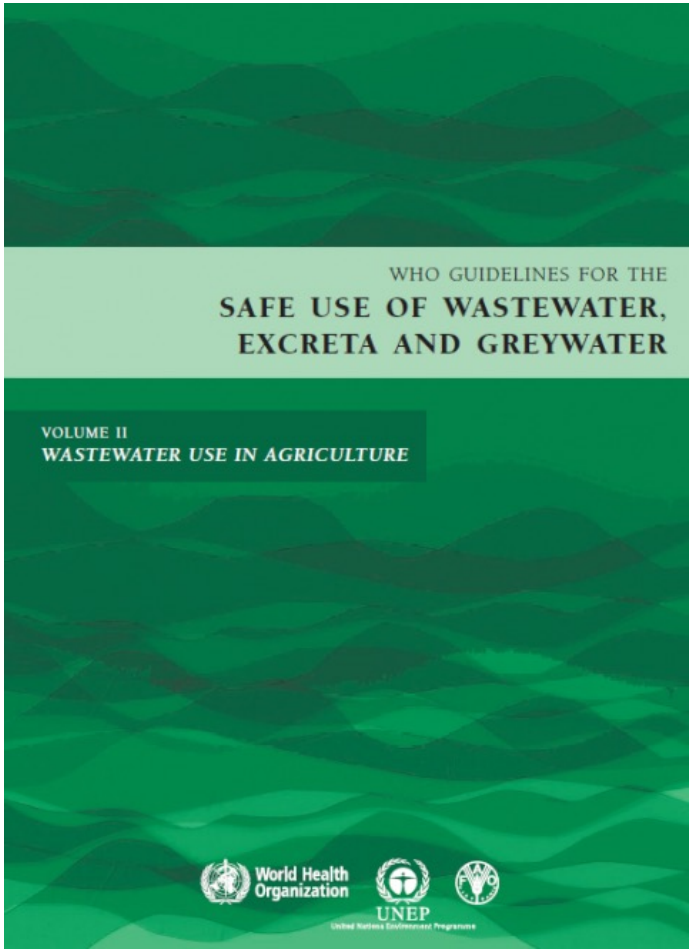
Regularized Regression Modeling of Rotavirus Disinfection in Wastewater for Predictive Environmental Microbiology in Sanitation Safety Planning

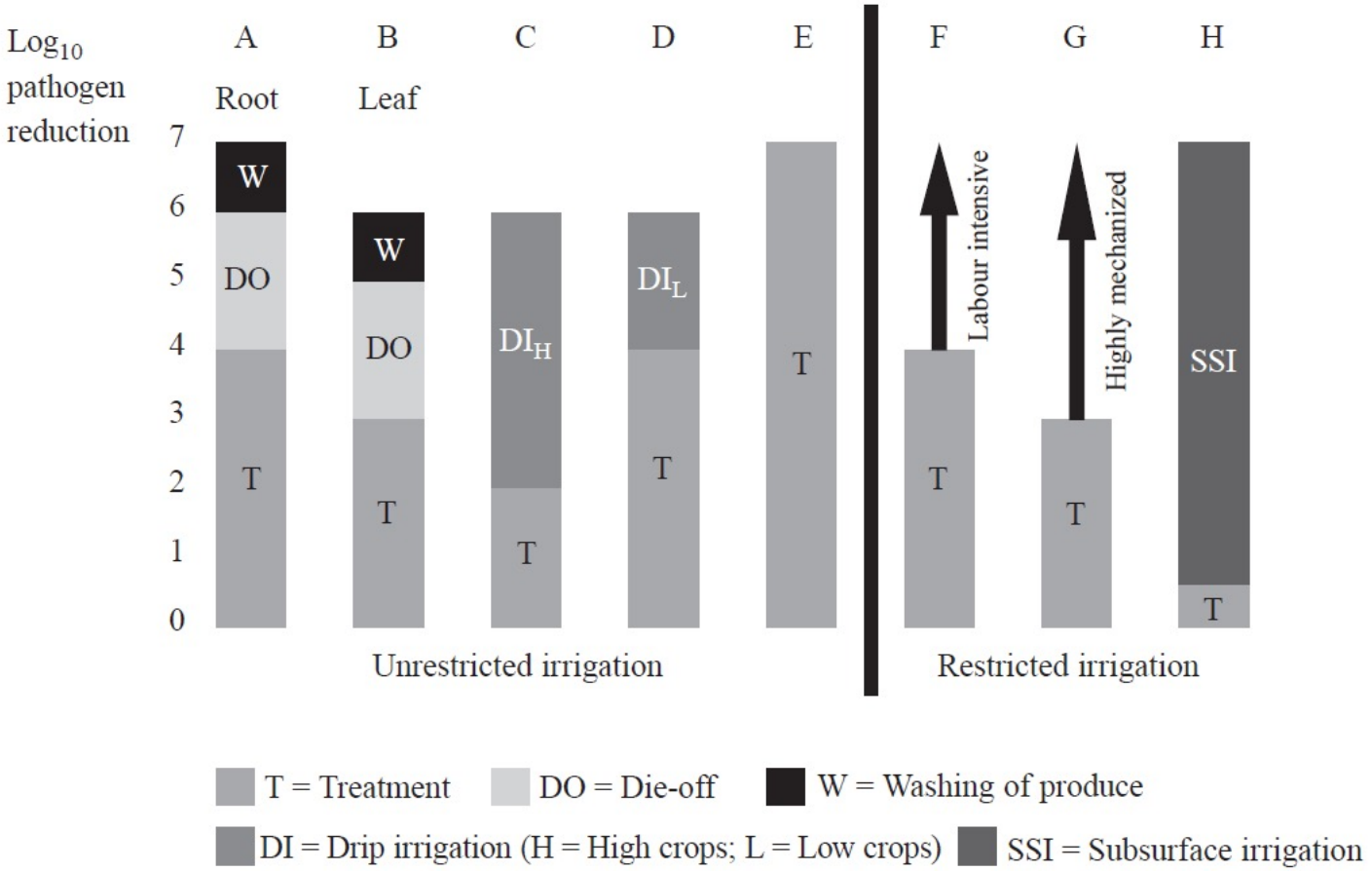
Daisuke Sano, PhD
Department of Civil and Environmental Engineering
Tohoku University, Japan

WHO guidelines for the safe use of wastewater…



- Determine the required degree of pathogen reduction, by QMRA, to achieve the tolerable disease burden, such as 10^{-6} disability adjusted life year per person per year ($DALY_{pppy}$)
- The required pathogen reduction is achieved by wastewater treatment alone or wastewater treatment in conjunction with other measures

Target $\log_{10}$ reduction value (LRV)



Ref. WHO guideline for the safe use of wastewater, 2006

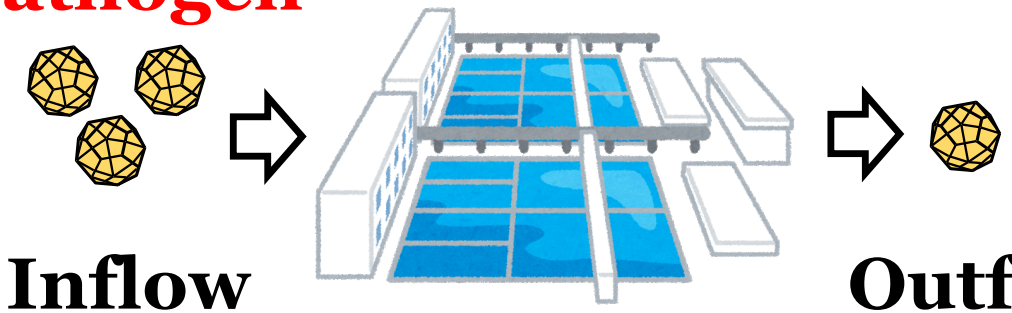
Sanitation Safety Planning (SSP)

A scheme for the safe use of excreta, wastewater and greywater (World Health Organization, 2016)

Hazard Analysis and Critical Control Point (HACCP)

WWTP

Pathogen



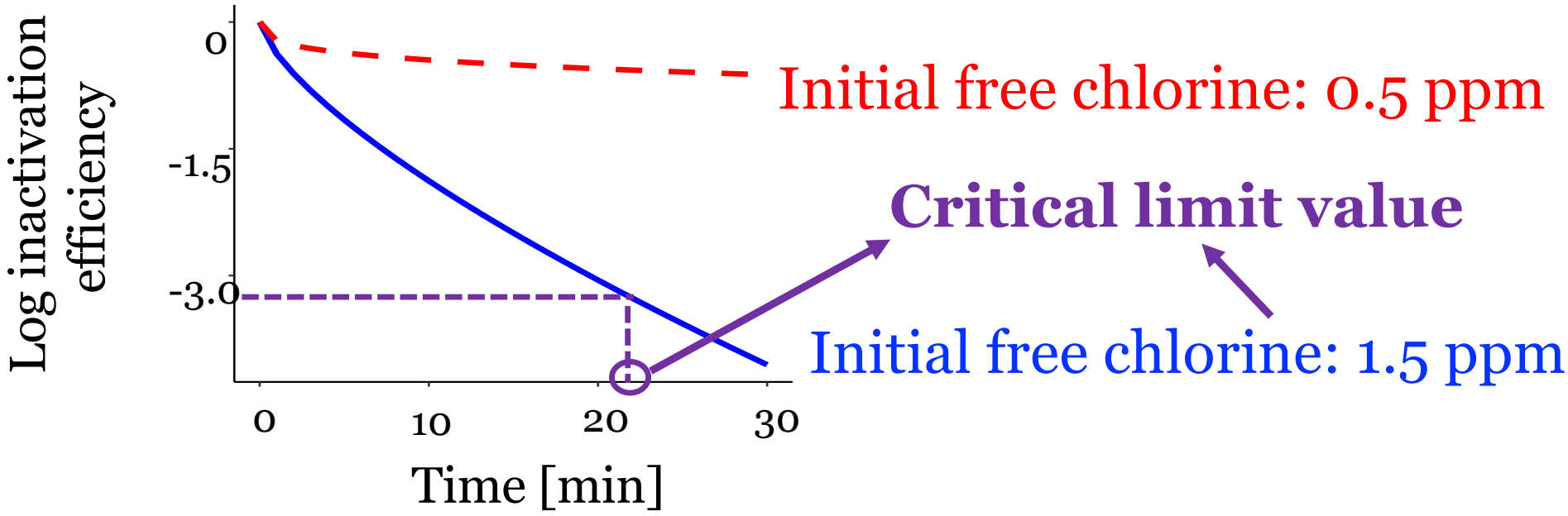
Inflow **Outflow**

Critical limit values are:
Disinfectant concentration
Contact time

Environmental Predictive Microbiology

How can a critical limit value be determined?

Inactivation model (Hom model) $Log(N_t/N_0) = -kC_0t^m$

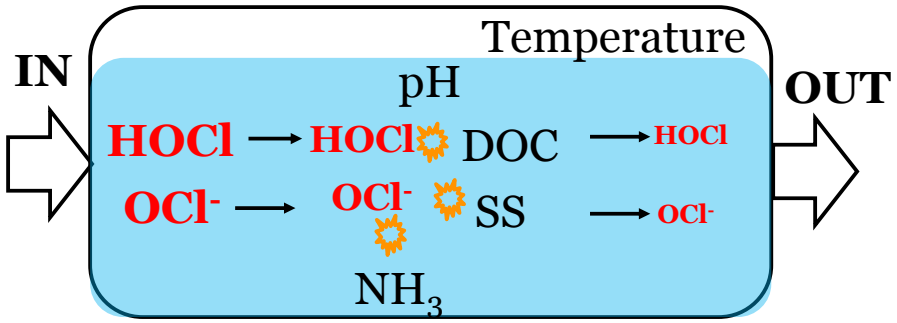


Issues in Environmental Predictive Microbiology

Variety of water quality among WWTPs

pH, Temperature, DOC,...

Disinfection reactor



Disinfectant decay

Genotype/Strain-dependent sensitivity of viruses to disinfectants

- Among genotypes

  Chlorine sensitivity of poliovirus Brunhilde strain vs Mahony strain (Sharp and Leong, 1980)

- Within a genotype

  Murine norovirus (Rachmadi *et al.*, 2018)

  MS2 phage and Echovirus (Zhong *et al.*, 2016 & 2017)

Study Objective

Construction of a framework for determining the CL value of the critical control point (CCP) that can ensure the achievement of target $\log_{10}$ reduction in a disinfection unit process of wastewater reclamation

Study Overview

1. Systematic review: Collection of inactivation efficiency & water quality data

2. Modeling with Regularized Regression Analyses: Construction of combined chlorine models for enterovirus based on literature values

3. Construction of combined chlorine concentration models for calculating decay constant: Chlorine decay data were obtained and used

4. Model verification: Inactivation experiment using Enterovirus 71 (EV71)

5. CL value determination: Calculate LRV using the combined chlorine disinfection model and determine the CL value of critical control point

Regularized Regression Analysis

$$min_\beta S_\lambda = min_\beta \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda R(\beta) \right\}$$

λ : regularization parameter(≥ 0)

$R(\beta)$: regularization term(≥ 0)

If $\lambda = 0$, the formula is equal to the OLS

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + \cdots$$

Ridge

$$R(\beta) = \frac{1}{2}\|\beta\|_2^2 \qquad (\|\beta\|_2 = \sum_{i=1}^{n} {\beta_i}^2)$$

Sparse estimation

Lasso

$$y = x_1\beta_1 + x_2 0 + x_3 0 + x_4 0 + \cdots$$

$$R(\beta) = \|\beta\|_1 \qquad (\|\beta\|_1 = \sum_{i=1}^{n} |\beta_i|)$$

Elastic net

$$y = x_1\beta_1 + x_2 0 + x_3\beta_3 + x_4 0 + \cdots$$

$$R(\beta) = \alpha\|\beta\|_2^2 + (1-\alpha)\|\beta\|_1$$

α determines the proportion of ridge to lasso

Bayesian Regression Analysis

Bayesian ridge

Use all variables as well as Ridge regression

prior distribution

$$p(\beta|f) = N\left(\beta_i|0, {f_i}^{-1}\right)$$

$$p(\beta|f) = N(\beta_i|0, A^{-1})$$

$$\left(diag(A) = f = \{f_1, \cdots, f_p\}\right)$$

Automatic relevance determination (ARD)

Regularization parameters

$$p(f) = \prod_{i=0}^{\rho} Gamma(f_i|a, b)$$

$$p(\sigma^2) = Gamma(\sigma^2|c, d)$$

Variables

Objective variable y

$$LRV = -log_{10}(N_t/N_0)$$

LRV：Log10 reduction value

N_o：intial virus concentration

N_t：virus concentration at contact time t[min]

Virus：enterovirus

Disinfectant: Combined chlorine

Explanatory variables x

Disinfection conditions

Conc: initial disinfectant concentration (mg/L)

k: disinfectant decay constant（min^{-1}）

t: contact time（min）

$LogCt$: Log_{10} Ct-value（min・mg/L）

Water quality

pH: pH (-)

$Temp$: temperature (˚C)

$Assay$: assay type (PFU/TCID50/qPCR)

WQ: water type (1:environ, 0:pure water)

Validation process

$$LRV \sim \beta 1 conc + \beta 2 k + \beta 3 t + \beta 4 LogCt + \cdots$$

Data split

The datasets was split into 80% for training and 20% for test. This was repeated 100 times.

Standardization

Regularized regression requires the standardization of explanatory variables

Feature Engineering

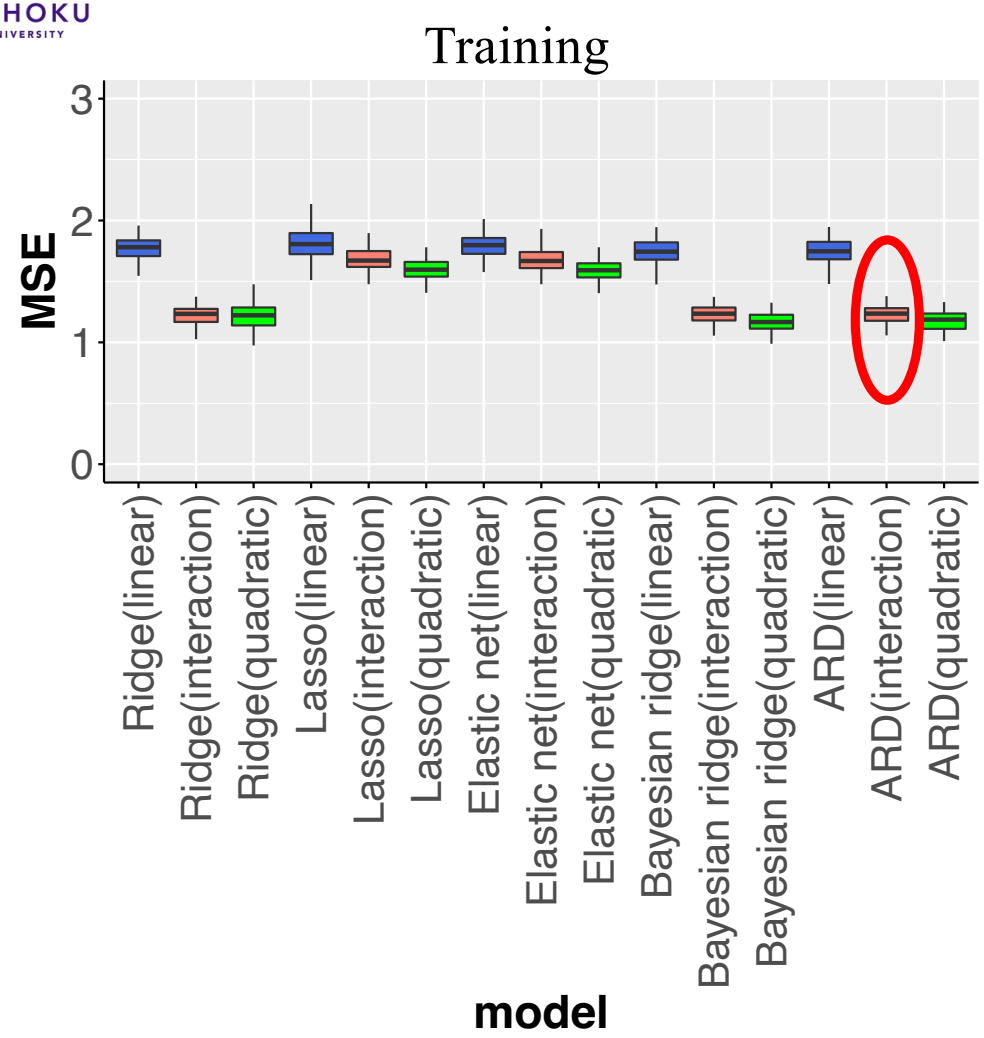
Polynomial terms (interaction and quadratic) were added and the prediction performance was compared.

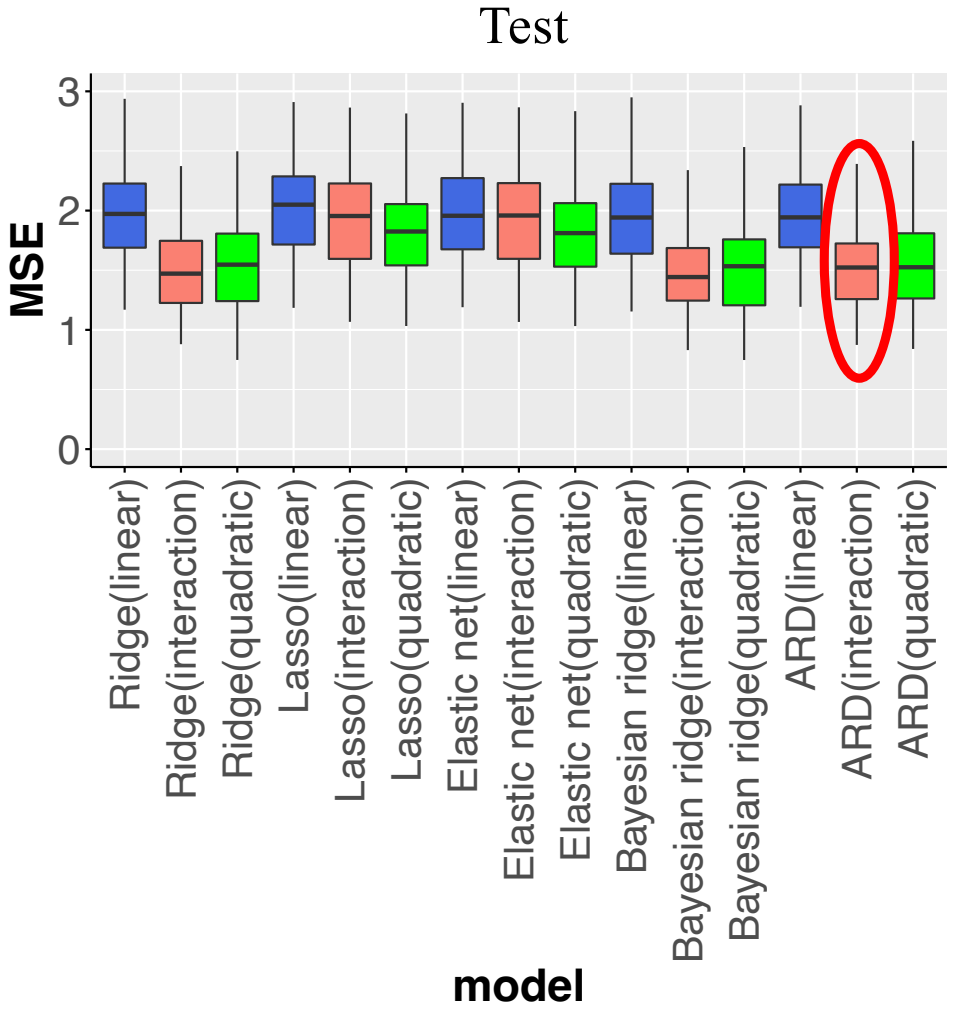
Regularized/Bayesian Regression

Estimate the regression coefficients

Compare the prediction performance with MSE and R^2

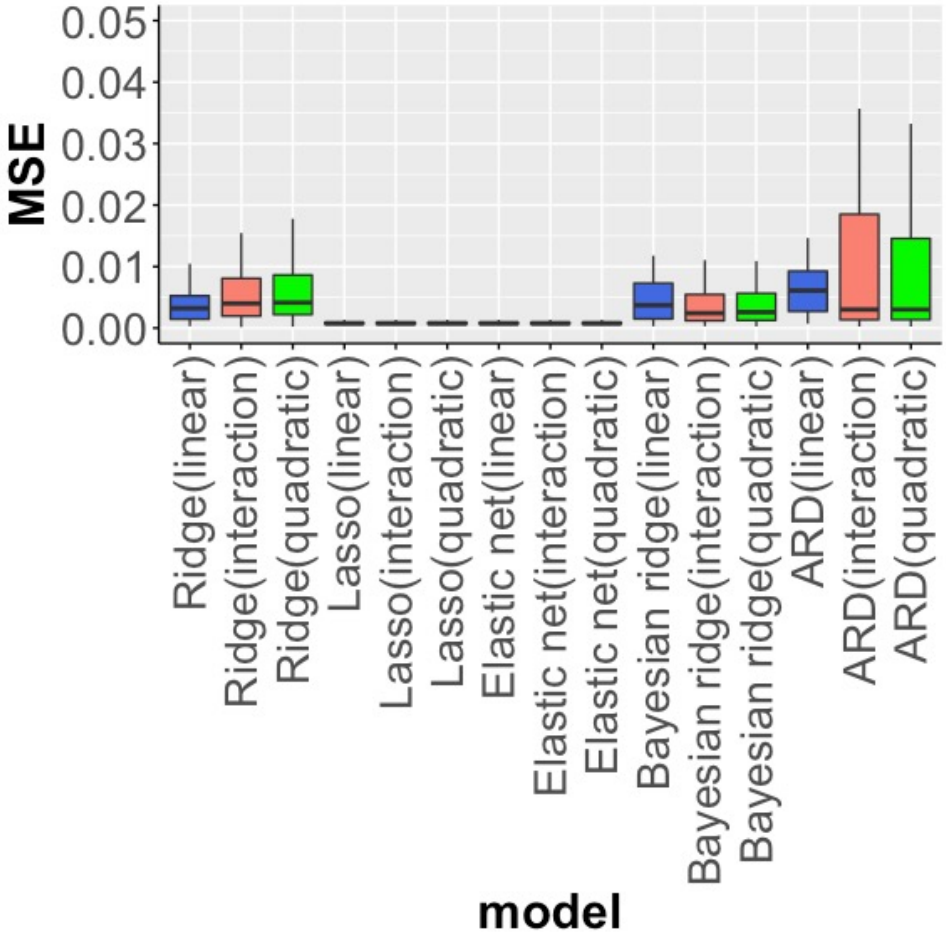
Model selection: Inactivation



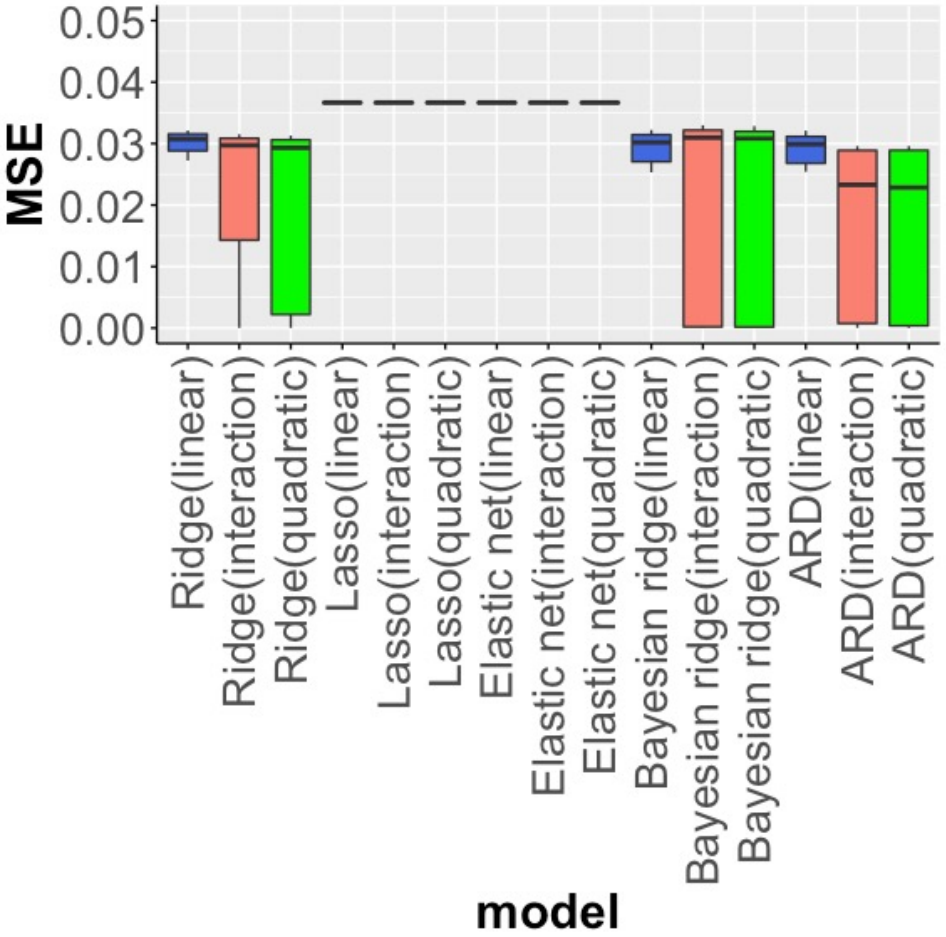


Model selection: Combined chlorine decay

Training

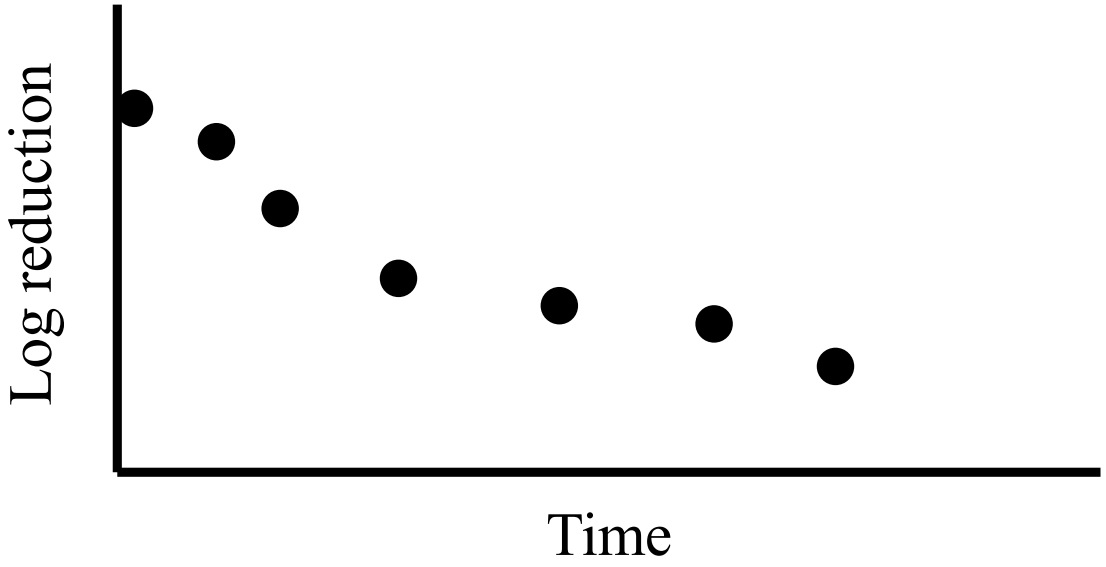


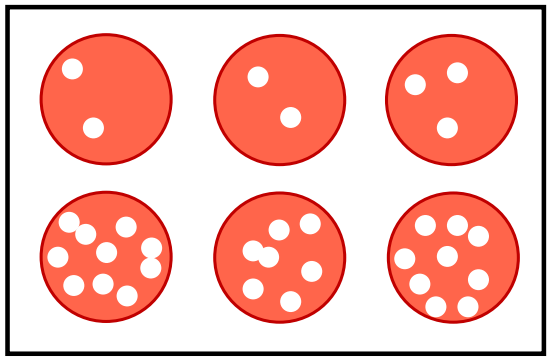
Test



Model verification

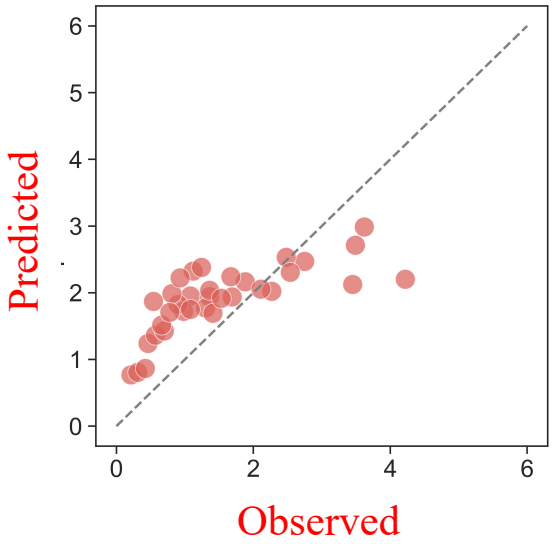
Chlorine disinfection experiment was performed using EV71 and secondary treated wastewater

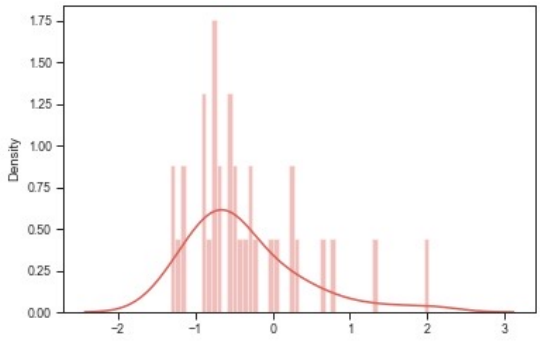




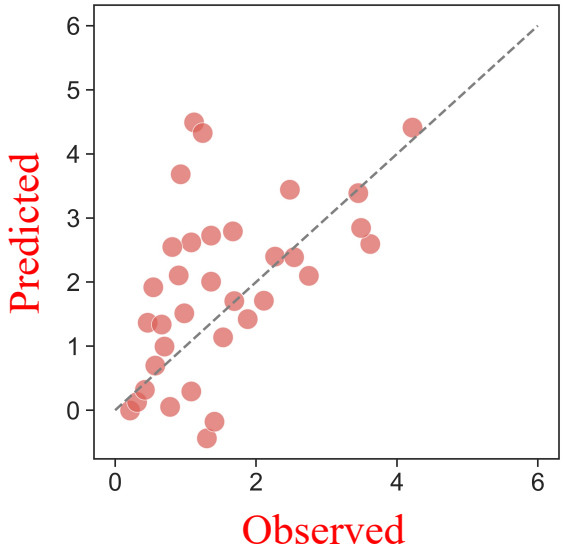
Model verification

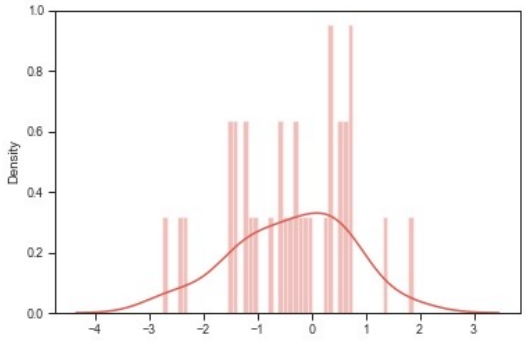
Ridge MSE=0.679



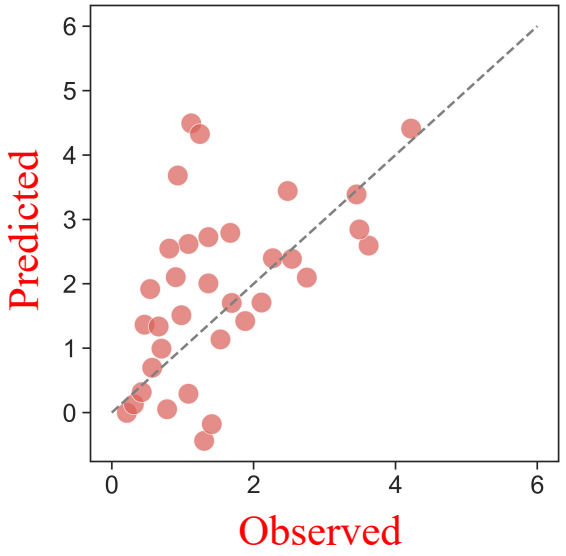


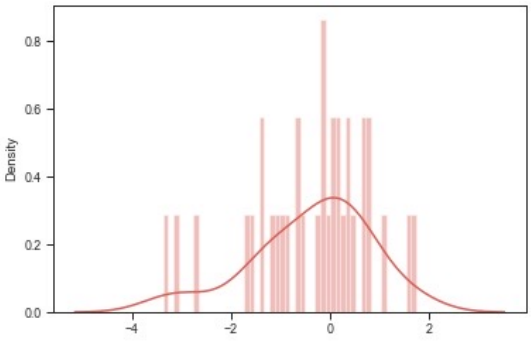
Bayesian Ridge MSE=1.266



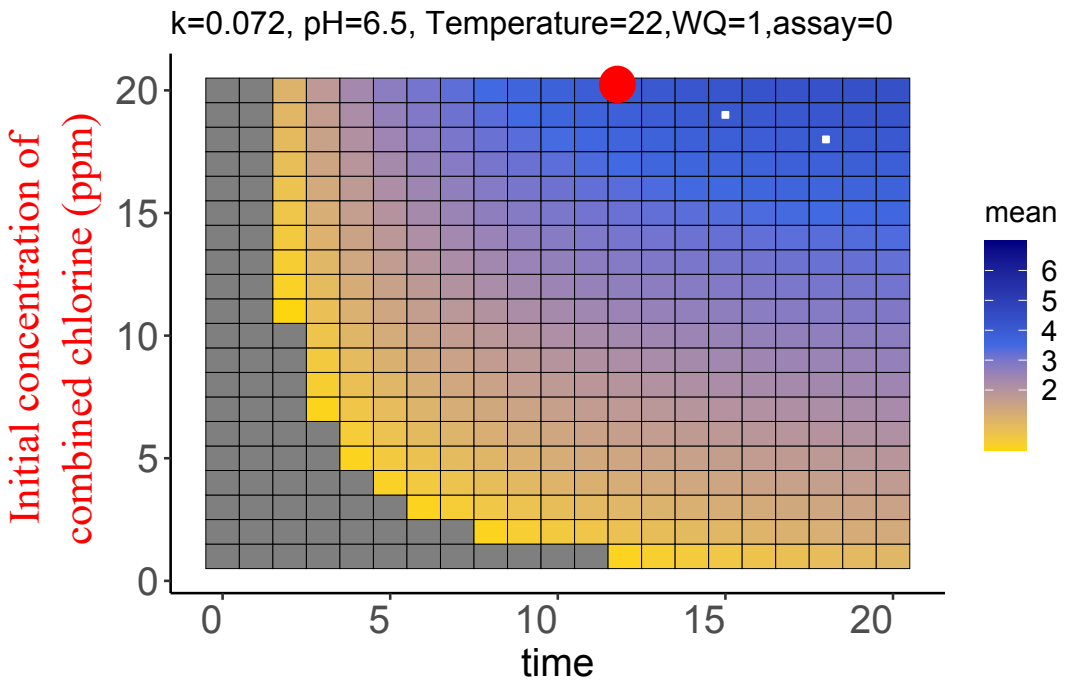


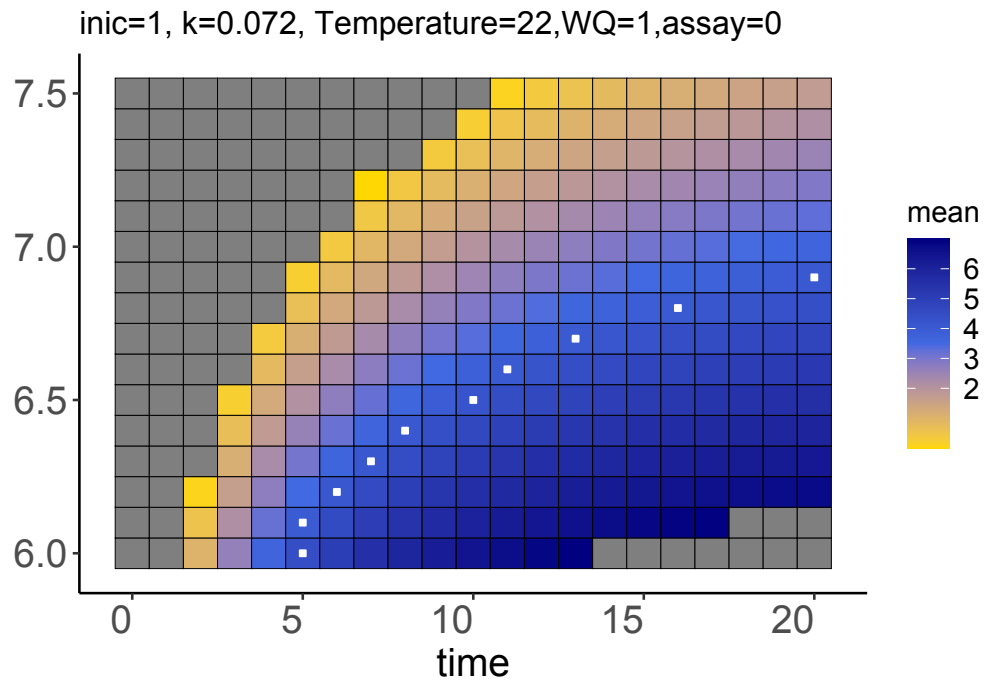
ARD MSE=1.548





Model verification





TOHOKU UNIVERSITY

Conclusions

Virus disinfection model was developed using a regularized regression, which allowed us to determine a critical limit value under various water quality conditions.

The accuracy of chlorine decay model needs to be improved.