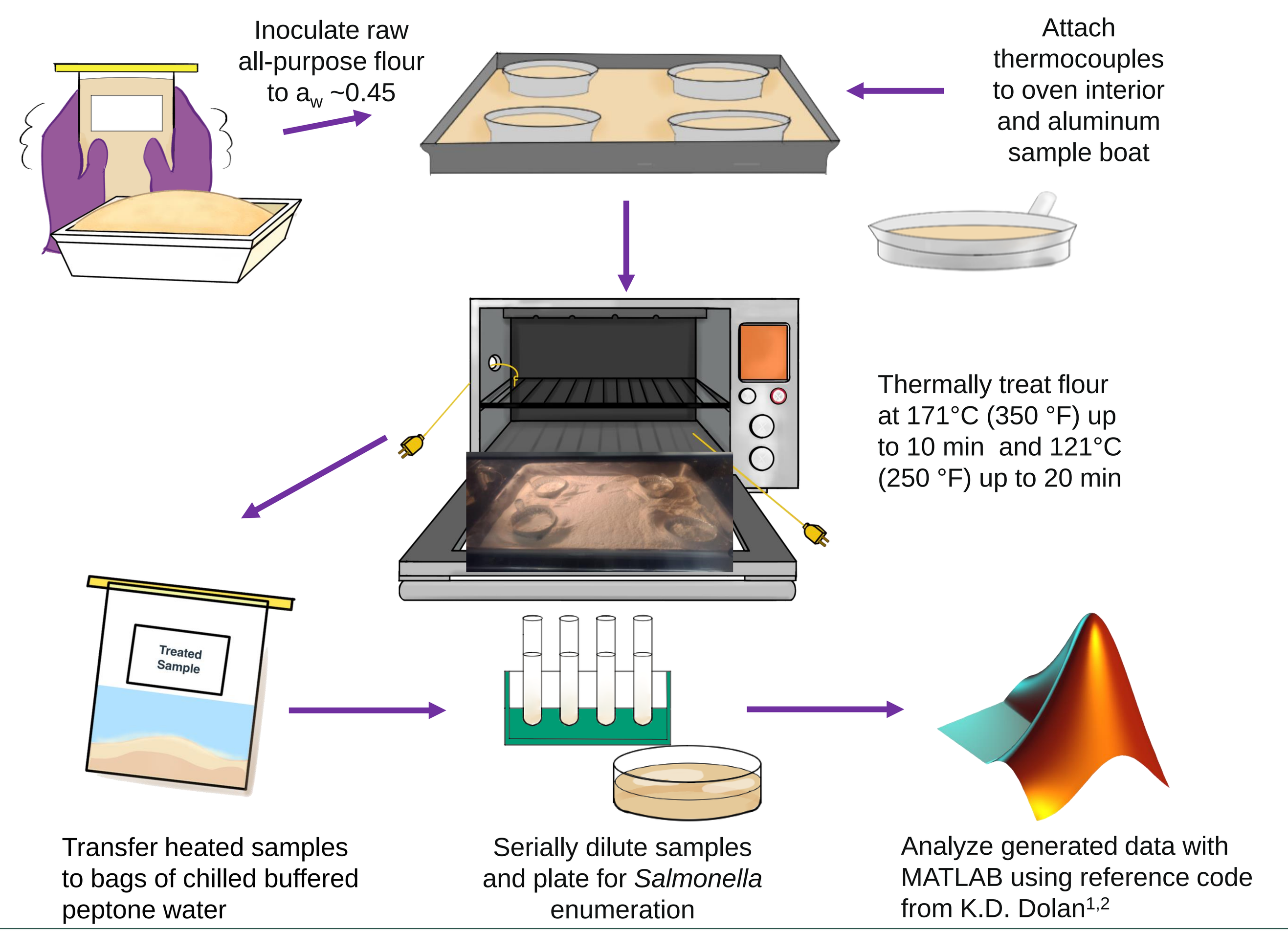
Objectives

- Thermally treat raw all-purpose flour at two target temperatures (121 and 177°C) to evaluate the efficacy of consumer-based thermal treatments.
- Estimate inactivation parameters (D_r, Z, log N_0, and Ra_w) for a multi-variable *Salmonella* inactivation model, given dynamic data.
- Compare and contrast the impact of multiple parameter estimation approaches.
- Contextualize the nature of the dataset and its impact on model parameter estimatability.

Methods and Materials

Data collection

Treatment methodology taken from Nelson et. al, 2021 (2).



Modeling

Rationale:

1. Iteration & multiple estimation methods can converge model parameters toward an optimum solution.
2. Each method has benefits and drawbacks.
   1. When to apply, and where?
   2. Can we be *confident* in generated parameters (i.e. do they make sense?)
   3. How does the nature of the dataset impact generated results?

Model Set-Up

Combination of primary log-linear model and modified Bigelow secondary models:

$$\frac{d(\log(N))}{dt} = -\frac{1}{D_{T,a_w}} \qquad D(T,a_w) = D_r * 10^{\left[\left(\frac{T_r - T}{z_T}\right) + Ra_w(a_{w,ref} - a_w)\right]}$$

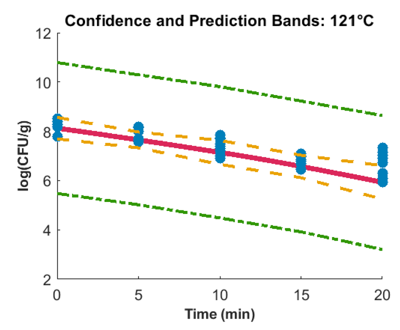
$$\log(N) = \log(N_0) - \int_0^t \left(\frac{1}{D_r}\right) 10^{\left[\left(\frac{T_r - T}{z_T}\right) + Ra_w(a_w, ref - a_w)\right]} dt \qquad (3)$$

Parameters	Initial guess
D_r (min)	10
z_T (°C)	70
log N_0 (log(CFU/g))	8.67
Ra_w (unitless)	0.5

Insert initial guess values into the model as beta (β) parameters.

Ordinary Least Squares *& Scaled Sensitivity Coefficients*

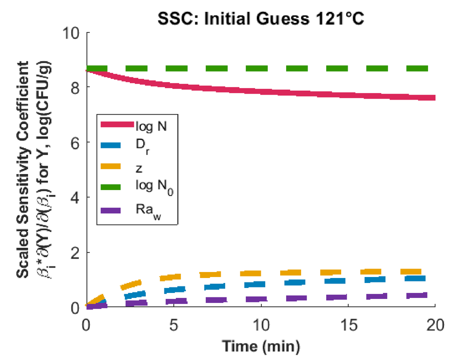
Ordinary Least-Squares Estimation (OLS) is a standard practice that uses the least squares method to estimate parameters by minimizing the sum of squares between the dataset, observations, and the provided model.



Benefits and drawbacks:
- **Depending on the model and parameters, relative error can be high**
- Confidence bands are asymptotic and symmetrical
  - OLS can only approximate asymptotic and symmetrical bounds
- Relative error can impact understanding (over and underestimation) of *Salmonella* survival at certain timepoints

Modeling Practice	Parameter Estimations	Quality of Models	Improvement of Data Collection
OLS	+	~	0
+ +: Highly useful, +: Useful, ~: Use depends on certain factors, 0: Not useful			

Sensitivity coefficients are a way to visually represent the uncertainty of individual model parameters in relation to the calculated results.

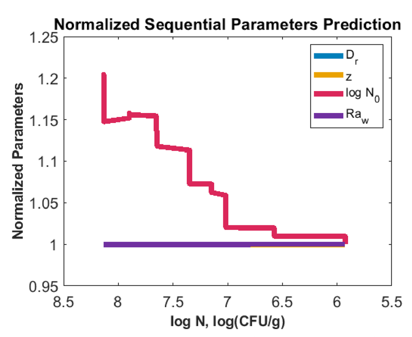


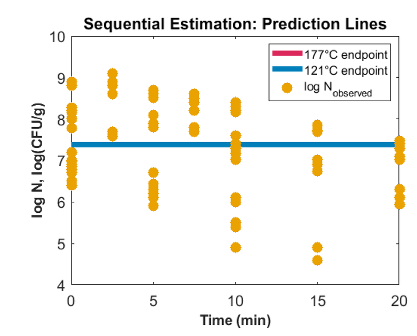
Benefits and drawbacks:
- Can describe the uncertainty of parameter estimates and relationships between them and the model outputs
  - Which ***β*** parameter impacts the results the most?
- Can describe if something is wrong with the model or chosen parameter values
- **Not a regression method**

Modeling Practice	Parameter Estimations	Quality of Models	Improvement of Data Collection
SSC	0	0	+ +
+ +: Highly useful, +: Useful, ~: Use depends on certain factors, 0: Not useful			

Sequential Estimation

Sequential Estimation updates parameter estimation results as each datum is added. The estimation is considered accurate if all the parameters converge to the normalized value = 1.0. .



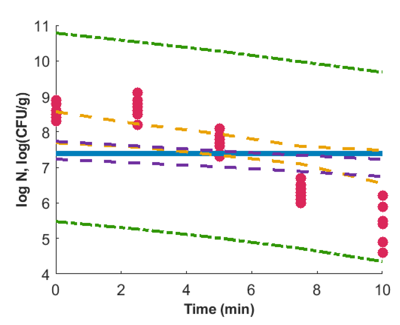


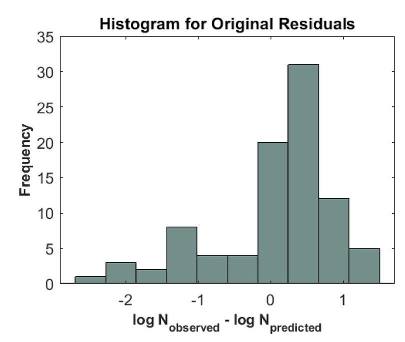
Benefits and drawbacks:
Can show if model parameters are converging
- **Cannot specify if the model form is incorrect or if there is missing data**
  - Only indicates that there is "something off" with the model
  - Model quality and improvement of data collection are coupled

Modeling Practice	Parameter Estimations	Quality of Models	Improvement of Data Collection
Sequential	+	~	~
+ +: Highly useful, +: Useful, ~: Use depends on certain factors, 0: Not useful			

Bootstrapping

Bootstrapping uses randomly drawn data points from the existing data set to estimate parameters, confidence intervals, and descriptive statistics. Additionally, a histogram of the bootstrapped residuals is generated.

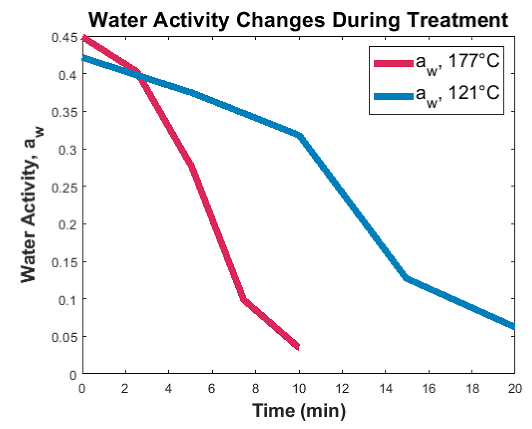




Benefits and drawbacks:
- "Garbage in, garbage out."
- **Results depend on which samples are randomly drawn for analysis**
  - Too few data points can result in a drastically different model if drawn data discards lower and upper 5% of values
- Bootstrap can have asymmetrical variance estimates

Modeling Practice	Parameter Estimations	Quality of Models	Improvement of Data Collection
Bootstrap	+ +	+	0
+ +: Highly useful, +: Useful, ~: Use depends on certain factors, 0: Not useful			

Results



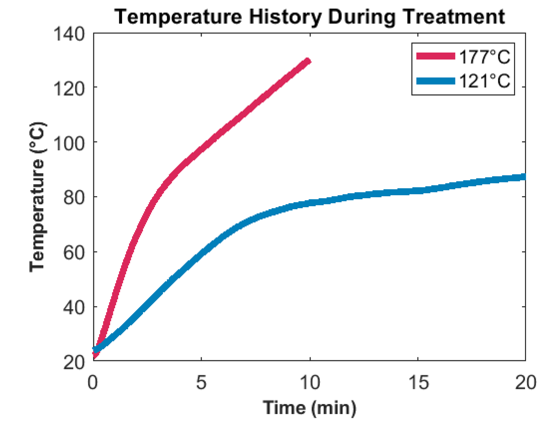


Table 1: Generated Parameter Estimates and Standard Error per Modeling Method

Modeling Method	Parameter Estimate [*Standard Error*]			
	D_r (min)	Z_T (°C)	Log N_0 (log CFU/g)	Ra_w (unitless)
OLS	6.24 [*6.03*]	2.11E+11 [*3.5E-16*]	8.19 [*0.16*]	0.42 [*1.03*]
Sequential	-128 [*407.20*]	328 [*1.04E+3*]	7.39 [*0.06*]	33.3 [*105.30*]
Bootstrap	10.7 [*0.03*]	2.11E+11 [*4.23E+7*]	7.24 [*0.05*]	0.89 [*0.01*]

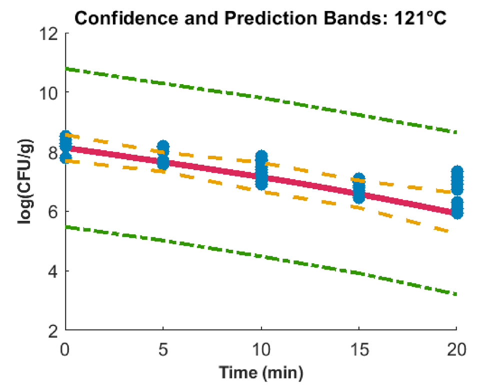
No treated samples achieved greater than an average 5 log reduction.

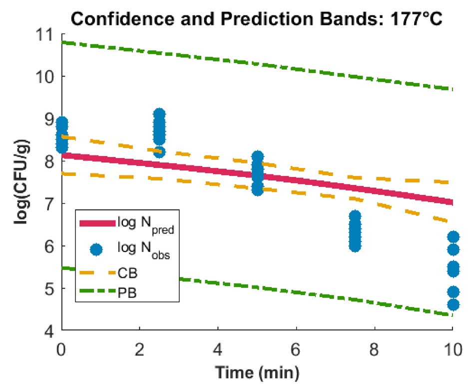
Discussion

Outputs:

Inferences:

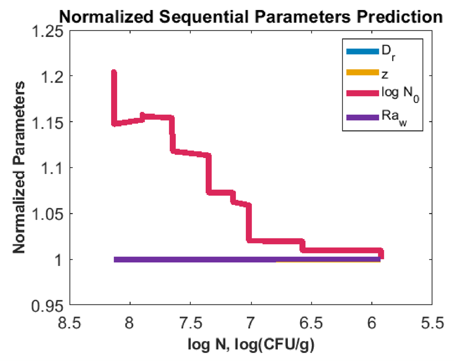
Ordinary Least Squares





- 2.5 min log counts at 177°C are higher than initial log counts
- Figures report: Temperature doesn't matter over time
- Predicts half as much reductions because "only time matters"
- RMSE (=1.10 log) is weighted by the 177°C datapoints

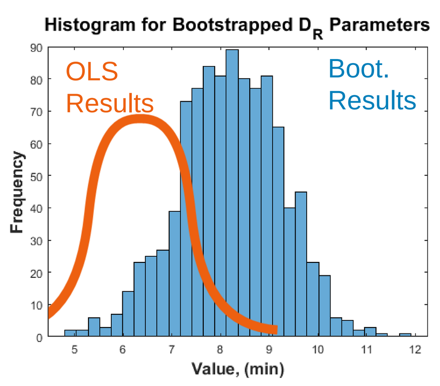
Sequential Estimation

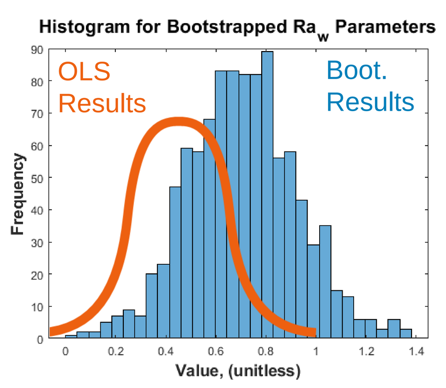


Parameters	Initial guess
D_r (min)	10
z_T (°C)	70
log N_0 (log(CFU/g))	8.67
Ra_w (unitless)	0.5

- D value should not be negative
- Parameter predictions for Z_T, log N_0, and Ra_w reflect OLS predictions
- Again, "the only thing that matters is time"
- The Log N_0 plot is essentially the average of the log populations every 2.5 min

Bootstrapping





- Bootstrapped parameters D_r and Ra_w are wider than the OLS "bell-curve" estimates
- OLS bell-curve is an "optimal" estimate
- The bootstrap results were inclusive of OLS results

Takeaways:

- According to the estimated model parameters, temperature does not affect inactivation. **This does not match reality.**
- Temperature and a_w are quasi-linear with time and may "counter-affect" each other in the model outputs.
- Simultaneous parameter estimation of dynamic inputs can greatly impact the reported value.

Table 2: Utility of Modeling Practices

Modeling Practice	Parameter Estimations	Quality of Models	Improvement of Data Collection
SSC	0	0	+ +
OLS	+	~	0
Sequential Estimation	+	~	~
Bootstrapping	+ +	+	0
+ +: Highly useful, +: Useful, ~: Use depends on certain factors, 0: Not useful			

References & Acknowledgements

This work is supported by the Agriculture and Food Research Initiative, Sustainable Agricultural Systems Program grant no. 2020-68012-31822 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

USDA National Institute of Food and Agriculture
U.S. DEPARTMENT OF AGRICULTURE

References

1. Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Foodborne, Waterborne, and Environmental Diseases (DFWED). (2023, June 7). *CDC: Salmonella outbreak linked to flour*. Centers for Disease Control and Prevention. https://www.cdc.gov/salmonella/infantis-03-23/index.html
2. Nelson, K., Hildebrandt, I. M., James, M., & Marks, B. (2021). Assessment of Consumer Flour Thermal Treatments on the Reduction of *Salmonella.* Poster presentation. International Association for Food Protection, Cleveland, OH, USA. (Presented online).
3. Jin, Y., Pickens, S., Hildebrandt, I. M., Burbick, S. J., Grasso-Kelley, E. M., Keller, S. E., & Anderson, N. M. (2018). Thermal inactivation of salmonella agona in low–water activity foods: Predictive models for the combined effect of temperature, water activity, and food component. *Journal of Food Protection*, *81*(9), 1411–1417. https://doi.org/10.4315/0362-028x.jfp-18-041