Spring 4-11-2022

Knowledge-informed Data-driven Modeling of Microbial Inactivation in Food

Steve Zhang
University of Nebraska-Lincoln, szhang35@unl.edu

Firnaaz Ahamed
University of Nebraska-Lincoln, firnaaz.ahamed@unl.edu

Hyun-Seob Song
University of Nebraska-Lincoln, hsong5@unl.edu

Knowledge-informed Data-driven Modeling of Microbial Inactivation in Food

Steve Zhang, Firnaaz Ahamed, Hyun-Seob Song

MOTIVATION

- Accurate modeling of microbial inactivation in food is a difficult task due its complex dependence on numerous internal and external food processing conditions.
- Past models rely on empirical methods, which lack generalization and quickly become unmanageable with increasing number of input variables.

Generalizable Pipeline for Data-driven Modeling of Complex Nonlinear Systems

- Sparse Identification of Nonlinear Dynamics (SINDy)[1] **automatically identifies the best and simplest equation** predicting **D-value** (the time taken to reduce microbial population to 10% of the initial level) from a function library calibrated with domain knowledge.
- Global Sensitivity Analysis (GSA) **identifies key process variables** and their interactions

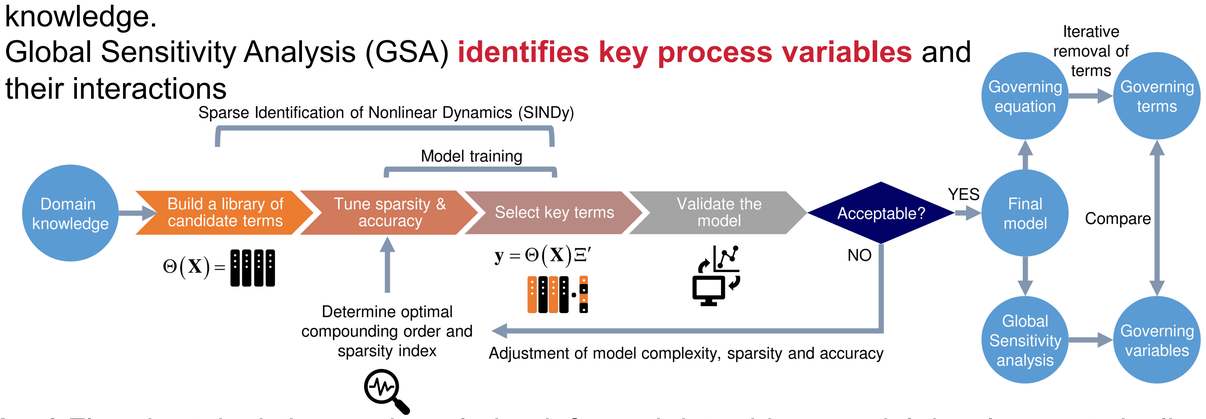

Fig. 1 Flowchart depicting our knowledge-informed data-driven model development pipeline.

Stepwise Tuning Balances Accuracy and Sparsity

- Statistical metrics mean-squared error (MSE) and Akaike information criterion (AIC) **penalize overfitting**.

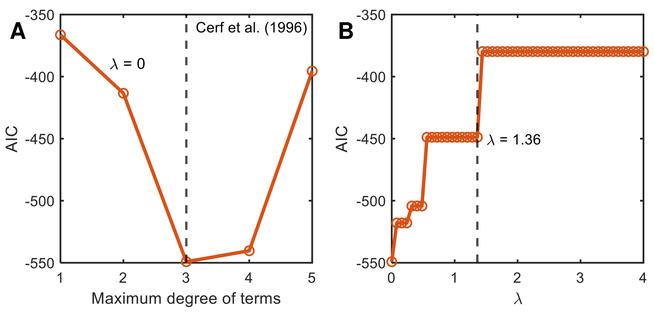

Fig. 2 Model tuning of (A) complexity followed by (B) sparsity using AIC.

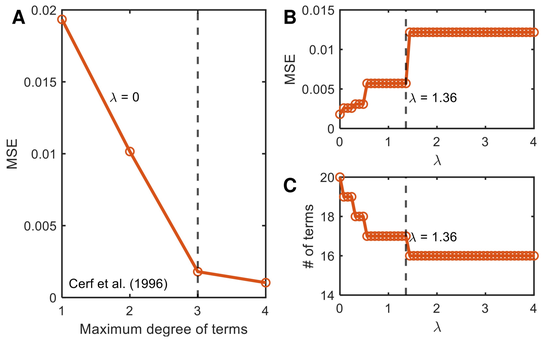

Fig. 3 Model tuning of (A) accuracy followed by (B) sparsity using MSE and (C) model equation's number of terms.

Our Model Shows Enhanced Accuracy and Minimizes Overfitting Compared to Literature Models

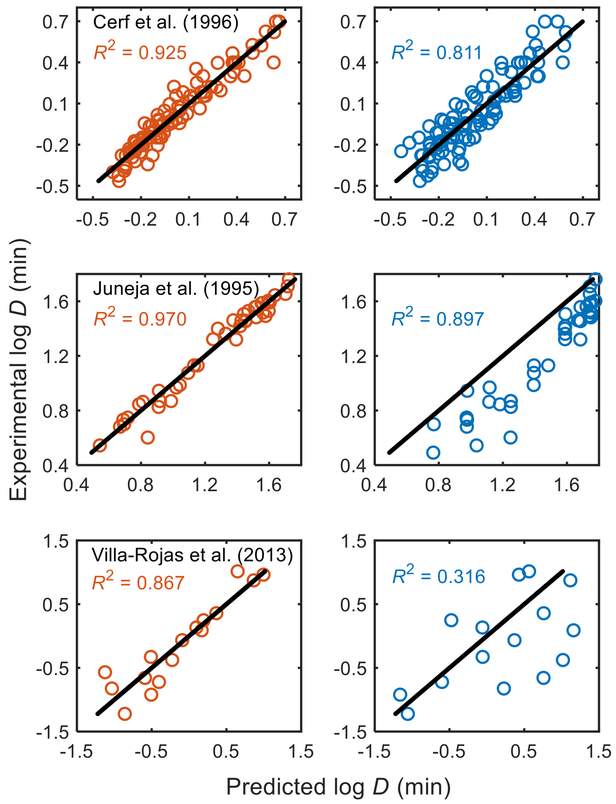

Fig. 4 Plots with predicted and observed log D values forming the data points' x and y values.

Ours		Theirs	
MSE	AIC	MSE	AIC
0.006	-448.85	0.014	-392.24
0.004	-193.46	0.071	-65.96
0.06	-36.96	0.559	27.26

Table 1 Comparison of MSE and AIC values between our model and literature models

Sensitivity Analysis Suggest Variables' Effects Are Highly Environmentally Dependent

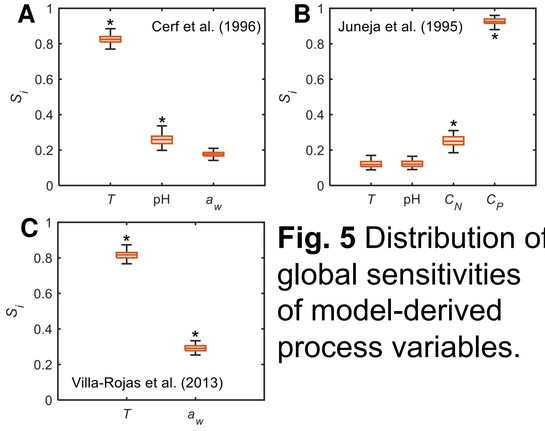

Fig. 5 Distribution of global sensitivities of model-derived process variables.

CONCLUSIONS

- GSA results are model-derived and contextual and shouldn't influence model selection.
- Combined use of data-driven modeling and GSA is also useful for rational model-based optimization of operating conditions and any non-linear systems for a wide range of applications.

REFERENCE

- [1] Brunton, S. L. et al. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. S. A. 113, 3932–3937.

University of Nebraska Lincoln