"STEAMS" Approach of Preparing Freshest STEAMed Dumplings

Mason Chen[1], Yvanny Chang[2], Patrick Giuliano[3], and Charles Chen[3]
Stanford University, Palo Alto, USA[1]
Prospect High School[2]
Morrill Learning Center, San Jose, USA[2]
Mason05@ohs.stanford.edu, yvanny.chang@gmail.com, patrick.giuliano@gmail.com, Charles.chen.training@gmail.com

Abstract

This paper would demonstrate the STEAMS (Science, Technology, Engineering, Artificial Intelligence, Mathematics, Statistics) methodology on how to customize the Dumpling Cooking process based on the Dumpling product types. Most foods like dumplings are made without precise control of cooking parameters. During the dumpling cooking process, the water temperature and cooking duration are the most important factors to determine whether Dumplings are under, fully or over cooked. The dumpling type, dumpling weight, batch size would also impact the cooking process. A specially designed and structured Design was conducted to build a predictive model of estimating the cooking duration. The HACCP (Hazard Analysis Critical Control Point) and ISO 22000 Food Safety Management were adopted. Modern Data Mining Neural technique was conducted. The results have observed both main effects mainly from the Boiling Temp, Dumpling Product Type, Dumpling Size/Batch and interaction effects which are constrained by the mixture of the dumpling composition.

Keywords
STEAMS, Data Mining, HACCP, Dumpling, Neural, Partitioning, Project Management

1. Introduction
The objectives of this paper are to utilize the STEAMS (Science, Technology, Engineering, Artificial Intelligence, Mathematics and Statistics) to minimize dumpling preparation time and standardize cooking parameters for various dumpling products (meat, seafood, veggie…). The STEAMS methodology is an enhanced STEM/STEAMS methodology [1,2]. STEAMS method has three core visions: (1) replace Art with Artificial Intelligence embracing in modern Big Data AI world, (2) separate Statistics from Mathematics to draw practical decision and conduct risk management, and (3) interdisciplinary STEAMS integration through thorough systematic Engineering Problem Solving, AI Machine Learning and Statistical Modeling to discover the Science Insights and Technology Applications seamlessly. Authors have previously published several STEAMS literatures in different fields [3-5]. Therefore, the detailed STEAMS vs. STEM/STEAM comparison won't be addressed in detail here.

2. STEAMS METHODOLOGY
This section would break down STEAMS Methodology to six elements: (1) Science, (2) Technology, (3) Engineering, (4) Artificial Intelligence, (5) Mathematics, and (6) Statistics.

2.1 Science
Figure 1 has demonstrated the Dumpling Heat Transfer Physics. There are three heat transfer mechanisms happening during dumpling boiling: (1) Thermal Conduction: energy transfer in the form of thermal energy (heat) goes from the cooker wall into the water, (2) Natural Convection boiling in the water generates cavitation bubble collapse and turbulent flow, creating heat energy gradients, and (3) Diffusion energy transfer in the form of thermal energy (heat) goes into the dumpling mass, changing the configuration of the meat proteins.. Dumpling cooking physics and modeling has also been demonstrated through spherical thermal conductivity model [6]. The other literature provides

a basic physical representation of cooking meat through boiling, where the coefficient of temperature conductivity is derived on the basis of considering the difference in the temperature of denaturation of meat and the boiling point of water [7]. Understanding these boiling heat transfer physics was the first and the most important step in conducting the STEAMS project which could identify the critical process parameters for building Dumpling Process modeling.

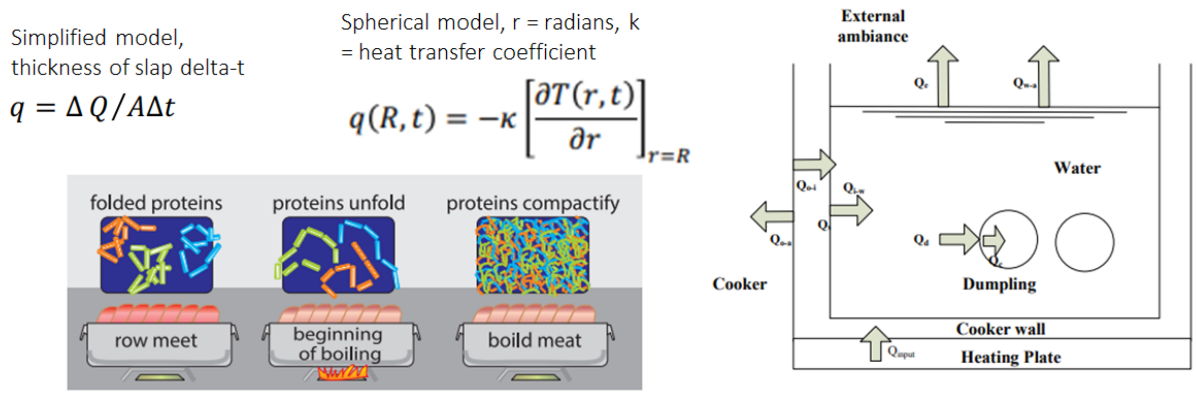

Figure 1. Dumpling Cooking Heat Transfer Physics

2.2 Technology

The second part of STEAMS would address the Dumpling Cooking Experimental Technology. To build dumping cooking model, three key experimental elements are considered: (1) Boiling Cooker: Composition of cooking vessel (metal vs stone), energy delivery method (electric vs gas), (2) Temperature/Time Regulation: Temperature/Time Regulation: Cooking medium temperature regulation by Infrared (IR) and time by digital timer, and (3) Dumpling Composition: Composition controlled by weighing balance. Carefully choosing the right technology is critical to the experimental success. The required accuracy and precision of these technologies/Equipment would be addressed in the next Engineering Section.

2.3 Engineering

This Engineering section would consist of two parts: (1) how to apply project management and (2) how to adopt Quality Control Engineering method HACCP and ISO 22000 Food Safety Management.

The project management would be deployed by the following five steps:

Phase I: Decide on Tools (Balance Scales, Utensils, Cooking Pots, Mixing Bowls). Consider equipment accuracy/ precision as applicable)

Phase 2: Dumpling Assembly Process Line Layout, including headcount allocation and utilization (takt) among each cook & master Chef.

Phase 3: DSD DOE – Optimal Orthogonality, Power, and Uniformity

Phase 4: Proper experimental execution: Preparing the dumplings to the indicated experimental settings, measuring/recording the response (cooking time).

Phase 5: Analysis and effective interpretation of data in JMP, summary of results, conclusions, and next steps.

The 2nd Engineering part would adopt the HACCP (Hazard Analysis Critical Control Point) and ISO 22000 Food Safety Management [8, 9]. 7 Principles of HACCP were adopted when standardizing the Dumpling Process: (1) Conduct Hazard Analysis, (2) Establish the Critical Control Points, (3) Establish the Critical Limits, (4) Establish the Monitoring Procedures, (5) Establish the Correct Action, (6) Verification, and (7) Recordkeeping. The associated regulation ISO 22000 was referred to when planning the experiment. Figure 2 has shown the project's HACCP progress. Hazard analysis was conducted through identifying their functions and associated variables. The process teams were also listed to monitor the process. 8 CCP points and their limits were identified, validated and verified. The monitoring/detection methods were specified. The communication from upstream to downstream would be facilitated by two internal messengers through the material/process transfer card for the entire dumpling process.

	Function	Variables	Leader	Member	CCP	Validation	Monitoring	Verification
A	Prepare Vegi Ingradient	Dry Duration	Sean	Alan	Container Labeling	Define how to cut vegi	Vegi Container Labeling	Dry all vegi pieces
B	Mix, Stir (2 meat x 2 vegi x 3 ratios)	1.Meat, 2.Vegi, 3. Ratio (25%, 50%, 75%)	Leo/Matt	Kathy's Boy	Weight Ratio +/- 1%	Weight Tolerance +/1 gram	Check Meat/Vegi Ratio	Visual Inspection on Ratio/Stir Efficiency
C	Dumpling Weight (3) X Pi Type (2)	4. Weight (20, 25,30),5. Pi	Kathy	Kathy's Daughter	Dumpling Weight +/- 0.5gm	Weight tolerance within +/-0.1 gram	Pi Type Labeling	Meet Dumpling Weight Target within +/-0.5 gram
D	Make Dumplings (Shape, Number)	6. Shape/Maker, 7. Batch Size (5,10,15)	Julianne's Mom	Julianne	Zero Tolerance on the Batch Size	Trainging between two Dumpling Makers	Count the Number of Dumpings	Any broken Dumpling (record how many)
E	Cooking (Water Temp)	8. Initial Water Temp (75C, 85C, 95C), 9. Pan Size (Water Depth)	Christina	Brianna	Control Water Temp +/- 2C?	Define Infrared Measurement	Take Multiple Readings every 30 seconds?	Check Dumplings are fully cooked
F	Dumpling Duration (Rising time)	3 more mins after last piece rised up?	Allan	Julianne's Sister	Rising up time within 0.3 second between two operators	Measure the Pan Contact Area of the Bottom Surface	Count the duration of the last one rising up	Check all Dumplings are fully cooked
G	Traveller	Define Traveller Template or use Laptop	Mason	Charles	No Missing or typo Information	All A-F Categories are listed	Traveller with each Dumpling Plate Treatment	Check all informaiton filled by each Group Leader
H	JMP Data Entry		Patrick	JeAnne	No Typo on Data Entry	Create JMP DSD Datasheet	Check all information available on the traveller	Double Check the Data Entry (no trpo or missing)

Figure 2. Dumpling HACCP Method.

2.4 Artificial Intelligence

Two modern AI techniques Neural [10, 11] and Partitioning were utilized to help build the Dumpling Process Modeling. Neural modeling is based on a Black Box transformation to enhance the goodness fit of modeling. The partitioning modeling is to recursively split the data points in binary at each step to identify the critical parameters and its separation boundary. Both methods would be further addressed in the Result section as compared to classical regression models.

2.5 Mathematics

Most Mathematics involved are in Heat Transfer Physics equations, Neural TanH transformation, Partitioning Binary calculation, and DOE Statistics. Authors won't address the detailed mathematics in this paper. Instead, the focus would be on how to apply mathematics in the other STEAMS elements practically.

2.6 Statistics

Definitive Screening Design (DSD) is exploited in order to reduce your sample size (does all corner pairing to increase Power and orthogonality). Advantages of using DSD [12-16] are: (1) Small number of DOE runs, (2) Main effects Orthogonal (No Resolution II Confounding), (3) Main effects uncorrelated with 2-way Interactions (No Resolution III Confounding), (4) 2-Way Interactions are not fully confounded with each other (Resolution IV Confounding), (5) Estimable quadratic effects – in a three-level design (Non-Linear). DSD Data Collection plan and results are shown in Figure 3. Collect Information on: Meat Type, Meat Mass, Veggie Mass (Cabbage/Mushroom), Total Dumpling Mass, Batch Size, Water Temp. Collect 18 DSD runs and measure the response – Rising Time.

STEAMS methodology is interdisciplinary and is much powerful if all six STEAMS elements could be integrated seamlessly and complimentarily. The experimental results and modeling analysis would be demonstrated in the next result and discussion section.

Run No.	Randomized Run No	Ratio (meat%)	Meat Type	Meat (g)	Vegi Cabbage (g)	Vegi Mushroom (g)	Dumpling Weight (g)	Batch Size (Count)	Water Temp (deg C)	Dumpling Rising Time (min:s)
1	9	0.8	Pork	80	16	4	25	4	95	1:21
2	11	0.8	Shrimp	140	28	7	25	7	75	3:20
3	18	0.5	Shrimp	125	100	25	25	10	95	0:55
4	16	0.5	Shrimp	34	27.2	6.8	17	4	85	0:50
5	2	0.65	Pork	110.5	47.6	11.9	17	10	95	1:20
6	1	0.5	Pork	125	100	25	25	10	75	7:36
7	12	0.5	Shrimp	50	40	10	25	4	95	1:05
8	3	0.5	Pork	59.5	47.6	11.9	17	7	95	1:10
9	17	0.65	Shrimp	95.55	41.2	10.3	21	7	85	1:38
10	15	0.8	Shrimp	168	33.6	8.4	21	10	95	2:04
11	13	0.65	Shrimp	65	28	7	25	4	75	5:19
12	4	0.65	Pork	95.55	41.2	10.3	21	7	85	3:27
13	5	0.8	Pork	200	40	10	25	10	85	2:25
14	6	0.5	Pork	42	33.6	8.4	21	4	75	2:01
15	14	0.8	Shrimp	54.4	10.9	2.7	17	4	95	0:46
16	7	0.8	Pork	136	27.2	6.8	17	10	75	2:42
17	8	0.8	Pork	54.4	10.9	2.7	17	4	75	1:59
18	10	0.5	Shrimp	85	68	17	17	10	75	1:52

Figure 3. Dumpling DSD Data Collection Plan

3. Results and Discussion

Three JMP (Statistical Software) Modeling Data Analysis were conducted to study the Dumpling Process: (1) Stepwise Regression, (2) Neural Network, and (3) Recursively Partitioning. Three models were compared to discover more insights on Dumpling Physics Science.

3.1 Stepwise Regression Model

The experiment was built based on an orthogonal DSD structure which was good for building the Response Surface Methodology (RSM) model. Classical Stepwise Regression Model was conducted to derive the RSM model step by step by adding or/and removing the insignificant terms to enhance the model signal-noise ratio. The Stepwise model has observed both significant Main Effect and Interaction Effects which has indicated the Dumpling Process is complicated and coupling. The observed interaction effects may be due to two competing heat transfer physics or due to the intrinsic DSD confounding structure. Top four parameters are: (1) Water Temp, (2) Meat (g), (3) Vegi Cabbage, (4) Meat Type (Shrink or Pork) as shown in Figure 4. Batch Size is less important which may indicate the heat transfer was quite uniform and least impacted by Batch Size. Dumpling Weight factor was confounding with Meat and Vegetable.

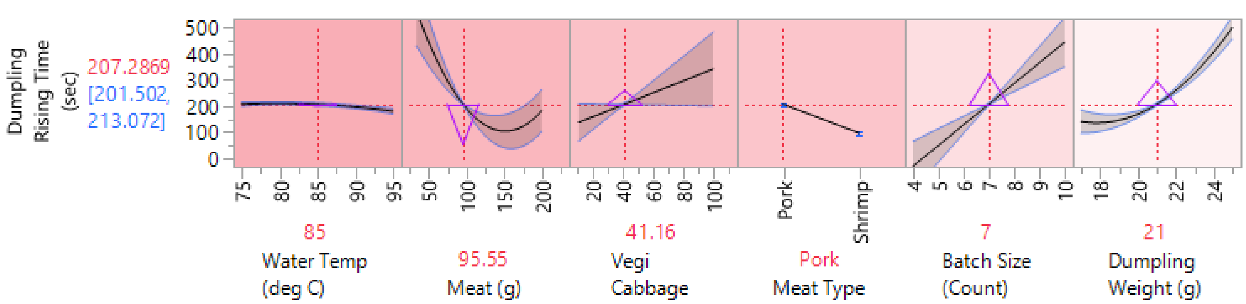

Figure 4. Stepwise Regression Model Sensitivity Analysis

To further illustrate the key cooking parameters, Figure 5 has shown the five different prediction profilers. The profiler can be utilized to predict the cooking duration at 95% confidence interval per each unique cooking recipe. The Dumpling business model should consider four performance criteria: (1) Faster Cooking Time, (2) Less Cooking Duration Variance, (3) Maximum Throughput, and (4) Power Saving. Though, this paper won't address the Dumpling Business Model.

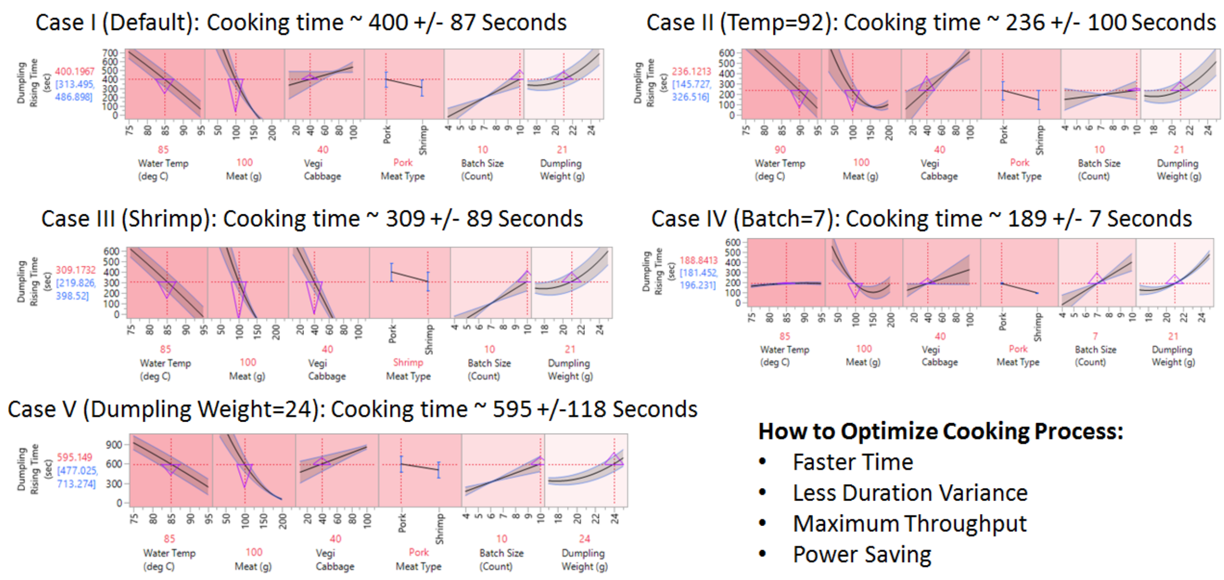

Figure 5. Five Case Study of Stepwise Regression Model Profiler.

3.2 Neural Network Model

The second modeling technique is Neural Network which is very powerful to add one or more middle layers to transfer the input variables through a TanH hyperbolic equation (Black Box). The Neural Modeling may get a higher Goodness of Fit if sample data size is larger. With the current 18 data points, it's curious to compare the Network model to the previous Stepwise RSM mode. The main drawback is Neural Network, after the transformation, won't maintain the intrinsic Physics and interaction effects in the eventual Neural Model which may make the model practitioners to interpret the scientific insights through the Neural Model. Top four parameters to predict the cooking duration are: (1) Water Temp, (2) Veggie Cabbage, (3) Dumpling Weight, and (4) Dumpling Size as shown in Figure 6. The Neural model sensitivity ranking is slightly different from the previous Stepwise RSM model. The Neural model would transform the input variables and empower the top 2 variables Water Temp and Veggie Cabbage. The Neural model may provide a better look on the R-Square. Though, would that make sense? This subject would be further addressed in later section 3.4 Model Comparison.

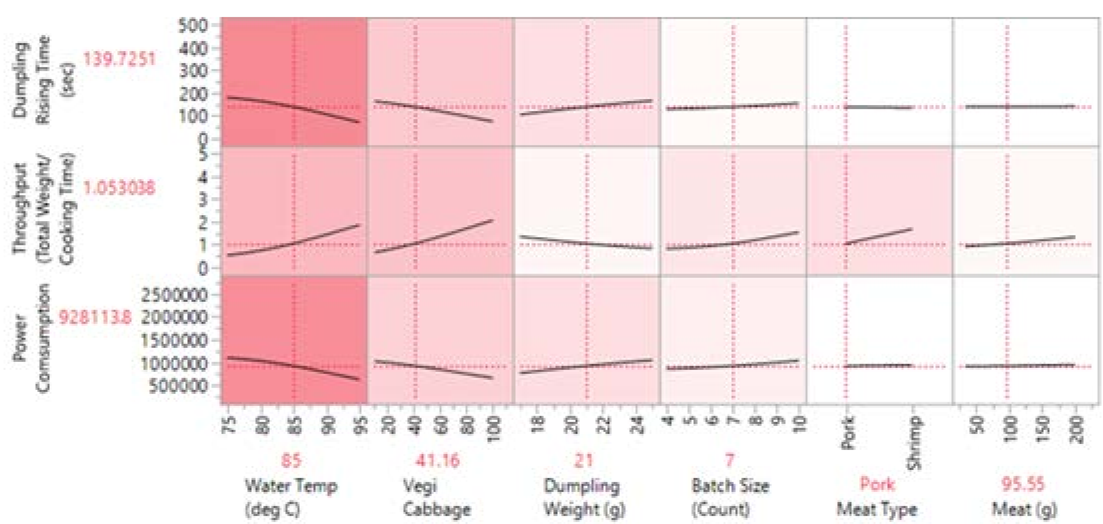

Figure 6. Neural Network Model

3.3 Recursively Partitioning Model

The third technique is a recursively partitioning model which would split the data points recursively through binary split. As shown Figure 7. The Water Temp is the dominator factor. Due to limited 18 data points, the discrete partitioning algorithm won't provide all modeling details like continuous stepwise or Neural modeling. Which modeling limitation can be indicated by a much lower R-Square. Therefore, the partitioning modeling is limited. Though, the partitioning model could identify the best split boundary of water temp at 85ºC. This discrete split method may actually overcome any Normality violation or overfit concerns from most continuous modeling. The model comparison would be addressed in the next section.

Faster Rising Time

RSquare	RASE	N	Number of Splits	AICc
0.726	53.188693	18	4	213.777

Water Temp (deg C)>=85

		LogWorth	Difference
Count	11		
Mean	92.818182	0.9985476	50.5714
Std Dev	48.799218		

Water Temp (deg C)<85

		LogWorth	Difference
Count	7		
Mean	212.71429	1.4889354	196.5
Std Dev	129.45748		

Figure 7. Partitioning Model

3.4 Model Comparison

Three models were compared in Figure 8. Several observations would be addressed below:

(1) All the models have identified the Water Temp as the top parameter to predict the cooking duration time. Though, three models have shown very different rankings after the top one.

(2) The Stepwise model has indicated a significant gap between main effect and total effect which has indicated the significant interaction effects may exist in the dumpling process. The Neural model has shown little gap between main effect and total effect which has transformed the model favored the main effects. The partitioning model could not separate the gap due to its discrete binary split algorithm.

(3) Among the ranking of six input parameters, the Stepwise model would provide a more uniform distribution pattern which may be due to complicated Heat Transfer Physics Mechanisms. Neural models have transformed the model and empowered the top parameter Water Temp to dominate the model. The partitioning would empower the top two factors water temp and dumping weight due to its limited split capacity.

(4) It seems the Stepwise RSM model could match the Physics Knowledge Domain and provide more process insights based on a well structured DSD experiment. Though, most time, there may be luxury to conduct the structured experiment and may only collect the observational data. Neural or Partitioning methods may win with larger observed sample runs which may not be structured. Classical Stepwise RSM models may not be effective to derive the interaction effects if data structure is not near orthogonal.

(5) Authors may suggest deriving all three models and conduct model comparison in details based on Scientific Knowledge Domain and Statistics Goodness Fit, Overfit risk management as demonstrated in this paper.

RSM (Stepwise)

Dumpling Rising Time (sec)

Column	Main Effect	Total Effect
Water Temp (deg C)	0.121	0.406
Meat (g)	0.079	0.354
Vegi Cabbage	0.103	0.306
Meat Type	0.114	0.263
Batch Size (Count)	0.034	0.151
Dumpling Weight (g)	0.065	0.075

Neural

Dumpling Rising Time (sec)

Column	Main Effect	Total Effect
Water Temp (deg C)	0.543	0.576
Vegi Cabbage	0.235	0.267
Dumpling Weight (g)	0.128	0.154
Batch Size (Count)	0.016	0.031
Meat Type	0.001	0.004
Meat (g)	2e-4	0.001

Partitioning

Column Contributions

Term	Number of Splits	SS	Portion
Water Temp (deg C)	2	68003.3016	0.5040
Dumpling Weight (g)	1	66192.4286	0.4905
Batch Size (Count)	1	744.047619	0.0055
Meat Type	0	0	0.0000
Meat (g)	0	0	0.0000
Vegi Cabbage	0	0	0.0000

Figure 8. Model Comparison

4. Conclusion

This paper has demonstrated the STEAMS methodology on the Dumpling Cooking application. Through the STEAMS method, the Heat Transfer Physics Science would be explored through literature research to identify the critical parameters during the dumpling process. The Technology section would emphasize on how to utilize the modern technologies to conduct a successful Dumpling experiment. The Engineering would focus on both the Engineering Project Management and Quality Control HACCP and ISO 22000 to ensure a smooth project management and successful engineering experiment. Two modern AI modeling techniques were adopted and compared against the Statistical Stepwise RSM model based on a structured DSD experiment. Three modeling techniques have drawn the same conclusion that Water Temp would be the most dominant factor to impact the dumpling cooking duration. Though, three models would also demonstrate different modeling patterns on the main effects and interaction effects. It would be powerful to compare three models and make a better judgement accordingly and collectively. The same STEAMS and Modeling Techniques could be applied to other Foods Processing fields and other Scientific Research applications.

Acknowledgements
The author would like to thank Dumpling team members Julianne Chiu, Alan Sean, and Brianna Zheng.

References
1. Rita Colwell, (2016) “Compare: Guest commentary: A "STEM" in Collier County to reach their future".
2. Charette, Robert N, (2013) "The STEM Crisis Is a Myth". *IEEE Spectrum*.
3. Chen, Mason (2019) “Introduce a Novel “STEAMS” Methodology of Conducting Scientific Research”, *IEEE FUZZ Proceedings*, P.320-325
4. Chen, Mason, (2019) “Apply “STEAMS” Methodology on Managing Europe Travel”, *ASA JSM Proceedings*, P.946-958
5. Chen Mason, (2020) “STEAMS” Methodology of Altitude Sickness and Fatigue Research”, *HUIC AHSE Proceedings*
6. Zhu Qiang (2015), “Dumpling Cooking – Modeling and Simulation” 9th International Symposium on Advanced Control of Chemical Processes, *The International Federation of Automatic Control*, Whistler, British Columbia, Canada
7. Andrey Varlamov, Zheng Zhou, Yan Chen, (2018) “Boiling, steaming or rinsing?”, *Physics of the Chinese Cuisine*
8. Introduction to HACCP: https://www.slideshare.net/Adrienna/introduction-to-haccp-57715003
9. HACCP Description Chart: https://www.researchgate.net/figure/HACCP-DESCRIPTION-CHART_tbl5_260401936
10. Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". *Neural Networks*. 61: 85–117
11. Ruslan, Salakhutdinov; Joshua, Tenenbaum (2012). "Learning with Hierarchical-Deep Models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 35 (8): 1958–71.
12. Errore, A., Jones, B., Li, W., and Nachtsheim, C. (2016), “Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects,” forthcoming, *Journal of Quality Technology*.
13. Jones, B. and Nachtsheim, C.J. (2011), “Efficient Designs with Minimal Aliasing,” *Technometrics*, 53:1, 62-71.
14. Jones, B. and Nachtsheim, C.J. (2013), “Definitive Screening Designs with Added Two-Level Categorical Factors,” *Journal of Quality Technology*, 45, 121-129.
15. Jones, B. and Nachtsheim, C.J. (2016), “Blocking Schemes for Definitive Screening Designs,” *Technometrics*, 58:1, 74-83.
16. Miller, A. and Sitter, R. R. (2005), “Using Folded-Over Nonorthogonal Designs,” *Technometrics*, 47(4), 502-513.