

MODELING OXYGEN CONCENTRATION IN
WASTEWATER-IRRIGATED SOIL

By

Shuai Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biosystems Engineering – Master of Science

2013

ABSTRACT

MODELING OXYGEN CONCENTRATION IN WASTEWATER-IRRIGATED SOIL

By

Shuai Zhang

Food processing wastewater generally contains non-toxic materials and is sometimes applied to soil by irrigation as a treatment method. However, the biochemical oxygen demand (BOD) is relatively high. In an aerobic soil environment (i.e., the oxygen in soil is greater than 10% and aerobic microorganisms are active) these pollutants can be effectively assimilated with no environmental impacts. When wastewater flows through the anaerobic soil environment, metals such as Fe and Mn in soil replace the free oxygen to become electron acceptors. These metals are dissolved into the wastewater and can cause groundwater pollution. Strategic loading time, concentration, and frequency may allow for the disposal of wastewater while protecting groundwater from metal leaching. The objective was to identify the conditions in which aerobic soil environment can be maintained. Three statistical models (response surface methodology, mixed-effects model, and time series analysis) were developed to predict oxygen concentration in soil based on data from previously conducted soil column experiments. The developed models were compared using Akaike information criterion and Bayesian information criterion. Among the studied models, the mixed-effects model was the best fit model, and the conditions to keep soil aerobic were determined using response surface and contour plots.

ACKNOWLEDGMENTS

I would like to thank everyone who has helped me during my graduate studies. First of all, I would like to thank Dr. Pouyan Nejadhashemi to give me the opportunity to study in this fabulous lab and university. Under his guidance, I become more motivated and knowledgeable. Moreover, I would like to thank Dr. Steve Safferman for his support and advice to help me to continue my graduate study and do my interested research. Also, I would like to thank my committee member, Dr. Wei Zhang, and the visiting instructor, Mr. Steve Miller, for the thesis accomplishment and the research project processing, as well as Ryan Julien and Leila Saber, the students who performed the lab experiments.

I want to specifically thank my mom and dad, Lan and Hou, for support, advice, and unlimited love to acquire this master degree. I also want to thank my friends. Without their understanding and assistance, my graduate study life would be much more difficult.

The last but not the least, I would like to thank my labmates, Umesh, Sean, Yaseen, Giri, Edwin, Mike, Georgina, Melissa, Matt, and Fariborz for all of our happy times in the past.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
KEY TO ABBREVIATIONS.....	xii
1. INTRODUCTION	1
2. LITERATURE REVIEW	4
2.1 OVERVIEW	4
2.2 MICROBES IN FOOD PROCESSING WASTEWATER.....	4
2.2.1 Aerobic Microorganisms.....	4
2.2.2 Anaerobic Microorganisms.....	5
2.2.3 Free Oxygen in Soil	5
2.3 MODELS FOR OXYGEN CONCENTRATION.....	6
2.3.1 Physical Models	6
2.3.1.1 TOUGHREACT	6
2.3.1.2 HYDRUS.....	7
2.3.1.3 PHREEQC	8
2.3.2 Statistical Simulations	8
2.3.2.1 Response Surface Methodology	9
2.3.2.2 Mixed-Effects Model.....	11
2.3.2.3 Time Series Analysis	12
3. METHODOLOGY AND RESULTS	14
3.1 EXPERIMENT DESIGN.....	14
3.2 TECHNIQUES TO SIMULATE OXYGEN CONCENTRATION IN SOIL	28
3.2.1 TOUGHREACT	28
3.2.1.1 Governing Equations	28
3.2.1.2 Input and Output.....	29
3.2.1.3 Summary.....	31
3.2.2 HYDRUS	32
3.2.2.1 Governing Equations	32
3.2.2.2 Input and Output.....	33
3.2.2.3 Summary.....	36
3.2.3 PHREEQC	36
3.2.3.2 Governing Equations	36
3.2.3.3 Input and Output.....	37
3.2.3.4 Summary.....	38
3.2.4 Response Surface Methodology.....	39
3.2.4.1 Governing Equations	39
3.2.4.2 Input and Output.....	41
3.2.4.3 Comparative Methods.....	48

3.2.4.4	Result	52
3.2.5	Mixed-Effects Model.....	80
3.2.5.1	Governing Equation.....	80
3.2.5.2	Input and Output.....	84
3.2.5.3	Result	87
3.2.6	Time Series Analysis.....	89
3.2.6.1	Governing Equation.....	89
3.2.6.2	Input and Output.....	91
3.2.6.3	Discussion.....	99
3.3	MODEL COMPARISON	100
3.3.1	Criterion for Model Fitness	100
3.3.2	Comparative Results under Submerged Condition	102
3.3.3	Comparative Result under Submerged Condition.....	103
4.	CONCLUSION.....	104
5.	FUTURE RESEARCH.....	107
	APPENDIX.....	109
	REFERENCES	121

LIST OF TABLES

Table 1: Content of the synthetically produced organic loading	16
Table 2: Sand properties	17
Table 3: The loading conditions of acceptable data.....	20
Table 4: Input of TOUGHREACT	30
Table 5: Output of TOUGHREACT.....	31
Table 6: Input of HYDRUS	34
Table 7: Output of HYDRUS	35
Table 8: Input of PHREEQC	38
Table 9: Output of the full RSM under submerged condition	43
Table 10: Output of the full RSM under un-submerged condition.....	44
Table 11: RSM sub-model comparisons under submerged condition	52
Table 12: RSM sub-model comparisons under un-submerged condition.....	66
Table 13: Output of the full mixed-effects model under submerged condition.....	86
Table 14: Output of the full mixed-effects model under un-submerged condition	87
Table 15: Mixed-effects sub-model comparisons under submerged condition	88
Table 16: Mixed-effects sub-model comparisons under un-submerged condition.....	88
Table 17: Data used for model calibration.....	95
Table 18: Regression statistics for the model created using two experimental conditions and three time series-estimated parameters	96
Table 19: Data used for model validation.....	98
Table 20: Model comparison under submerged condition	103
Table 21: Model comparison under un-submerged condition	103

LIST OF FIGURES

Figure 1: Dimensions of the (a) long column, (b) short column, and (c) bottom of the column (all in cm). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.	15
Figure 2: Unacceptable data for Column 2 in phase A with a 7 g BOD/m ² /day loading concentration and 6 hour loading frequency under submerged condition	19
Figure 3: Unacceptable data for Column 3 in phase A with a 28 g BOD/m ² /day loading concentration and 12 hour loading frequency under submerged condition	20
Figure 4: Acceptable data for Column 1 in phase A with a 7 g BOD/m ² /day loading concentration and 6 hour frequency under submerged condition	21
Figure 5: Acceptable data for Column 4 in phase A with a 28 g BOD/m ² /day loading concentration and 12 hour frequency under un-submerged condition	21
Figure 6: Acceptable data for Column 7 in phase A with a 28 g BOD/m ² /day loading concentration and 12 hour frequency under submerged condition	22
Figure 7: Acceptable data for Column 8 in phase A with a 28 g BOD/m ² /day loading concentration and 12 hour frequency under un-submerged condition	22
Figure 8: Acceptable data for Column 6 in phase B with a 112 g BOD/m ² /day loading concentration and 6 hour frequency under un-submerged condition	23
Figure 9: Acceptable data for Column 7 in phase B with a 112 g BOD/m ² /day loading concentration and 12 hour frequency under submerged condition	23
Figure 10: Acceptable data for Column 8 in phase B with a 112 g BOD/m ² /day loading concentration and 12 hour frequency under un-submerged condition	24
Figure 11: Acceptable data for Column 6 in phase C with a 112 g BOD/m ² /day loading concentration and 24 hour frequency under un-submerged condition	24
Figure 12: Acceptable data for Column 7 in phase C with a 112 g BOD/m ² /day loading concentration and 56 hour frequency under un-submerged condition	25
Figure 13: Acceptable data for Column 8 in phase C with a 112 g BOD/m ² /day loading concentration and 24 hour frequency under un-submerged condition	25

Figure 14: Oxygen concentration response surface plot under the submerge condition showing the influence of soil depth and loading time (constant loading frequency and concentration)54

Figure 15: Oxygen concentration contour plot under the submerge condition showing the influence of soil depth and loading time (constant loading frequency and loading concentration)55

Figure 16: Oxygen concentration response surface plot under the submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)56

Figure 17: Oxygen concentration contour plot under the submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)57

Figure 18: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)58

Figure 19: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)59

Figure 20: Oxygen concentration response surface plot under the submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)60

Figure 21: Oxygen concentration contour plot under the submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)61

Figure 22: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)62

Figure 23: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)63

Figure 24: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)64

Figure 25: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)65

Figure 26: Oxygen concentration response surface plot under the un-submerged condition showing the influence of the soil depth and loading time (constant loading frequency and loading concentration)67

Figure 27: Oxygen concentration contour plot under the un-submerged condition showing the influence of the soil depth and loading time (constant loading frequency and loading concentration)68

Figure 28: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading frequency and loading time (soil depth and loading concentration)69

Figure 29: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)70

Figure 30: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)71

Figure 31: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)72

Figure 32: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)73

Figure 33: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)74

Figure 34: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)75

Figure 35: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)76

Figure 36: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)77

Figure 37: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)78

Figure 38: Periodogram of 112 g BOD/m ² /day loading concentration, 56 hour loading frequency, and 30.48 cm soil depth	92
Figure 39: Periodogram of 112 g BOD/m ² /day loading concentration, 24 hour loading frequency, and 111.76 cm soil depth	93
Figure 40: Illustration of seasonal decomposition for time series	94
Figure 41: Linear model from time series analysis.....	97
Figure 42: Validation for linear model from time series analysis	99
Figure 43: Validation for linear model from time series analysis	99
Figure 44: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 24 hour frequency, 10.16 cm soil depth.....	110
Figure 45: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 24 hour frequency, 30.48 cm soil depth.....	111
Figure 46: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 24 hour frequency, 50.80 cm soil depth.....	112
Figure 47: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 24 hour frequency, 71.12 cm soil depth.....	113
Figure 48: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 24 hour frequency, 91.44 cm soil depth.....	114
Figure 49: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 56 hour frequency, 10.16 cm soil depth.....	115
Figure 50: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 56 hour frequency, 30.48 cm soil depth.....	116
Figure 51: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 56 hour frequency, 50.80 cm soil depth.....	117
Figure 52: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 56 hour frequency, 71.12 cm soil depth.....	118
Figure 53: Seasonal decomposition for time series at 112.08 g BOD/m ² /day concentration, 56 hour frequency, 91.44 cm soil depth.....	119

Figure 54: Seasonal decomposition for time series at 112.08 g BOD/m²/day concentration, 56 hour frequency, 111.76 cm soil depth.....120

KEY TO ABBREVIATIONS

AIC	Akaike information criterion
BIC	Bayesian information criterion
BOD	Biochemical oxygen demand (mg/L)
L	Sub-model only including the linear variable
L+I	Sub-model only including the linear and interaction variables
L+Q	Sub-model only including the linear and quadratic variables
L+I+Q	Full model including the linear, quadratic and interaction, and variables
MAPE	Mean Absolute Percentage Error
ORP	Oxidation reduction potential (V)
RSM	Response Surface Methodology
stl	Seasonal Decomposition of Time Series by Loess
VMC	Volumetric water content (%)

1. INTRODUCTION

In the United States, 1.4 billion liters of wastewater is produced by roughly 20000 food industries every year (Elitzak, 2000; Oh and Logan, 2005). This wastewater contains very little to no toxic materials (Oh and Logan, 2005). However, in comparison with the wastewater produced by other industries, the biochemical oxygen demand (BOD) is relatively high due to the presence of simple sugars and starch (Oh and Logan, 2005). When this wastewater, which is high in organic content, is used for irrigation, organics are removed by the soil due to filtration, adsorption, and biodegradation. The rate of biodegradation is faster with a higher oxygen reduction potential (ORP); occurring when there is an abundance of oxygen in the soil. When the oxygen concentration is greater than 10%, the soil environment is assumed to be aerobic (King et al., 1998). In aerobic respiration, oxygen is the electronic acceptor. If organic materials remain in the water as it moves downward past the aerobic soil, biological activity still takes place under anaerobic conditions if enough nutrients are available. Anaerobic activity is significantly slower (King et al., 1998). Furthermore, when oxygen is not present and the ORP is low, metals such as iron and manganese in soil can serve as the electronic acceptors instead of oxygen. When reduced, these metals become soluble and can pollute groundwater (Brown and Caldwell, 2007).

In order to protect groundwater from metal contamination, reduce BOD, and save cost from treating food processing wastewater, aerobic conditions should be maintained until the carbon in the wastewater is completely oxidized. Keeping the soil aerobic

depends on soil type, soil depth, loading concentration, loading frequency, and loading time. However, no physical modeling tools are available to directly simulate this process. Several modeling tools were evaluated to simulate oxygen levels in soils without success. TOUGHREACT and HYDRUS simulate solute movements with multiple compositions in porous media based on a mass balance and Darcy's law (Simunek et al., 1999; Xu et al., 2006). These models are incapable of relating solutes to oxygen content in soil, and therefore, oxygen distribution cannot be predicted. A model that has been extensively used to predict gas fate is PHREEQC. PHREEQC is a geochemical model that simulates a large number of equilibrium reactions of water, minerals, ion exchangers, surface complexes, solid solutions, and gases. Additionally, PHREEQC can simulate non-equilibrium reactions of dissolution and precipitation of minerals, reactions of microorganisms, decay of organic complexes, and other reactions such as kinetic, solid-solution equilibrium, and fixed-volume gas-phase equilibrium (Charlton and Parkhurst, 2011). This model estimates the oxygen content in soil using water flow, solute, pressure, and the chemical reactions between the factors. In our study, total moles and total pressure of all gases in air, and partial pressure of free oxygen were not known. Consequently, the equilibrium of gas phase was not detectable, rendering PHREEQC inapplicable.

Due to the lack of a robust technique to predict the oxygen diffusion in soil, statistical simulations were used to make predictions. First, response surface methodologies (RSM) and mix-effects models were utilized using a second-order polynomial regressions technique. It was hypothesized that the oxygen content is affected by soil depth and the wastewater's organic loading duration, organic loading

concentration, and frequency of application. A best fit model was identified according to Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Ultimately, the oxygen concentration with any loading conditions at any depth within the condition range of our study was able to be calculated by the developed model. The suggested conditions to keep soil aerobic with an oxygen concentration greater than 10% were determined through response surface plots and contour plots based on RSM. Time series analysis including periodogram and seasonal decomposition were used to discover the conditions to sustain aerobic conditions in the soil. The frequency, amplitude, and frequency of oxygen changed were predicted by time series analysis. The problem was that this method was unable to be applied to all research conditions. Moreover, the correlation between the predicted results and aerobic soil conditions was not found.

2. LITERATURE REVIEW

2.1 OVERVIEW

This review discusses the activities of aerobic and anaerobic microorganisms in wastewater flowing through soil. Free oxygen is available in an aerobic soil environment, but is minimal under anaerobic conditions. Wastewater in anaerobic soil can cause metals such as iron (Fe) and manganese (Mn) to become electron acceptors instead of oxygen. As a result, the metals are mobilized and can ultimately contaminate the groundwater. Models are needed to predict the oxygen level in soil to determine loading conditions necessary to keep soil aerobic. Therefore, the modeling tools of TOUGHREACT, HYDRUS, and PHREEQC, as well as statistical approaches are introduced.

2.2 MICROBES IN FOOD PROCESSING WASTEWATER

Fresh water is used to wash food residues such as sugars, salts, starch, fruits, and vegetables. The wash water is regarded as food processing wastewater. Because of the rich nutrients in food processing wastewater, microbes readily grow and deplete free oxygen.

2.2.1 Aerobic Microorganisms

To survive, aerobic microorganisms require free oxygen for aerobic respiration. Free oxygen is molecular oxygen uncombined with other substances such as normal diatomic oxygen (O₂). In food processing wastewater, microorganisms use free oxygen to consume food residues, which contains sugars and fats. This process requires electron donors and acceptors. The electron transport chain transfers electrons from an electron

donor to an acceptor as well as producing protons (H^+) to generate energy. Carbon dioxide and water are released. In aerobic respiration, oxygen (O_2) is the terminal electron acceptor (Claus et al., 2006; Hoorman et al., 2011).

2.2.2 Anaerobic Microorganisms

Anaerobic microorganisms do not use free oxygen and for some, oxygen is actually toxic. The anaerobic respiration process uses nitrates (NO_3^-), sulfate (SO_4^{2-}), sulfur (S), and certain metal hydroxide complexes as electron acceptors instead of free oxygen (Farrar et al., 2003; Degelmann et al., 2009). The reduction of metals when used as electron acceptors can result in their mobilization to groundwater. Metals (such as manganese, copper, and zinc) exist in different forms depending on the oxidation and reduction potential (ORP), often represented by the level of oxygen in the soil (Lee and Kittrick, 1984). When the soil environment is aerobic, iron and manganese interact with oxygen and form insoluble compounds such as $Fe(OH)_3$ and MnO_2 (Tiedje et al., 1984). However, when the soil environment lacks oxygen, Fe^{3+} and Mn^{4+} can change to be Fe^{2+} and Mn^{2+} , all of which are soluble in water and can be transported into groundwater (Masscheleyn et al., 1991; Lipson et al., 2012; Schellenberger et al., 2011).

2.2.3 Free Oxygen in Soil

The majority of free oxygen in soil is located in pore spaces between the soil's solid particles. Soil moisture also occupies these pore spaces. Consequently, with more moisture there is less oxygen (Moncayo, 2003; RKB, 2009). Free oxygen also dissolves

in water contained in pore spaces. However, this amount is small, and the rate of oxygen transport between water and soil is slow (Erickson and Tyler, 2001).

2.3 MODELS FOR OXYGEN CONCENTRATION

In order to simulate the oxygen concentration to find aerobic soil conditions, both physical and statistical methods were examined. TOUGHREACT, HYDRUS, and PHREEQC are physical modeling tools used to simulate the transportation of liquid, gas, and heat through soils. Statistical approaches include response surface methodology, linear mixed-effect model, and time series analysis. The complexities of each vary from simple empirical analytical to more complex numerical analysis. In the following section, the fundamental knowledge and applications of all these physical and statistical methods are presented.

2.3.1 Physical Models

TOUGHREACT, HYDRUS, and PHREEQC were developed to simulate oxygen concentration in soil. TOUGHREACT simulates chemically reactive transport of non-isothermal fluids with multiple compositions in porous and fractured media. HYDRUS models water, heat, and solute movement in different mediums that are saturated in two or three dimensions. PHREEQC simulates both equilibrium reactions and non-equilibrium reactions. Further details for each are provided below.

2.3.1.1 TOUGHREACT

The program TOUGHREACT is written in Fortran 77 and improved through introducing reactive geochemistry using a simulator called TOUGH2. Several thermodynamic, physical, and chemical processes are simulated including pressure,

temperature, water saturation, ionic strength, and pH. Soil porosity and permeability can be altered by precipitation and dissolution. The program is capable of being used to model geothermal systems, weathering processes, subsurface waste disposal, acid mine drainage remediation, contaminant transport, and groundwater quality (Xu et al., 2006; Gu et al., 2010). The applications of TOUGHREACT can be one, two, and three dimensions. The code is compatible with numerous chemicals existing in liquid, gas, and solid states. A diversity of chemical reactions for equilibrium involving aqueous complexation, gas dissolution/exsolution, and cation exchange are taken into account. In unsaturated systems, due to local equilibrium or kinetic controls, dissolution or precipitation of minerals can be related to porosity, permeability or capillary pressure. Linear absorption and radioactive decay can also be modeled. The source codes, input data files, and a comprehensive user's guide are available for this program (Xu et al., 2004; Xu et al., 2006).

2.3.1.2 HYDRUS

HYDRUS is applied for saturated-unsaturated water flow, and heat and solute transport. Water taken in through plant roots is modeled by a sink term coupled with the transport equation (Simunek et al., 1999). The heat transport equation concentrates on the flow of water conducting and convecting heat. The solute transport equations are responsible for convective-dispersive transport when the media is in liquid phase and solute diffusion when the media is in gas phase. There are numerous rules included in the transport equations for non-equilibrium as well as nonlinear reactions (Simunek et al., 1999; Simunek et al., 2009). In this model, the transportation of physical non-equilibrium solutes is performed by dividing the liquid phase into mobile and immobile regions. The

attachment/detachment theory containing the filtration theory is used to model the movements of viruses, colloids, and bacteria. HYDRUS is able to control flow districts formed through boundaries with irregular shapes.

2.3.1.3 PHREEQC

PHREEQC simulates a large number of equilibrium reactions associated with gases, water, minerals, ion exchangers, surface complexes, and solid solutions. It can also simulate non-equilibrium reactions of dissolution and precipitation for minerals, reactions of microorganisms, decay of organic complexes, and other kinetic reactions. In addition, PHREEQC is able to simulate one-dimensional reactive transport processes, for example, multi-component diffusion and transport of complex species in surface media (Charlton and Parkhurst, 2011). PHREEQC is used to model leaching including metals such as lead (Pb), cadmium (Cd), arsenic (As), and chromium (Cr) derived from cement wastes. It also supplies data about leachate and precipitate speciation (Halim et al., 2005).

2.3.2 Statistical Simulations

For statistical simulation, experimental data must be collected. Two statistical methods were used to simulate the oxygen concentration in soil, the response surface methodology (RSM) and mixed-effects model. The models developed from these two methods were then compared according to AIC (Akaike information criterion) and BIC (Bayesian information criterion) derived from log-likelihood. Time series analysis was also applied but was found to be not applicable. Each method is explained in the following subsections.

2.3.2.1 Response Surface Methodology

The relationships between explanatory variables and response variables are developed based on RSM. The goal of RSM is to find the optimal responses to the explanatory variables. The values of experimental parameters are input data, which are referred as explanatory variables. RSM assumes the explanatory variables are independent so that variations are explained by measurement error, such as the instrumental error and artificial error (Zhang et al., 2007). Explanatory variables are divided into categorical variables and quantitative variables. The categorical variables are qualitative (category names or labels), while the quantitative variables are numerical values of measurable quantity. The conditions of soil depth, loading concentration, loading time, and loading frequency stated in this study with accurate values are quantitative (Lenth, 2009). The experimental results, effects or output data are response variables, of which there can be multiple. If there is only one response variable, the *rsm* package in R software is used to predict the optimal response (Wu et al., 2008; Sin et al., 2006; Zhang et al., 2007). When there are only quantitative variables, first or second-order functions are applied to develop the model. The order of the function is determined by comparing the fitted response to the surface and contour plots introduced below. The second-order polynomial function for RSM model is suggested by Box and Wilson (1951). If the RSM model is first-order, it is simple linear regression. The first-order linear regression presents a monotonic line, i.e., increasing, or decreasing line. The monotonic line extends infinitely with no maximum or minimum, where the explanatory variables are infinite. The second-order RSM creates a curve rather than a linear line, and may present local maximal or minimal values. This matches the objective to find the

maximum or minimum values as optimal responses. If the optimal response is only one critical value, the second-order function is still preferred. If the trend is linear rather than quadratic, the second-order coefficients will be zero and the surface will be most likely flat, i.e., reduced to a first-order regression.

The response surface method involves at least one experiment. Once the first experiment is analyzed, the result is used to guide subsequent experiments. These experiments may be conducted under either the same or different conditions (Lenth, 2009). The researchers repeat the same experimental conditions to avoid random errors. The objectives of experiments with different conditions may discover the range of parameters, or contribute to practical applicability.

To intuitively display the fitted RSM results, response surface plots in three dimensions and contour plots in two dimensions are created. In response surface plots, the data of a response variable is applied to the z axis; explanatory variables are placed on the x and y axes. The surface plot is in the same color with thick gridded lines. There are hills, valleys, and ridge lines to show changes in the response variable. Plots can be rotated to different angles to find a suitable view for investigation. Response surface plots visualize the trends of the fitted response variable. Contour lines may be added at the base of response surface plots to clearly show the structure of the response surface. In contour plots, contour lines with labeled numbers indicate the levels of the response variable. Each contour line represents the same level of response variable. Explanatory variables are shown by the x and y axes. The change of response variable can be presented by a color gradient as the background combined with the overlaid contour lines of contour plots. Every color corresponds to a certain value of response variable. Contour

plots allow for identification of an estimate value for optimal response variables. For both response surface plots and contour plots, if there are multiple explanatory variables, the relationship between any two can be examined by setting the other ones at their mean value (Wu et al., 2008; Lenth, 2009).

2.3.2.2 Mixed-Effects Model

Like RSM, the mixed-effects model describes the relationship between response variables and covariates to find the optimal responses, but uses a function consisting of fixed and random effects. The effects are represented by coefficients in the governing equation determined by repeated experiments. The repeated experiment is under either the same or different experimental conditions, but the experimental design should be the same and only the values of covariates are changed. The fixed effects correspond to the measured data. There can be multiple fixed effects in an experiment. Meanwhile, the random effects are the unexpected errors described by the random mutual influences (Fox, 2002). The definitions of fixed effects and mixed effects proposed by Gelman (2005) are “We define effects (or coefficients) in a multilevel model as constant if they are identical for all groups in a population and varying if they are allowed to differ from group to group”. For example, in this study, oxygen concentration measurement was repeated with different loading conditions and soil depths. The loading concentrations, loading time, loading frequency, and soil depth, in which values were definite and the effects on oxygen concentration were certain, contributed to fixed effects. However, some factors affecting the oxygen change might not be considered in the experiment, for example, the temperature and microorganism distribution. Temperature affects microorganism activity and moisture evaporation. Meanwhile, the initial microorganism distribution in the soil

column is unknown, and wastewater application will add more uncertainty. The added microorganisms likely consume different amounts of oxygen with the same loading. These unexpected factors were ignored and create the random effects. For each experiment, because the factors contributing to the random effects are unexpected, the value of its random effect is different. The distribution of random effects is assumed to be normal with mean of zero and a known variance (Cheung, 2008).

Comparing the mixed-effects model to RSM, their experimental times can be different for each experiment. However, the mixed-effects model is better at simulating the response variable with missing covariate values. These missing values may be a result of various factors, such as the experimental error (Howell, 2012). Moreover, the covariates of mixed-effects model are not independent. Therefore, the error from probability of correlation between the data sets from repeated experiments should be taken into consideration in addition to the measurement error. This error is corrected by the random effect.

2.3.2.3 Time Series Analysis

Successive measurements over time with the same time intervals form a time series. Under each condition, the oxygen concentration was collected every ten minutes, so these oxygen concentration data points were regarded as a time series. In our study, the frequency, amplitude, and slope of the oxygen concentration change were predicted by the time series analysis. The *periodogram* package in R was used to predict the frequency, and the *stl* (Seasonal decomposition of time series by loess) package of seasonal decomposition methods described by Cleveland et al. (1990) was used to predict the oxygen concentration time series. Based on the predicted time series, the amplitude and

slope of the oxygen concentration change were predicted. The amplitude was the difference of the maximum and minimum data points in each cycle of the smoothed time series. The slope was predicted based on the mean values of the maximum and minimum points in each time cycle of the smoothed time series. These mean values were fitted with a regression. The slope of this regression was regarded as the slope of the smoothed time series (Box and Jenkins, 1994; Cleveland et al., 1990). It was hypothesized that these parameters (period, amplitude, and slope) could be correlated to conditions to keep soil aerobic.

First, the best fit frequency was determined by the *periodogram* package in R software. The best fit period (T) was obtained by the reciprocal of the best fit frequency (Vaughan and Uttley, 2006). Then the *stl* package was applied to make the prediction. The original time series was broken down into sub-series. These sub-series were produced according to time division, e.g. t , $t+T$, $t+2T$, $t+3T$... The t was random, and was assumed as T . The original time series was decomposed into three components: seasonal, trend, and the remainder. This process was named as seasonal decomposition. The time series was smoothed during the decomposition by separating the random errors from the original time series. In the *stl* package, the loess regressions were built inside to smooth the data. Finally, all three components were added to produce the new smoothed time series showing the predicted oxygen concentration (Box and Jenkins, 1994; Cleveland et al., 1990).

3. METHODOLOGY AND RESULTS

Research was performed to observe how different soil depth, organic loading conditions, and the presence of a confining groundwater table affect soil oxygen content and water content. This was achieved by constructing soil columns to hold sand media with a submerged or un-submerged bottom (representing groundwater) that were monitored through the use of sensors at multiple depths. The sand was then subjected to different organic loadings and frequencies. Further details into this process are explained below.

3.1 EXPERIMENT DESIGN

Six short columns (Column 1 to 6) and two long columns (Column 7 and 8) were constructed. The lengths of the short and tall columns were 96.52 cm and 157.48 cm, respectively. Each column was made from single-wall, 45.72 cm inner diameter corrugated drainage pipes. The base of each column consisted of a split-end cap with a 1.2 cm layer of pea gravel. Eleven 0.32 cm holes in each cap allowed drainage into a bucket. A dimensional schematic of the short and long columns is presented in Figure 1. The distance between the highest position sensor and the surface of sand and between the lowest position sensor and the bottom of the sand was 10.16 cm.

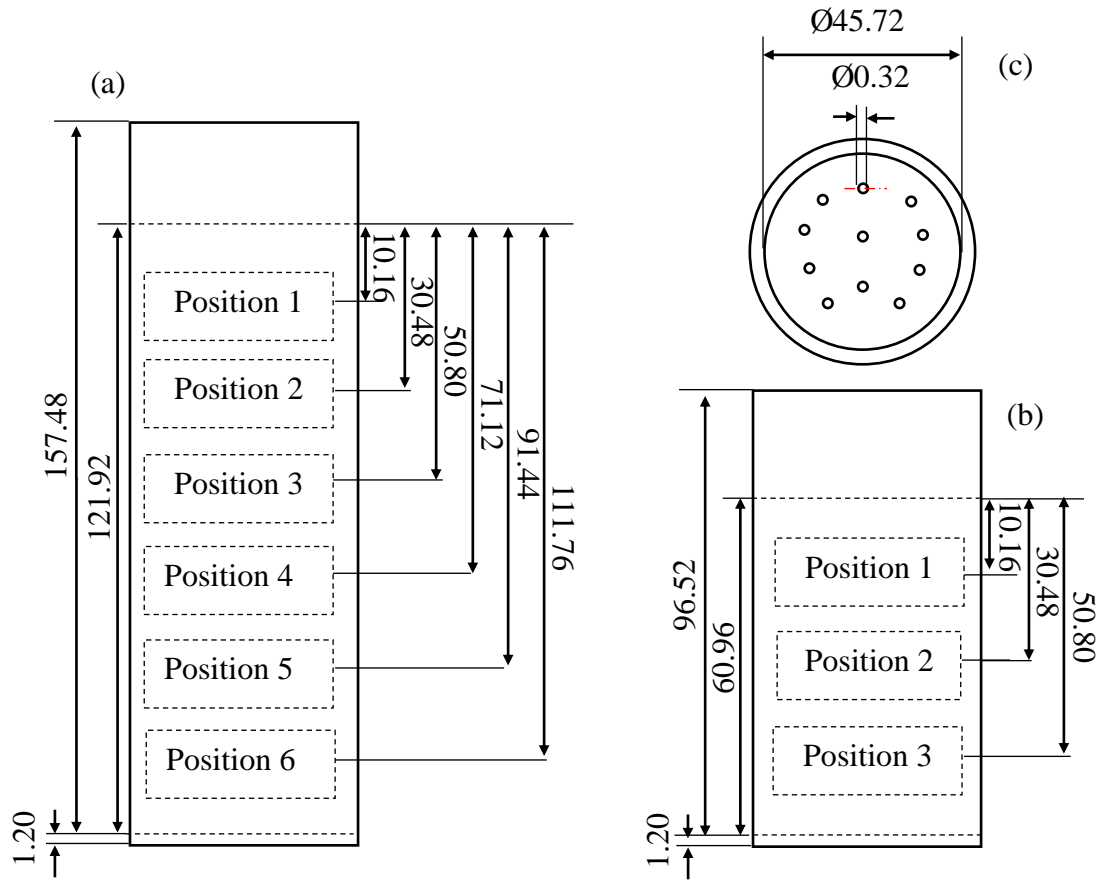


Figure 1: Dimensions of the (a) long column, (b) short column, and (c) bottom of the column (all in cm). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Organic loading was sprayed inside the columns from the top. The constituents in the synthetically prepared organic loading (consisting primarily of glucose) were based on research by Trulear and Characklis (1982) presented in Table 1. Organic loading flowed down through the soil to the bucket at the bottom of columns. When the leachate stayed in the bucket and covered the bottom of column, this situation was classified as submerged. When the leachate in the bucket did not cover the bottom of column, this condition was un-submerged. The submerged condition was used to simulate the soil with groundwater under it, while the un-submerged condition simulated the soil without the influence of groundwater or the artificial provision of oxygen.

Table 1: Content of the synthetically produced organic loading

BOD Loading Concentration (g /m ² /day)	7	28	56	112
Stock Solids (g/L)				
Na ₃ C ₆ H ₅ O ₇	0.0204	0.0786	0.1573	0.3145
Glucose	0.5000	1.9270	3.8540	7.7080
KH ₂ PO ₄	0.1767	0.6808	1.3616	2.7233
NaH ₂ PO ₄	0.2254	0.8687	1.7374	3.4747
Stock Solution(g/L)				
FeCl ₃	0.0023	0.0087	0.0173	0.0347
MnCl ₂ •4H ₂ O	0.0005	0.0021	0.0042	0.0085
ZnSO ₄ •2H ₂ O	0.0004	0.0015	0.0031	0.0062
CuCl ₂ •2H ₂ O	0.0002	0.0010	0.0019	0.0038
CoCl ₂ •6H ₂ O	0.0004	0.0013	0.0027	0.0054
(NH ₄) ₆ Mo ₇ O ₂₄	0.0002	0.0010	0.0019	0.0038
Na ₂ B ₄ O ₇ •10H ₂ O	0.0002	0.0006	0.0012	0.0023
Na ₃ C ₆ H ₅ O ₇ •2H ₂ O	0.0288	0.1107	0.2214	0.4428
NaH ₂ PO ₇ •H ₂ O	0.0184	0.0707	0.1413	0.2826
(NH ₄) ₂ SO ₄	0.1709	0.6578	1.3156	2.6312
NH ₄ Cl	0.0154	0.0593	0.1186	0.2372
CaCl ₂	0.0283	0.1088	0.2175	0.4351
MgCl ₂ •6H ₂ O	0.0023	0.0087	0.0173	0.0347

The height of the sand media in the short and tall columns was 60.96 and 121.92 cm, respectively, and their masses were 205 kg and 410 kg, respectively. Laboratory measurements of sand properties are presented in Table 2. The bulk density and hydraulic conductivity were tested with the originally dry sand without compaction. Using the assumed particle density and the bulk density, the porosity was calculated.

Table 2: Sand properties

Sand Property	Value	Method
Texture	Pure sand	Specified by manufacturer
Particle Density (g/cm ³)	2.650	Assumed as the general value of fine sand
Bulk Density (g/cm ³)	1.546	Mass/Volume
Porosity (%)	41.640	1- (Bulk Density/Particle Density)
Hydraulic Conductivity (cm/s)	0.013	Measured by constant-head method

Three different sensors were installed at each position within the columns: oxygen, water content reflectometer, and a thermistor. Soil-air oxygen saturation was measured by an O2S-D oxygen sensor (Apogee Instruments, Inc., Logan, UT). The accuracy of the oxygen sensor is <0.01% O₂ drifts with an operational temperature from 0 to 50 °C, as reported by the manufacture. Before use, the calibration of all oxygen sensors was verified by testing in an oxygen saturated environment, air, anaerobic environment, and nitrogen. Soil volumetric water content (VMC) was measured by a CS616 water content reflectometer (Campbell Scientific, Logan, UT). The soil VMC is defined as the volume of the water in the soil divided by the total volume of soil. The manufacturer reported that the accuracy, resolution, and precision are ±2.5% VMC, 0.1% VMC, and ±1.5% VMC, respectively. Soil and environmental temperature were measured by a T108 temperature probe, utilizing a BetaTherm 100K6A thermistor (Campbell Scientific, Logan, UT). If operated within the range of -5 to 95 °C, the interchangeability error is less than ±0.2 °C. Temperatures within the soil columns were maintained between 13 to 28 °C during

experimentation. All sensors were connected to a data logger, which recorded data every ten minutes.

The experiment was divided into three phases (A, B, and C) in which a different organic loadings and loading frequencies were examined. The durations of the experiments during phases A, B, and C were 254 days, 90 days, and 75 days, respectively. These durations were represented by loading time. The bottoms of columns 1, 3, 5, and 7 in phase A and B were submerged in water to represent groundwater. The remaining columns were un-submerged. The loading frequency was the time interval between loadings, which was constant for each experiment. The loading was added every 3, 4, 6, 12, 24, 56 hours between doses. During the experiments, some sensors were replaced because of malfunctions. Electrical noise also affected the accuracy of observed data. Therefore, the data of some periods is missing and only acceptable data were selected for further study.

The data was selected based on two rules. Primarily, the oxygen concentration should be under 21% because atmospheric oxygen content is about 21%, and there was no additional oxygen added in the soil other than oxygen from atmosphere and loading (Moncayo, 2003). Figure 2 presented the unacceptable data because the maximum oxygen concentrations were greater than 21%. Moreover, the oxygen concentration should not sharply change. As shown in Figure 3, two peak oxygen concentrations were around 20%, but the concentrations became almost 0% at the valley. This may be due to sand media saturation following wastewater loading and then becoming unsaturated as wastewater moves past the sensor location. The oxygen concentration changed sharply

between 0% to approximately 20% without a regular pattern, deeming this data unsatisfactory.

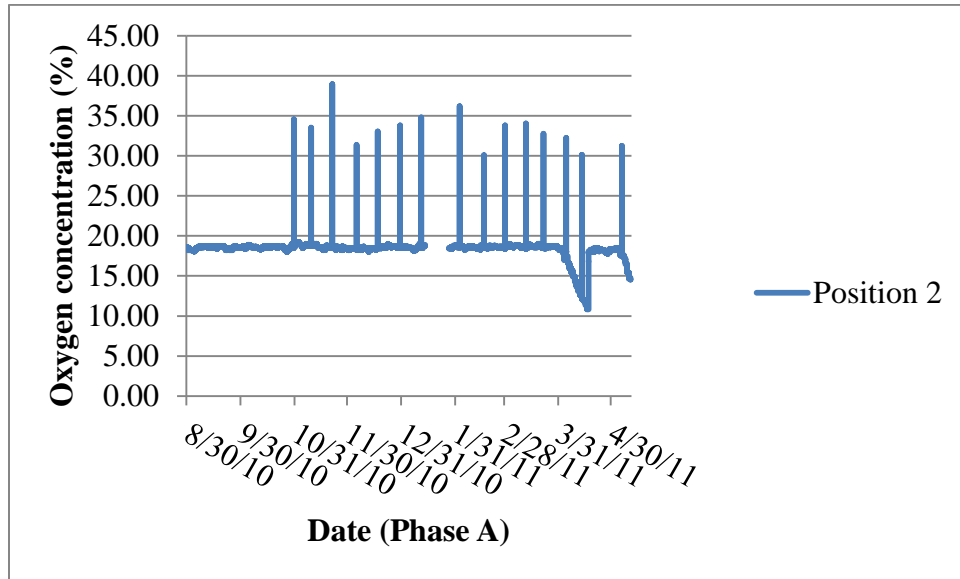


Figure 2: Unacceptable data for Column 2 in phase A with a $7 \text{ g BOD/m}^2/\text{day}$ loading concentration and 6 hour loading frequency under submerged condition

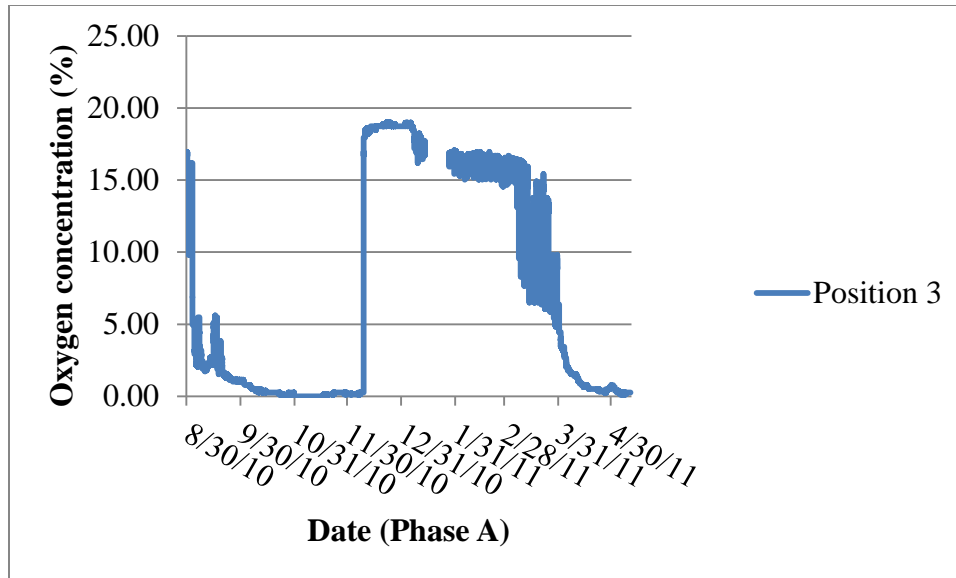


Figure 3: Unacceptable data for Column 3 in phase A with a 28 g BOD/m²/day loading concentration and 12 hour loading frequency under submerged condition

Table 3 lists the conditions where an adequate amount of acceptable data was available. The observed oxygen sensor output at each depth for the acceptable periods are presented in Figure 3.

Table 3: The loading conditions of acceptable data

Phase	A	A	A	A	B	B	B	C	C	C
Loading Concentration (g BOD/m ² /day)	7	28	28	28	112	112	112	112	112	112
Column No.	1	4	7	8	6	7	8	6	7	8
Loading Frequency (hour)	6	12	12	12	6	12	12	24	56	24
Bottom of Column Submerge	Yes	No	Yes	No	No	Yes	No	No	No	No

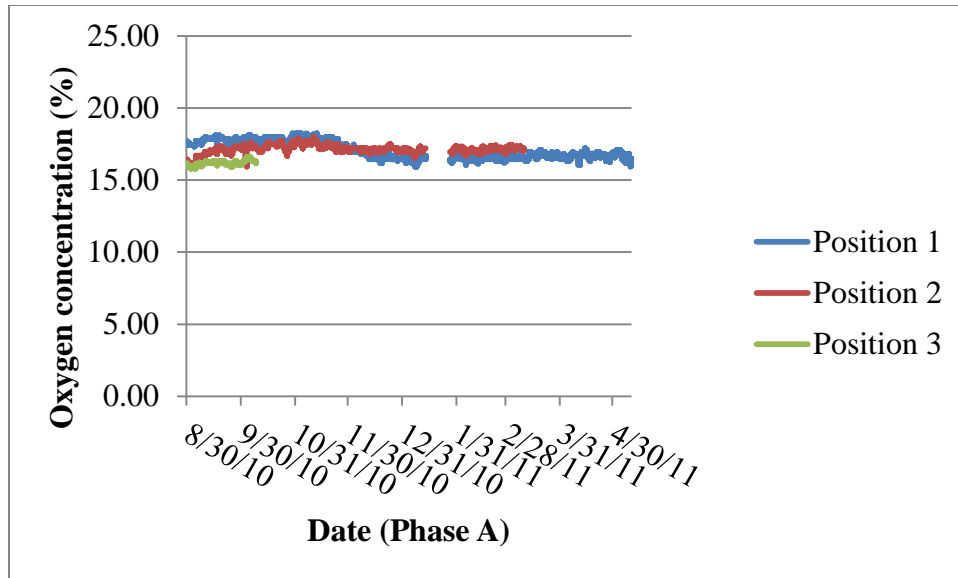


Figure 4: Acceptable data for Column 1 in phase A with a $7 \text{ g BOD/m}^2/\text{day}$ loading concentration and 6 hour frequency under submerged condition

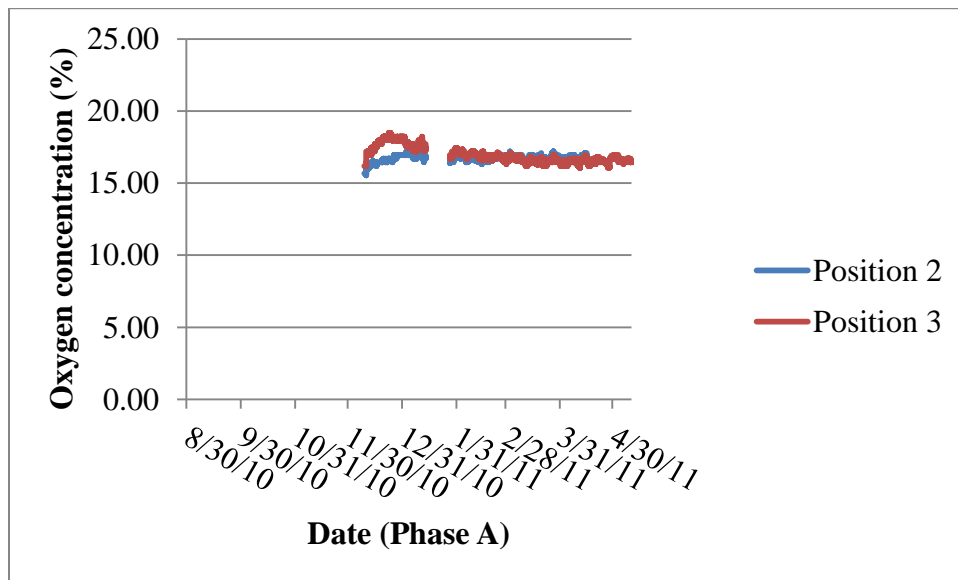


Figure 5: Acceptable data for Column 4 in phase A with a $28 \text{ g BOD/m}^2/\text{day}$ loading concentration and 12 hour frequency under un-submerged condition

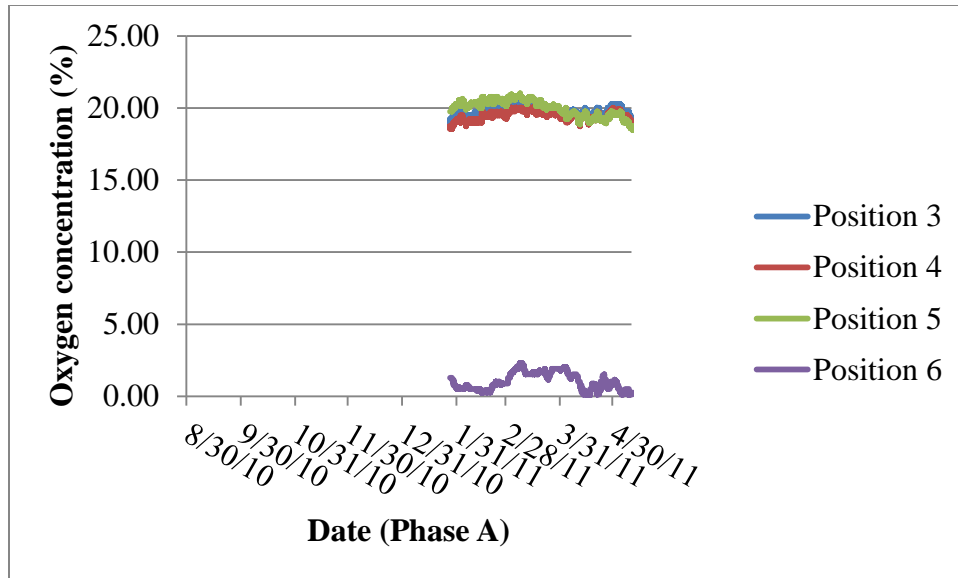


Figure 6: Acceptable data for Column 7 in phase A with a $28 \text{ g BOD/m}^2/\text{day}$ loading concentration and 12 hour frequency under submerged condition

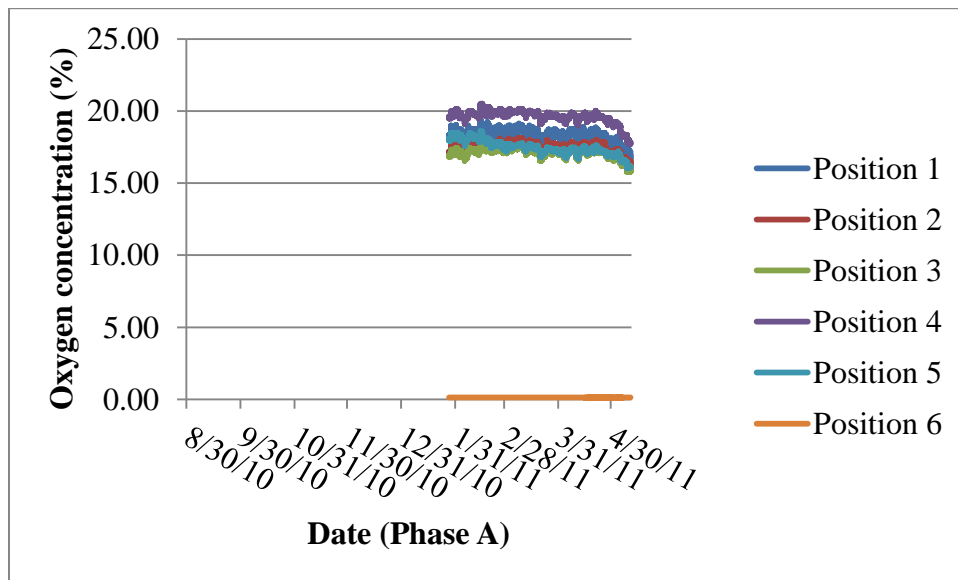


Figure 7: Acceptable data for Column 8 in phase A with a $28 \text{ g BOD/m}^2/\text{day}$ loading concentration and 12 hour frequency under un-submerged condition

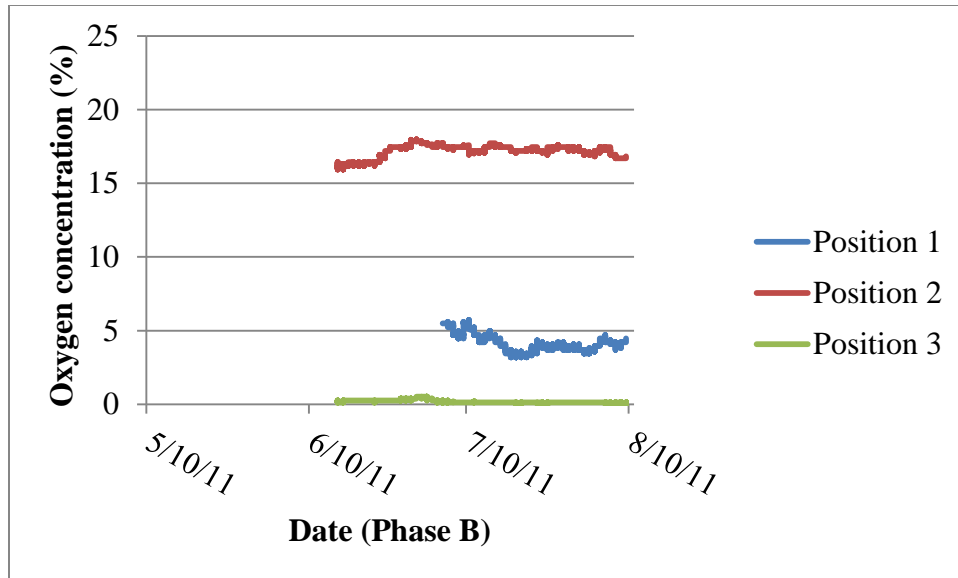


Figure 8: Acceptable data for Column 6 in phase B with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 6 hour frequency under un-submerged condition

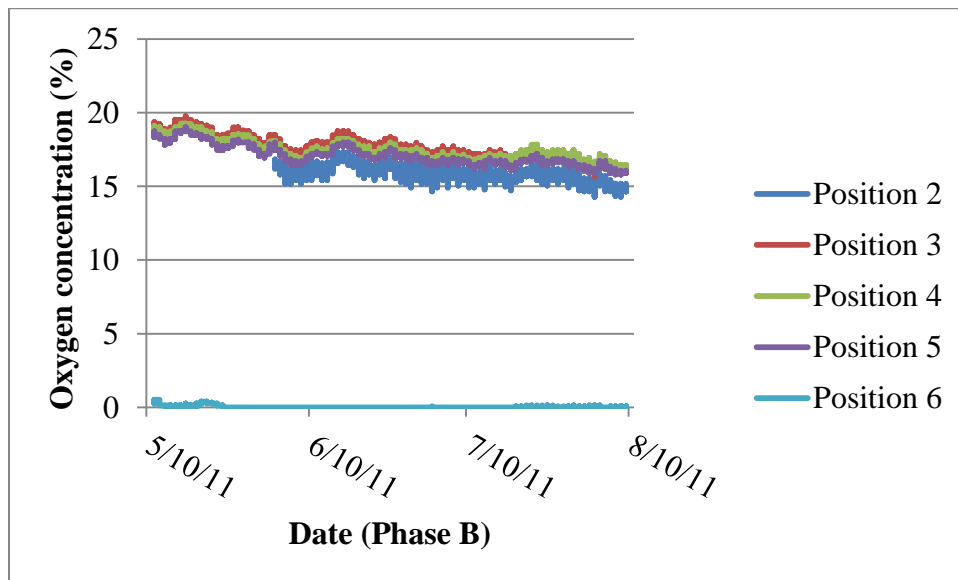


Figure 9: Acceptable data for Column 7 in phase B with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 12 hour frequency under submerged condition

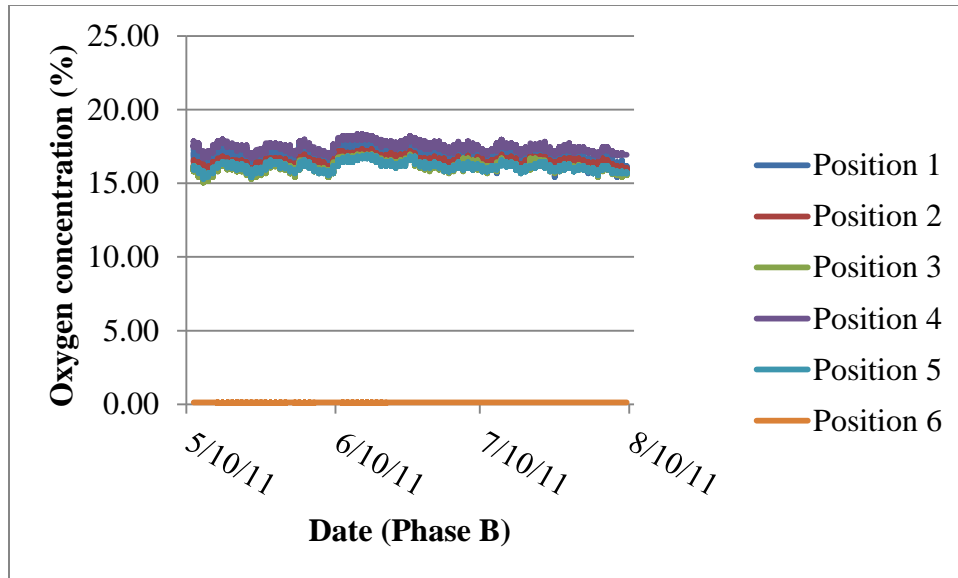


Figure 10: Acceptable data for Column 8 in phase B with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 12 hour frequency under un-submerged condition

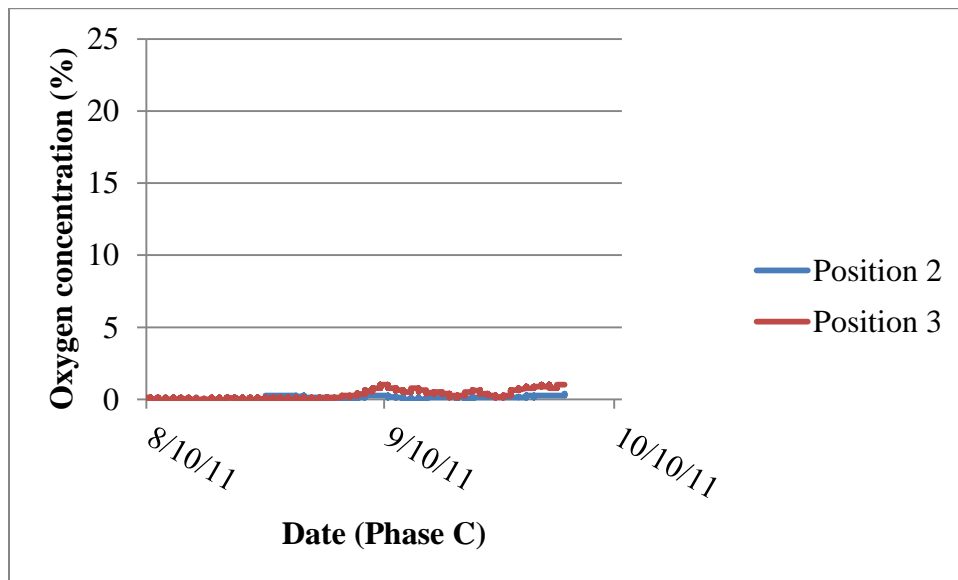


Figure 11: Acceptable data for Column 6 in phase C with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 24 hour frequency under un-submerged condition

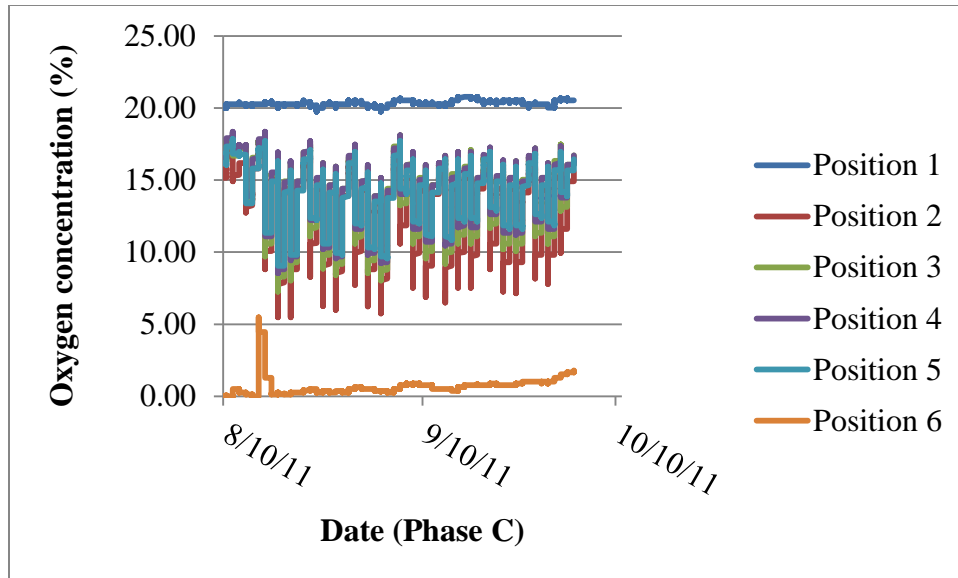


Figure 12: Acceptable data for Column 7 in phase C with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 56 hour frequency under un-submerged condition

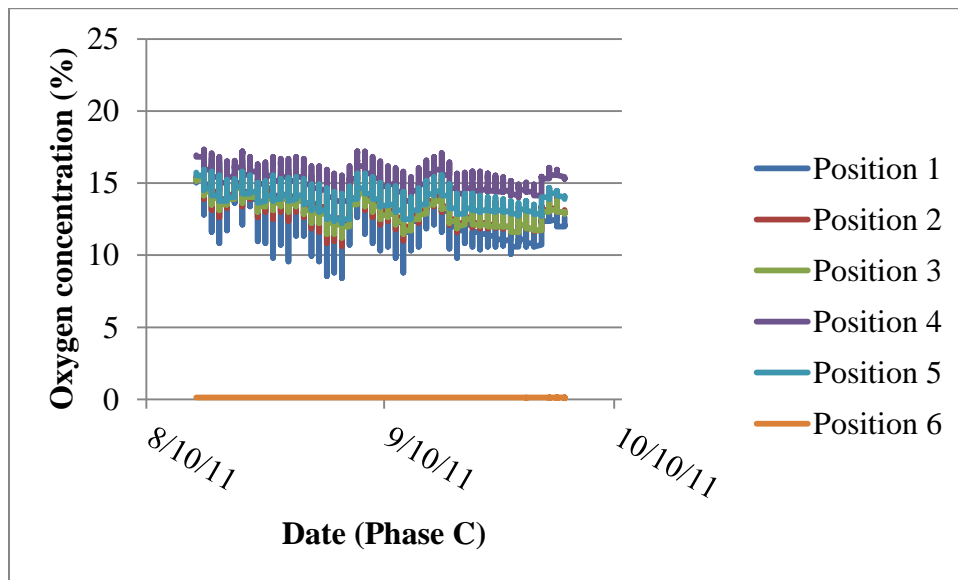


Figure 13: Acceptable data for Column 8 in phase C with a $112 \text{ g BOD/m}^2/\text{day}$ loading concentration and 24 hour frequency under un-submerged condition

In Figure 4, the curves were discontinuous. During the discontinuous period, the data was not recorded because the sensors were removed. In Figure 12 and 13, each curve demonstrated a clear oxygen concentration cycle except for the bottom curves, while curves in other figures do not exhibit this feature. From Figure 6 to 13, some of the

observed output of oxygen concentration from the bottom sensors (Position 3 in Figure 8 and 11, and Position 6 in Figure 6, 7, 9, 10, 12, and 13) was 0% regardless of being submerged or un-submerged. The columns with 0% oxygen had the following characteristics: long columns, or a high organic loading concentration of 112 g BOD/m²/day. In addition, the oxygen concentration from some sensors at a relatively low depth was higher than that from their upper sensors in the same column with the same loading under submerged or un-submerged conditions.

There were three primary factors causing oxygen reduction in the soil columns. First, free oxygen from the atmosphere went through the soil pores by diffusion, meaning oxygen molecules moved from high concentration to low concentration (RKB, 2009). Diffusion includes micro diffusion and macro diffusion. Micro diffusion is the short distance transport in the horizontal direction because of the barrier formed by the liquid loading and soil aggregates (soil clumps formed by soil practices). Macro diffusion is the major transport method referring to the long distance transport in the vertical direction (Moncayo, 2003). When the soil depth increases, the oxygen concentration decreases. Secondly, high organic loading concentration contributed to the oxygen depletion. The aerobic microorganisms require free oxygen and nutrients, especially carbon, for respiration to support their lives. Although oxygen dissolves in the organic loading, the dissolved amount was small. Additionally, the speed of oxygen transport between the organic loading and soil pore is slow (Erickson and Tyler, 2001). Most of this oxygen was consumed by aerobic microorganisms in the organic loading. The same volume of organic loading of high concentration provided more substrate than that of low concentration. With more substrate, the organisms respired more and depleted more

oxygen (Hoorman et al., 2011). Thirdly, the VMC was relatively high. The high VMC meant more of the pore space in soil was filled by water instead of air. When the soil was saturated, the oxygen concentration should be zero.

The columns with high loading concentration were divided into three regions according to the oxygen concentration in soil. The first region was the surface soil, where there was abundant oxygen from air and substrate from the loading. The aerobic microorganisms consumed the most oxygen and substrate. In the second region, as the loading decreased, little nutrients remained in the loading. Some air was still able to arrive at these depths by diffusion. Most of the microorganisms are likely aerobic in this region, but because of insufficient substrate, the microorganisms were not active. As a result, there was oxygen accumulation; oxygen measured at some deep sensors was greater than that from shallow sensors. In the third region, there was almost no substrate in the loading.

For the short column, the air would diffuse to the bottom of the column, as demonstrated in Figure 4 and 5. The exception was in Figure 8, where the oxygen was likely depleted by the high loading. For the long column, the air did not diffuse to the bottom of the column as the oxygen concentration at their bottom became almost zero shown in Figure 6 and 7. Under the un-submerged condition, there was also air diffusion from the bottom holes, resulting in higher oxygen concentration at the bottom sensor than that from the upper sensors, as presented in Figure 11. But from Figure 7, almost all the oxygen concentrations from the bottom sensor were zero, although there was oxygen added by air diffusion from the bottom. The reason might be that the region within the

bottom sensor was saturated in the long columns. The reason for the saturation was unclear.

3.2 TECHNIQUES TO SIMULATE OXYGEN CONCENTRATION IN SOIL

According to above analysis, the acceptable data was up and down in a small range with a relatively regular pattern. The acceptable data formed various curves of oxygen concentration over time. Each curve was divided into the smallest cycles which involved in only one peak or valley. Each divided part was fitted by a second-order equation. To integrate all the fitted second-order equations, a model with a final equation to predict the oxygen concentration was suggested, and the aerobic conditions (oxygen less than 10%) were found. There were two methods to develop a model: physical and statistical. Both physical (TOUGHREACT, HYDRUS, and PHREEQC) and statistical simulations (RSM, mixed-effects model, and time series analysis) were examined to simulate the oxygen concentration in soils. Each is discussed below.

3.2.1 TOUGHREACT

TOUGHREACT simulates the transport of chemically reactive, non-isothermal fluid with multiple compositions through porous and fractured media. The mass balance for organic loading in soil was found. However, the oxygen concentration in soil could not be simulated by TOUGHREACT.

3.2.1.1 Governing Equations

Equation 1 is the governing equation for the mass and energy balances. The basic integral form of Equation 1 could be utilized for water, air, and chemical components. To

attain the unconditional stability, the first-order implicit finite difference method is used to discretize the time parameter with irregular grids or regular grids (Xu et al, 2004).

$$V_n \frac{\Delta M_n}{\Delta t} = \sum_m A_{nm} F_{nm} + V_n q_n \quad (1)$$

Where,

m, n: grid block;

V_n : the nth volume element;

M_n : average mass or energy density in grid block n;

Δt : time step;

A_{nm} : surface segments;

F_{nm} : average flux (of mass or energy) over the surface segment A_{nm} between volume elements n and m;

q_n : average source/sink rate in grid block n per unit volume.

3.2.1.2 Input and Output

Three types of input files, including initial and boundary conditions, transport characteristics and geochemical properties can be introduced (Xu et al., 2004). The output files are divided into two types, fixed file names, which contain general model results,

and user-specified file names. Tables 4 and 5 represent the specific details for each file of input and output, respectively (Xu et al., 2004; Xu et al., 2006).

Table 4: Input of TOUGHREACT

Input File	File Name	Description
Flow input	flow.inp	Time step, geometric grid, initial and boundary conditions, simulation of multi-phase fluid and heat flow.
Reactive transport parameters	solute.inp	Diffusion coefficients, convergence of transport and chemical iteration tolerance limits, mineral and aqueous species flags.
Geochemical properties.	chemical.inp	Type and number of aqueous component species, minerals, gases, and sorbed species, the initial compositions of water, mineral, and gas zones.

Table 5: Output of TOUGHREACT

Output File	File Name	Description
Fixed-name	flow.out	Temperature, pressure, liquid saturation, and mass flux.
	solute.out	Transport parameters, and chemical zone configuration.
	chemical.out	Initial compositions of water, rock, and gas, equilibrium constants, and stoichiometry of chemical reactions.
	runlog.out	Run input parameters, and all run-related messages.
	chdump.out	Geochemical speciation calculations, and chemical mass balances.
	Savechem	Restart of the flow simulation.
User-specified	Iteration Data	Numbers of flow, transport, and chemical iterations of convergence at each time step.
	Aqueous species plot data.	Times, grid block coordinates (m), liquid saturation, temperature (°C), pH, and aqueous species concentrations.
	Solid phase plot data.	Time, grid point coordinates (m), temperature (°C), mineral abundance, and exchanged species concentrations.
	Gas phase plot data	Time, grid point coordinates (m), temperature (°C), and gaseous partial pressures.
	Plot data at specified grid blocks (time evolution).	Grid blocks identifier, time, liquid saturation, temperature, pH, aqueous species concentrations, mineral abundances, gas pressures, and exchanged species concentrations.

3.2.1.3 Summary

In this study, the objective was to find the oxygen concentration. In the vertical direction, block m was set up at the soil surface and block n was set up at any depth of soil. A_{nm} was the area of upper surface of the column, and that of every column was the same for every column. When we assumed V_n as the VMC at depth n, M_n was calculated by the product of A_{nm} and loading concentration. Δt was the time of loading leaching

from m to n. F_{nm} was loading flux of mass over A_{nm} between m and n. q_n was the leaching for n per unit volume. The values of q_n and F_{nm} were not known because wastewater flow velocity from the top to the bottom of the soil column was not measured. The q_n were tried to be assumed as constant in the soil column. However, the relationship between oxygen concentration and VMC at the same soil depth was not found. Consequently, the model was not developed by TOUGHREACT.

3.2.2 HYDRUS

HYDRUS models water, heat, and solute movement in different mediums that are saturated in two or three dimensions. The VMC distribution was modeled according to the data collected by the moisture sensors at different soil depths. However, an accurate relationship between the VMC and oxygen concentration was needed for further investigation. Thus, HYDRUS had no ability to simulate the oxygen concentration in soil directly.

3.2.2.1 Governing Equations

The governing equation, Richards equation, is originally derived from the mass continuity equation in combination with Darcy's law to describe water transport in unsaturated non-swelling soils (Equation 2) (Radcliffe and Simunek, 2010).

$$\frac{\partial \theta}{\partial t} = - \frac{\partial \theta}{\partial z} \left[K \left(\frac{\partial \psi}{\partial z} - 1 \right) \right] \quad (2)$$

Where,

θ : water content;

t: time;

z: elevation above a vertical datum;

K: hydraulic conductivity;

ψ : pressure head.

θ was known as predicted VMC, time was the loading time, z was the measured distance from soil surface, K was measured by the constant-head method as shown in Equation 3, and ψ was the distance from the bottom of soil to the measured level (Klute and Dirksen, 1986).

$$K = \left(\frac{V}{AT} \right) \left(\frac{L}{H_2 - H_1} \right) \quad (3)$$

Where,

V: volume of water;

A: cross sectional area of core sample;

t: time;

L: length of core;

$H_2 - H_1$: head gradient.

3.2.2.2 Input and Output

The input to the HYDRUS is divided into five files, each of which is further subdivided into multiple blocks. The blocks are divided into files as demonstrated in

Table 6. The information produced by this model is presented in output files (Table 7) (Simunek et al., 2009).

Table 6: Input of HYDRUS

File Name	Block	Description
SELECTOR. IN	A. Basic Information	Considering solute and heat, equilibrium or adsorptions in the solute transport.
	B. Water Flow Information	Absolute water content and pressure head tolerance for nodes in the saturated part of the flow region, boundary conditions, and drainage.
	C. Time Information	Initial and final time and time increment.
	F. Solute Transport Information	Temporal weighing coefficient, artificial dispersion, absolute and relative concentration tolerance, stability criteria, dummy variable, bulk density, and longitudinal dispersivity.
	L. Major Ion Chemistry Information	Maximum and minimum number of iterations allowed during any time step between the solute transport and chemical modules, length conversion factor, and molecular diffusion coefficient in free water.
	G. Root Water Uptake Information	Type of root water uptake stress response function, critical root water uptake index, pressure head, and potential transpiration rate.
	I. Atmospheric Information	Maximum and minimum allowed pressure head at the soil surface, precipitation rate, potential evaporation rate, and potential transpiration rate.
	J. Inverse Solution Information	Number of soil materials, soil hydraulic properties model, hysteresis in the soil hydraulic properties, parameter constraints, and initial estimate of parameter.

Table 7: Output of HYDRUS

File name	Information Level	Description
RUN_INF.OUT	T-level information	Time step, number of iterations necessary for solution of the water and the solute transport flow equation, cumulative number of iterations
SOLUTE.OUT	T-level information	Actual solute flux across the soil surface and the bottom of the soil profile, cumulative solute flux across the soil surface and the bottom of the soil profile, solute concentration at the soil surface and bottom of the soil profile, time, and cumulative mass transfer between the matrix and fracture domains of the dual-permeability model.
NOD_INF.OUT	P-level information	Nodal values of the pressure head, water content, and solution and sorbed concentrations.
BALANCE.OUT	P-level information	Total amount of water, heat and solute inside each specified subregion, inflow/outflow rates to/from each subregion, and absolute and relative errors in the water and solute mass balances.

3.2.2.3 Summary

HYDRUS is capable of predicting VMC. However, the equation to indicate the relationship between VMC and oxygen concentration in soil is still needed. Thus, using the HYDRUS to indirectly simulate oxygen concentration in soil was not practical.

3.2.3 PHREEQC

PHREEQC is a common geochemical model. Both equilibrium and non-equilibrium reactions are simulated. The equilibrium of oxygen in soil required the total pressure of air and partial pressure of oxygen. The pressure of air and oxygen was not measured in these experiments. Therefore, the oxygen concentration in soil was not modeled by PHREEQC.

3.2.3.2 Governing Equations

The governing equation, the general mass-action equation (Equation 4), which is the base to drive the mole-balance, charge-balance, and phase-equilibrium functions, model the activities of aqueous, exchange, and surface species (Parkhurst and Appelo, 1999).

$$K_i = a_i \prod_m^{M_{aq}} a_m^{-c_{m,i}} \quad (4)$$

Where,

m: varies over all master species, including exchange master species;

K_i : a constant depend on temperature equilibrium;

$c_{m,i}$: stoichiometric coefficient of master species m in species i;

M_{aq} : total number of aqueous master species;

a_i : activity of the unknowns for each aqueous species i ;

a_m : activity of the unknowns for each master species m .

For the gas-phase components, Equation 5 is derived to determine the oxygen level. Only the fixed-pressure gas phase is applied in the program to model the equilibrium between a gas phase consisting of different gases and an aqueous phase (Parkhurst and Appelo, 1999).

$$\frac{dn_g}{dN_{\text{gas}}} = \frac{P_g}{P_{\text{total}}} + \frac{\sum_m^{M_{\text{aq}}} n_g c_{m,g} d \ln a_m}{dN_{\text{gas}}} \quad (5)$$

Where,

n_g : number of moles of a gas component in the gas phase;

N_{gas} : total moles of gas components in the gas phase;

P_g : partial pressure of gas component g ;

P_{total} : total pressure.

3.2.3.3 Input and Output

For gas-phase components, the PHASES data block defines a gas-phase composition to a prescribed equilibrium of an aqueous phase together with pure-phase, surface

exchange, and solid-solution assemblages. These calculations are based on mass-action equations, Henry's law constant, and temperature dependence of the constant. The types of gas phase are divided into the fixed-pressure gas phase and the fixed-volume phase. The input is shown in Table 8 (Parkhurst and Appelo, 1999). The output file is the data block containing information regarding the total concentration and transfer of gas after each equilibrium calculation (Parkhurst and Appelo, 1999).

Table 8: Input of PHREEQC

Data Block	Description
fixed pressure	One gas phase as a fixed total pressure (gas bubble forms).
fixed volume	One gas phase as a fixed volume (not gas bubble forms).
pressure	Fixed pressure of the gas phase that applies during all batch reaction and transport calculations.
volume	Initial volume of the fixed-pressure gas phase.
temperature	Initial temperature of the gas phase.
phase name	Name of a gas component.
partial pressure	Initial partial pressure of this component in the gas phase and in atmospheres.

3.2.3.4 Summary

In our study, the oxygen concentration was able to be assessed by the moles of oxygen and total gas, i.e. dn_g/dN_{gas} (Equation 5). n_g was the number of moles of oxygen, N_{gas} was the total moles of air in soil., but N_{gas} was unknown. The partial pressure and the total pressure of gas were not measured during the experiment. As a result, the oxygen concentration was not successfully predicted. Future experiments could use atmospheric pressure as total pressure, and the partial pressure of oxygen calculated from its concentration.

3.2.4 Response Surface Methodology

Compared to the physical models above, RSM simply focused on the prediction of oxygen concentration, and effects of assumed factors on oxygen concentration in soil. No additional physical parameters measurements were necessary to develop the model.

3.2.4.1 Governing Equations

As introduced in the previous section, RSM was able to predict the optimal response based on explanatory variables. The first-order and second-order governing equations of models developed by RSM were shown in Equation 6 and 7, respectively (Wu et al., 2008).

$$Y = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (6)$$

$$Y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \beta_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \beta_{ij} x_i x_j \quad (7)$$

Where,

Y: response variable;

β_0 : constant coefficient;

β_j : linear coefficient;

β_{ii} : quadratic coefficients;

β_{ij} : interaction coefficient;

x_i, x_j : explanatory variables.

In this study, the response variable (Y) refers to the oxygen concentration in soil. It was hypothesized that the oxygen concentration was affected by loading time, soil depth, loading frequency, and loading concentration. The RSM included the interaction of aforementioned factors and used a sequence of designed experiments to predict the oxygen concentration and determine the desired aerobic condition in which oxygen concentration is greater than 10% (Wu et al., 2008; Sin et al., 2006). The RSM contained sub-models, which were named as L, L+I, L+Q, and L+I+Q, shown in Equations 8 through 11, respectively. L contained the linear term, $\beta_i x_i$. Q contained the quadratic term, $\beta_{ii} x_i^2$. I contained the interaction term, $\beta_{ij} x_i x_j$. L was in the first order. L+I, L+Q, and L+I+Q were in the second order. The governing equations of these sub-models applied for both submerged and un-submerged conditions. The full RSM was assumed as a second-order polynomial model, i.e. sub-model L+I+Q (Fox, 2002; Cheung, 2008).

$$\text{L: } Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (8)$$

$$\begin{aligned} \text{L+I: } Y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \\ & \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \quad (9) \end{aligned}$$

$$L+Q: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{44}x_4^2 \quad (10)$$

$$L+I+Q: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{44}x_4^2 \quad (11)$$

Where,

Y (%): predicted oxygen concentration;

x_1 (min): loading time;

x_2 (cm): soil depth;

x_3 (hour): loading frequency;

x_4 (g BOD/m²/day): loading concentration.

3.2.4.2 Input and Output

R software was utilized in order to estimate the coefficients in the model. The input data was based on raw data from the experiments. The data of submerged and un-submerged conditions was input separately by two Microsoft Excel files. For both

conditions, the input data was divided into five columns in Microsoft Excel: loading time, oxygen concentration, soil depth, loading frequency, and loading concentration. As the loading time was recorded every 10 min, the loading time input into Microsoft Excel was counted from the beginning as 1, 2, 3, 4, 5, ..., which represented the first 10 min, the second 10 min, the third 10 min, the fourth 10 min, the fifth 10 min, and so on. In this way, the input number was simplified. For other factors, the input data was their raw values. The data with the same conditions were regarded as a data set. There were five columns in a data set, which were loading time, soil depth, loading frequency, loading concentration, and oxygen concentration. The same conditions meant the same soil depth, loading frequency, loading concentration, and continuous collection without interruption. The loading time was counted from one for every data set. The continuity of the data should be emphasized, as demonstrated with data from column 1 in phase A with a 7 g BOD/m²/day loading and 6 hour frequency under submerged condition (Figure 3). Although the data had the same condition, it was split into two parts as two sets of data when data collection was not continuous due to operational issues.

Processing the input data from the Microsoft Excel files by *rsm* package in R software according to above four sub-model equations separately, all the unknown coefficients in each sub-model were predicted. The output of the full RSM, i.e. sub-model L+I+Q, is listed in Table 9 for the submerged condition, and Table 10 for the un-submerged condition.

Table 9: Output of the full RSM under submerged condition

Variables	Estimate	Std. Error	t statistic	p-value	Lower 95% limit	Upper 95% limit
Intercept	11.8582	0.059	204.40	< 2E-16	11.7445	11.9720
Time	2.83E-5	6E-6	4.70	3E-06	2E-5	4E-5
Depth	0.2702	0.001	269.14	< 2E-16	0.2683	0.2722
Frequency	0.1620	0.006	29.06	< 2E-16	0.1510	0.1729
Loading	0.0379	0.001	26.30	< 2E-16	0.0351	0.0407
Time ²	1.19E-9	2E-10	7.12	1E-12	8E-10	2E-9
Depth ²	-0.0030	6E-6	-481.08	< 2E-16	-0.0031	-0.0031
Frequency ²	0.0033	5E-5	70.58	< 2E-16	0.0032	0.0034
Loading ²	-0.0007	1E-5	-61.03	< 2E-16	-0.0008	-0.0007
Time: Depth	-1.70E-6	5E-8	-34.78	< 2E-16	-2E-6	-2E-6
Time: Frequency	-5.70E-6	2E-7	-27.28	< 2E-16	-6E-6	-5E-6
Time: Loading	3.07E-6	4E-8	71.27	< 2E-16	3E-6	3E-6
Depth: Frequency	-0.0014	2E-5	-86.97	< 2E-16	-0.0014	-0.0013
Depth: Loading	0.0006	6E-6	102.34	< 2E-16	0.0005	0.0006
Frequency: Loading	-0.0022	6E-5	-38.31	< 2E-16	-0.0024	-0.0021

Table 10: Output of the full RSM under un-submerged condition

Variables	Estimate	Std. Error	t statistic	p-value	Lower 95% limit	Upper 95% limit
Intercept	16.9804	0.037	456.98	< 2E-16	16.9076	17.0533
Time	-2.78E-5	4E-6	-7.26	4E-13	-4E-5	-2.E-5
Depth	0.0113	0.001	13.08	< 2E-16	0.0096	0.0129
Frequency	0.0265	0.003	7.60	3E-14	0.0197	0.0333
Loading	0.0481	0.001	50.99	< 2E-16	0.0463	0.04997
Time ²	-3.52E-10	1E-10	-3.43	0.0005	-6E-10	-2E-10
Depth ²	0.0002	7E-6	33.03	< 2E-16	0.0002	0.0002
Frequency ²	0.0058	4E-5	162.19	< 2E-16	0.0058	0.0059
Loading ²	-0.0010	8E-6	-131.62	< 2E-16	-0.0010	-0.0010
Time: Depth	-1.45E-7	4E-8	-3.59	0.0003	-2E-7	-7E-8
Time: Frequency	3.53E-6	2E-7	21.04	< 2E-16	3E-6	4E-6
Time: Loading	-1.74E-6	6E-8	-29.29	< 2E-16	-2E-6	-2E-6
Depth: Frequency	-0.0047	2E-5	-273.15	< 2E-16	-0.0048	-0.0047
Depth: Loading	0.0015	7E-6	205.64	< 2E-16	0.0015	0.0015
Frequency: Loading	-0.0016	4E-5	-42.02	< 2E-16	-0.0017	-0.0016

The first column presents the name of the estimated coefficients. Intercept corresponds to the constant coefficient, Time, Depth, Frequency, and Loading represents the linear coefficient, β_0 . Time², Depth², Frequency², and Loading² are the quadratic coefficient, β_{ii} . (Time: Depth), (Time: Frequency), (Depth: Loading), and (Frequency: Loading) were interaction coefficient, β_{ij} . The estimated values of these coefficients were yielded as “Estimate” in the output. “Std. Error” was the standard error for the mean of coefficients from each set of data. The calculation was based on Equation 12. The

standard error indicates the difference between the estimated mean and the true mean of the raw data. A smaller standard error indicates the greater accuracy estimation.

$$SE = \frac{s}{\sqrt{n}} \quad (12)$$

Where,

SE: standard error of the mean;

s: sample standard deviation;

n: sample size.

The standard error was still an estimated value because the true mean of the raw data was also an estimated value during the calculation. To make a more accurate assessment for the estimated coefficients, a t-test with a 95% confidence level was conducted. The yields of the t-test were t statistic (Equation 13) and the corresponding p-value. One t statistic determined one p-value.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (13)$$

Where,

t: t statistic;

\bar{x} : sample mean;

μ_0 : mean of the estimated coefficients;

s: sample standard deviation;

n: sample size.

Using a confidence level of 95%, the statistical significance level was set to $p < 0.05$. If the estimated coefficient was significant, it could be used in the model equation. If it was not, the term including this coefficient would be removed from the model equation. A 95% confidence level indicates that there was 95% confidence that the estimated coefficient was between the lower 95% limit (Equation 14) and the upper 95% limit (Equation 15). The value of the confidence interval was able to be any value. For example, when most of the p-values were less than 0.05, 95% confidence level was suitable; when most of the p-values were between 0.05 and 0.1, 90% confidence level was suitable. The confidence level was suggested to be 95% in practice (Zar, 1984). With a confidence level greater than 95%, the estimates were more significant. However, a 95% confidence level was suitable to identify the reliability of estimates while there was no higher significance level than 0.05 required. As shown in Table 9 and 10, all the p-values were far less than 0.05, which meant the estimated coefficients were all significant, assuming the 95% confidence interval was appropriate.

$$\text{Lower 95\% Limit} = \bar{x} - 1.96SE \quad (14)$$

$$\text{Upper 95\% Limit} = \bar{x} + 1.96SE \quad (15)$$

Where,

\bar{x} : sample mean;

SE: standard error of the mean.

RSM also produced response surface plots and contour plots to intuitively present the changes of oxygen concentration in soil as introduced in the last section. Because the data of the submerged and un-submerged conditions were input separately, the plots of these conditions were also separate. There were three variables for both response surface and contour plots. The response surfaces plots was in three dimensions. In the response surface plot, x and y axes contained two of the four factors, which were loading time, soil depth, loading frequency, and loading concentration. The other two variables not presented kept their mean value (Zhang et al., 2007). The calculation for the sample mean was shown in Equation 16.

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^n x_i \quad (16)$$

Where,

\bar{x} : mean;

n: sample size;

i: sample number;

x_i : value of the i^{th} sample.

Under the submerged condition, the mean values of loading time, soil depth, loading frequency, and loading concentration were 70296.9 min, 55.43 cm, 15.9 hour, and 68.86 g BOD/m²/day, respectively. Under un-submerged conditions the mean values of these

parameters were 77188.2 min, 42.96 cm, 17.39 hour and 52.49 g BOD/m²/day respectively. For the response surface plot, the z axis was the RSM-estimated oxygen concentration; the x-direction gridlines represented the values on y axis and vice versa. Each line indicated the same value. There were also lines under the bottom surface, i.e. the x-y surface. These lines were contour lines. The points on the response surface plot projected on the bottom surface. One point on the response surface corresponded to one point at the bottom surface. These projective points on the bottom surface formed the contour lines. On response surface plots, the contour lines were not labeled with number yielded by the software automatically. However, by focusing on the bottom surface containing x axis and y axis with contour lines, and labeling the lines with values of oxygen concentration, the contour plots were created. x axis and y axis were created the same as response surface plots. As contour plots were in two dimensions, the oxygen concentration was presented by the contour lines. One contour line presented the same oxygen concentration. In addition, the background of contour plots was in different color. One color represented the values in the same range of oxygen concentration. This range was assumed as 2%.

3.2.4.3 Comparative Methods

The results of all the coefficients in each sub-model were not yielded following the calculation. Instead, the sub-models were compared first. The objective of sub-model comparisons was to find the best model, which meant the best fit model. To compare the sub-models of RSM and mixed-effects, AIC and BIC were used to identify which sub-model had the best fit.

The four sub-models of RSM were compared in order to find the best sub-model. Before the comparison, the log-likelihood, AIC, BIC, and R^2 were calculated for each sub-model using the R software. Log-likelihood was the logarithmic transformation of likelihood. Likelihood reflected the probability that the predicted value fits the observed data. The log-likelihood was the maximized value of the likelihood function using the maximum-likelihood estimation method to calculate as shown in Equation 17. Using log-likelihood instead of likelihood was more convenient in mathematical calculation, which turned multiplication into addition. To maximize the probability of fit, a maximum log-likelihood value suggested the best fit model (Myung, 2003).

$$\ln \mathcal{L}(\theta | x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i | \theta) \quad (17)$$

Where,

θ : a vector of parameters for this family;

x_1, \dots, x_n : observations in the sample;

n : size of the sample;

x_i : the i^{th} observation in the sample.

AIC was calculated for the sub-model comparisons. AIC was the short of Akaike information criterion solved by Equation 18, and calculated according to the maximized value of the likelihood. Comparing with log-likelihood, AIC finds the best fit model. In Equation 18, the maximized value of the likelihood function was negative. A penalty, k , was introduced. When there were more parameters in the model (resulting in a larger k),

the likelihood of the model would be larger, but the model may be overfit, indicating the inclusion of too many parameters relative to the number of observations. Thus, in compare the models' AIC values, the smallest AIC indicates a best fit model (Akaike, 1974).

$$AIC=2k-2 \ln(L) \quad (18)$$

Where,

k: number of parameters in the statistical model;

L: the maximized value of the likelihood function for the estimated model.

In the model comparisons, BIC was also used to verify the comparative results. Like AIC, BIC was calculated using the maximized value of the likelihood function. BIC introduces an additional penalty (n) as shown in Equation 19. Because of the additional penalty, the penalty term in BIC is more powerful than that of AIC. AIC only considers how well the model fitted the observations, but BIC also considered the observation sample error. As the sample size was large, the model tended to increase the number of parameters to increase the goodness of fit, but may have resulted in overfit. A relatively small sample size, or small penalty n, could avoid these problems. BIC is an increasing function of n and a decreasing function of L. As a result, a minimal n and maximal L indicate the best fit model. The smallest BIC is desired in the model comparisons (Schwarz, 1978; Liddle, 2008).

$$BIC = k \ln(n) - 2 \ln L \quad (19)$$

Where,

n: the sample size;

k: number of parameters in the statistical model;

L: the maximized value of the likelihood function for the estimated model.

In all, AIC and BIC were based on the log-likelihood. Both of them include a term to correct this possible error, but they were different. AIC was preferred for the infinite-dimensional samples, which meant AIC was able to find the model to best fit the oxygen collected by all the conditions. While BIC was preferred for finite-dimensional samples, which meant BIC was able to find the model to best fit the oxygen collected by our existing experiments. Thus, both AIC and BIC were applied for the comparisons between sub-models (Yang, 2005).

During the calculation of the RSM sub-models, R^2 was a default output by the software. R^2 presented how observations fit the curve produced by model. The definition of R^2 was shown in Equation 20, which identifies the percentage of explained variance. R^2 ranges from 0 to 1, where 1 indicates a perfect fit (Steel and Torrie, 1960). The coefficient of determination was not used because it is not suitable for the mixed-effects model due to the random effects. One random effect corresponds to one experiment under specific experimental conditions, but the random effect is arbitrary. Depending on the definition, the variance of all the random effects was meaningless. So the R^2 for each model was not produced.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (20)$$

Where,

SS_{res} : sum of squares of residuals equal;

SS_{tot} : total sum of squares.

3.2.4.4 Result

The RSM sub-models were compared to determine the best sub-model before yielding the result (Table 11). The L+I+Q model had the highest log-likelihood and R^2 , as well as the lowest AIC and BIC. Thus, the sub-model L+I+Q was the best fit RSM sub-model. The model of the full RSM was developed as Equation 21.

Table 11: RSM sub-model comparisons under submerged condition

Model	Log-likelihood	AIC	BIC	R^2
L	-1699061	3398134	3398201	0.2803
L+I	-1686651	3373325	3373460	0.3126
L+Q	-1598069	3196158	3196270	0.5047
L+I+Q	-1587190	3174413	3174592	0.5242

$$Y = 11.8582 + (2.83 \times 10^{-6}x_1 + 0.2702x_2 + 0.1620x_3 + 0.0379x_4) + (1.19 \times 10^{-11}x_1^2 - 0.0030x_2^2 + 0.0033x_3^2 -$$

$$0.0007x_4^2) + (-1.70 \times 10^{-7}x_1x_2 - 5.70 \times 10^{-7}x_1x_3 + 3.07 \times 10^{-7}x_1x_4 - 0.0014x_2x_3 + 0.0006x_2x_4 - 0.0022x_3x_4) \quad (21)$$

Where,

Y (%): the predicted concentration of oxygen;

x_1 (min): loading time;

x_2 (cm): soil depth;

x_3 (hour): loading frequency;

x_4 (g BOD/m²/day): loading concentration.

Based on the full RSM, the response surface plots and contour plots under the submerged condition were determined for six scenarios: soil depth and loading time, loading frequency and loading time, loading concentration and loading time, loading frequency and soil depth, loading concentration and soil depth, and loading concentration and loading frequency. This was done to evaluate the importance of two factors, excluding the influence of the remaining factors. Figure 14 was response surface plot presenting the influence of soil depth and loading time under submerged condition, and its corresponding contour plot was shown in Figure 15. The loading frequency and concentration were kept at their mean values, 15.9 hour, and 68.86 g BOD/m²/day in Figure 14 and 15. In Figure 14, the oxygen concentration increased first and then decreased with increased soil depth at the constant loading time. The oxygen

concentration was greater than 10% when the soil depth was less than 100 cm, as demonstrated in Figure 14. At the constant depth of soil, the oxygen concentration increased with loading time but the rate of increase was slower in deeper soil. This is contrary to what was expected, where increasing loading time increases total BOD loading and decrease oxygen concentration. Higher oxygen concentration at increased loading time may be due to microorganism distribution and temperature and moisture gradient.

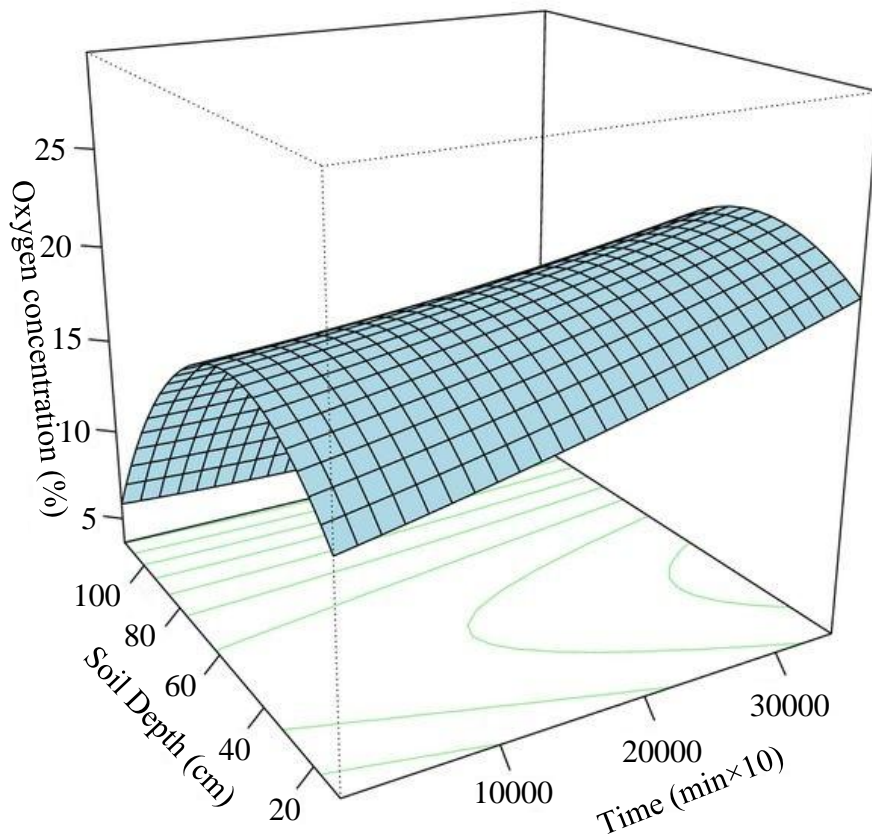


Figure 14: Oxygen concentration response surface plot under the submerge condition showing the influence of soil depth and loading time (constant loading frequency and concentration)

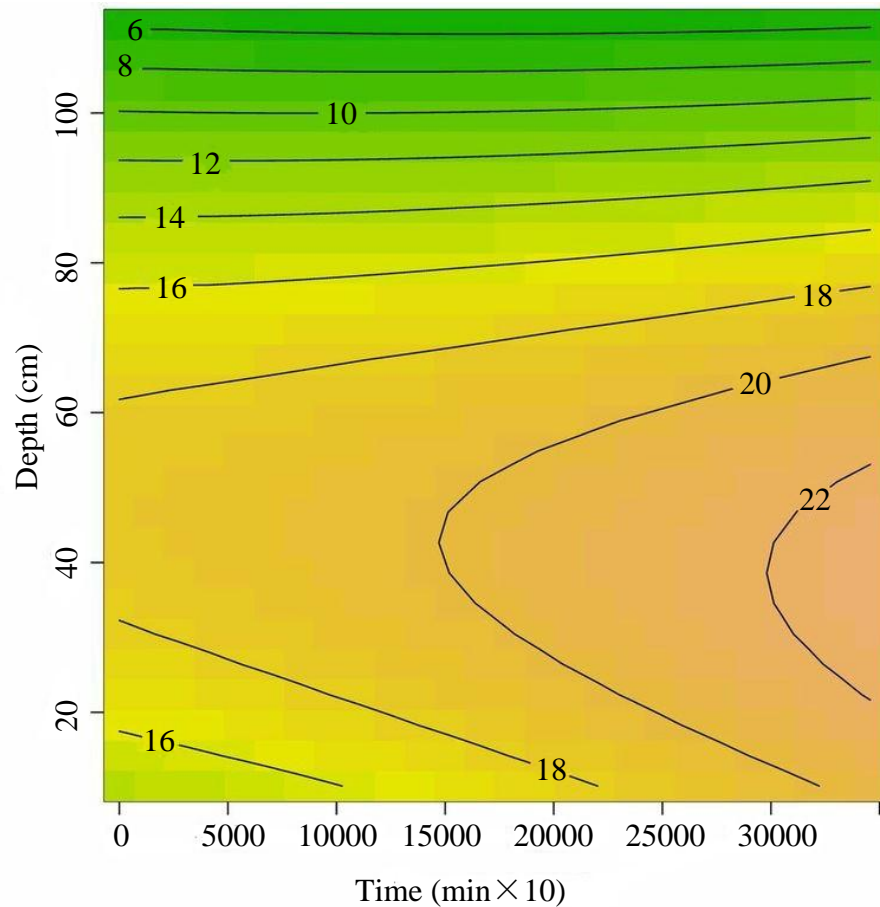


Figure 15: Oxygen concentration contour plot under the submerge condition showing the influence of soil depth and loading time (constant loading frequency and loading concentration)

In Figure 16 and 17, the soil depth and loading concentration were 55.43 cm and 68.86 g BOD/m²/day at their mean levels. Combining Figure 16 and 17, the oxygen concentration increased with increased loading frequency at the constant loading time when the loading time was less than 100000 min. The higher loading frequency means longer time interval between the loadings. By increasing loading frequency, the total BOD loading within the same time period decreases, resulting in increasing oxygen concentration. After this time, the oxygen concentration decreased first and then

increased with the increased loading frequency. The oxygen concentration increased when the loading frequency was less than 30 hour, but decreased when the loading frequency greater than 30 hour. The oxygen concentrations were all greater than 10% at the soil depth of 55.43 cm, which was less than 100 cm corresponding to the result from Figure 14 and 15.

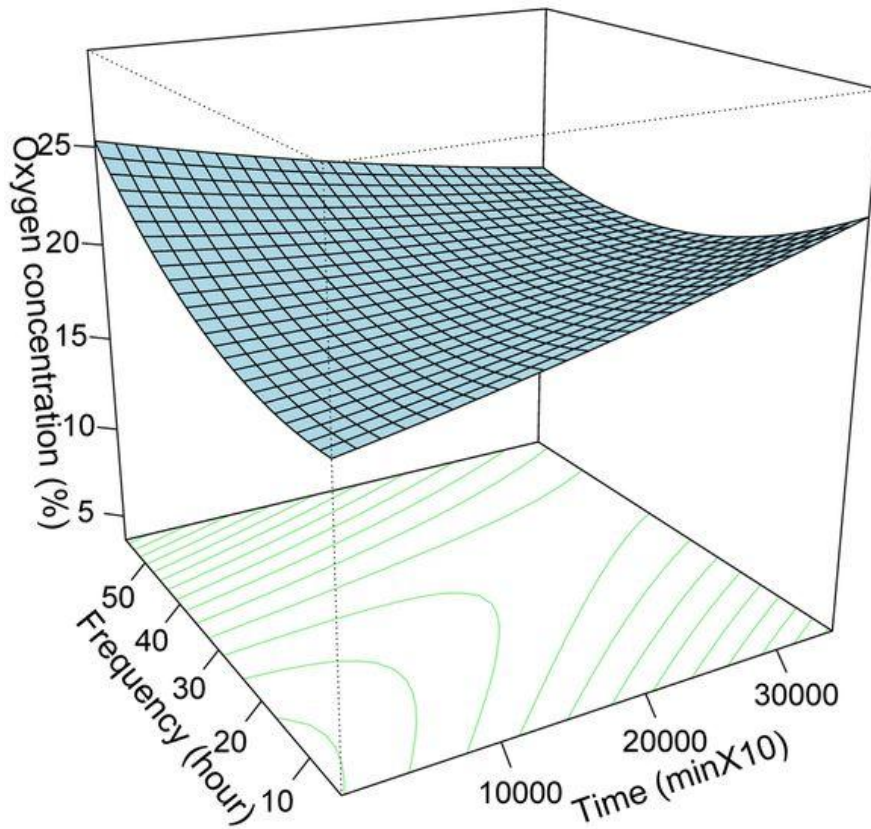


Figure 16: Oxygen concentration response surface plot under the submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)

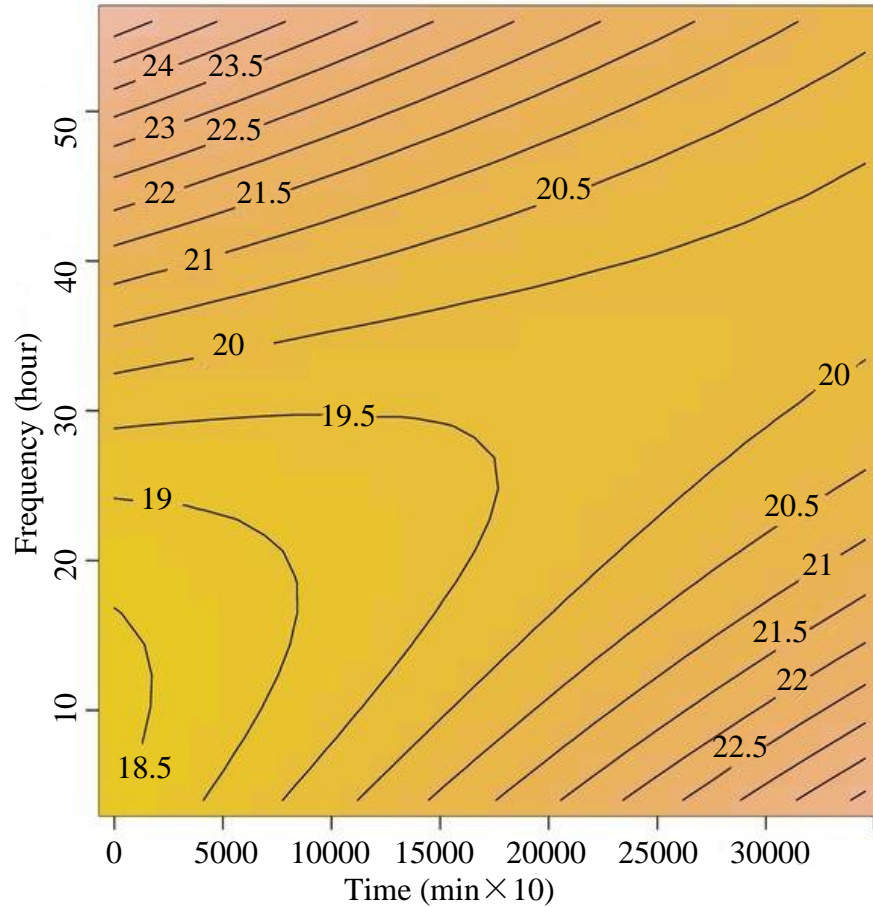


Figure 17: Oxygen concentration contour plot under the submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)

In Figure 18 and 19, the soil depth and loading frequency were 55.43 cm and 15.9 hour at their mean levels. Before 150000 min loading time, the oxygen concentration decreased with increased loading concentration. Between 150000 min to 250000 min, the oxygen concentration increased first and then decreased. After 150000 min, the oxygen concentration increased with increasing loading concentration. When the loading concentration was less than $40 \text{ g BOD/m}^2/\text{day}$, the oxygen concentration decreased over loading time. When the loading concentration was over $40 \text{ g BOD/m}^2/\text{day}$, the oxygen

increased over loading time. Based on the Figure 19, the oxygen concentration may be less than 10% over a $100 \text{ g BOD/m}^2/\text{day}$ loading concentration.

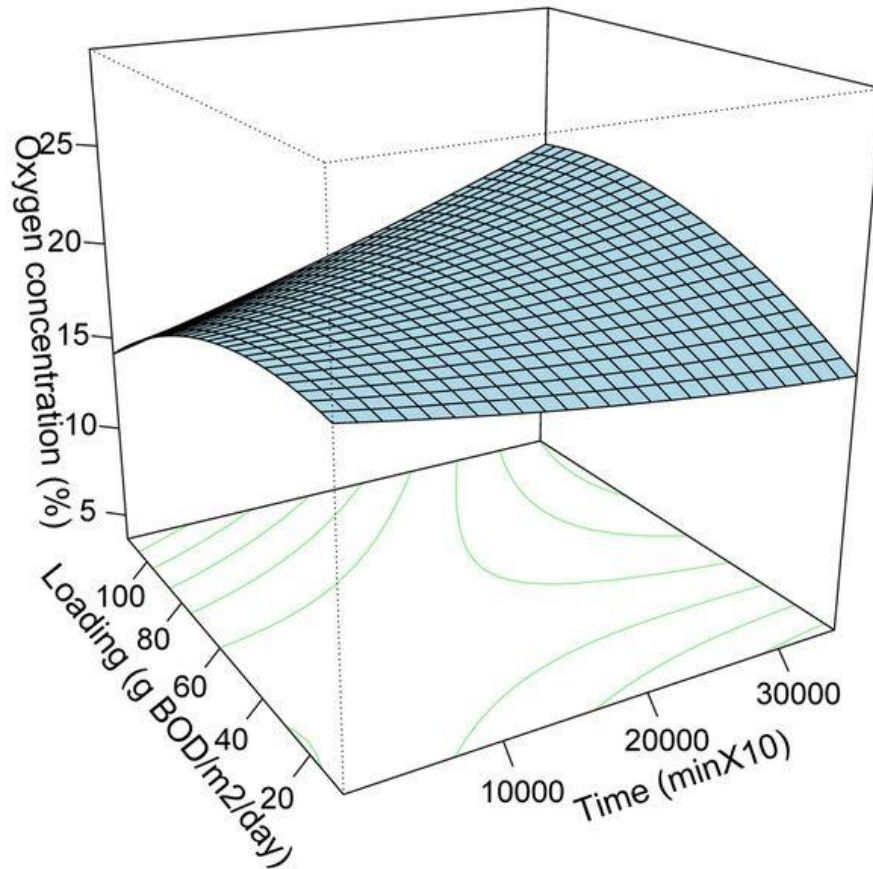


Figure 18: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)

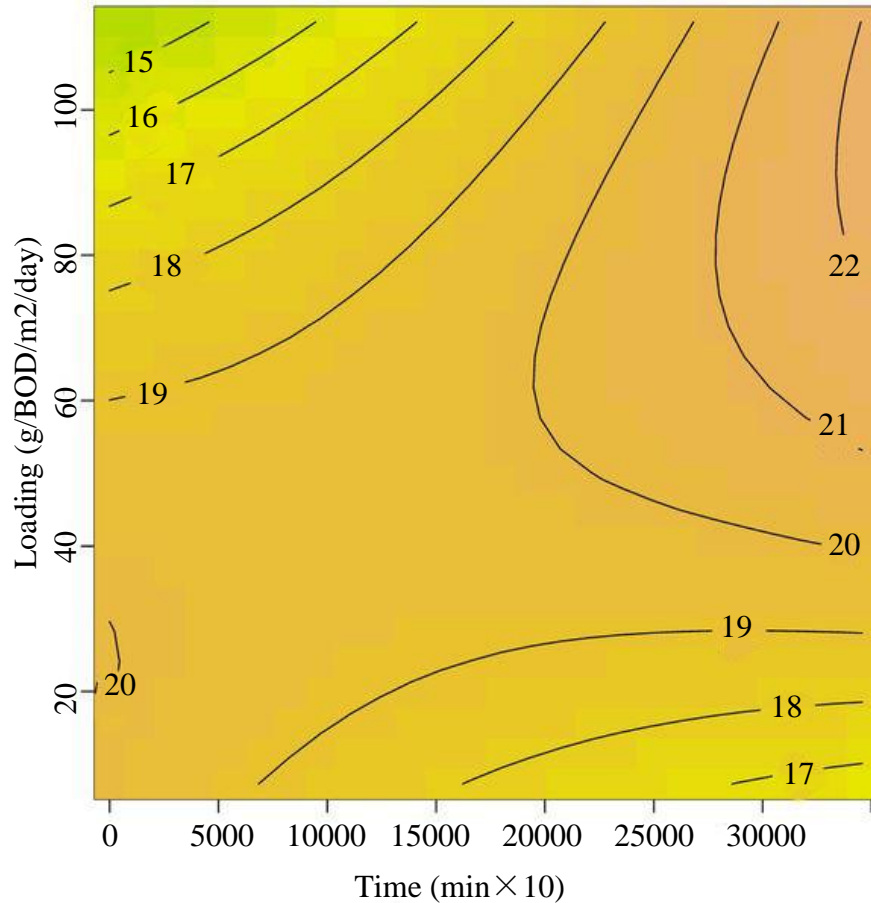


Figure 19: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)

In Figure 20 and 21, the loading time and loading concentration were 70296.9 min and 68.86 g BOD/m²/day at their mean levels. At the constant depth of soil, the oxygen concentration increased with increased loading frequency. At the constant loading frequency, the oxygen concentration increased first and then decreased with deeper soil. When the soil depth was less than 100 cm, the oxygen concentration was over 10% at all loading frequencies. However, this situation occurred when the loading time was up to 70296.9 min. How the oxygen concentration changed before and after 70296.9 min was not clear.

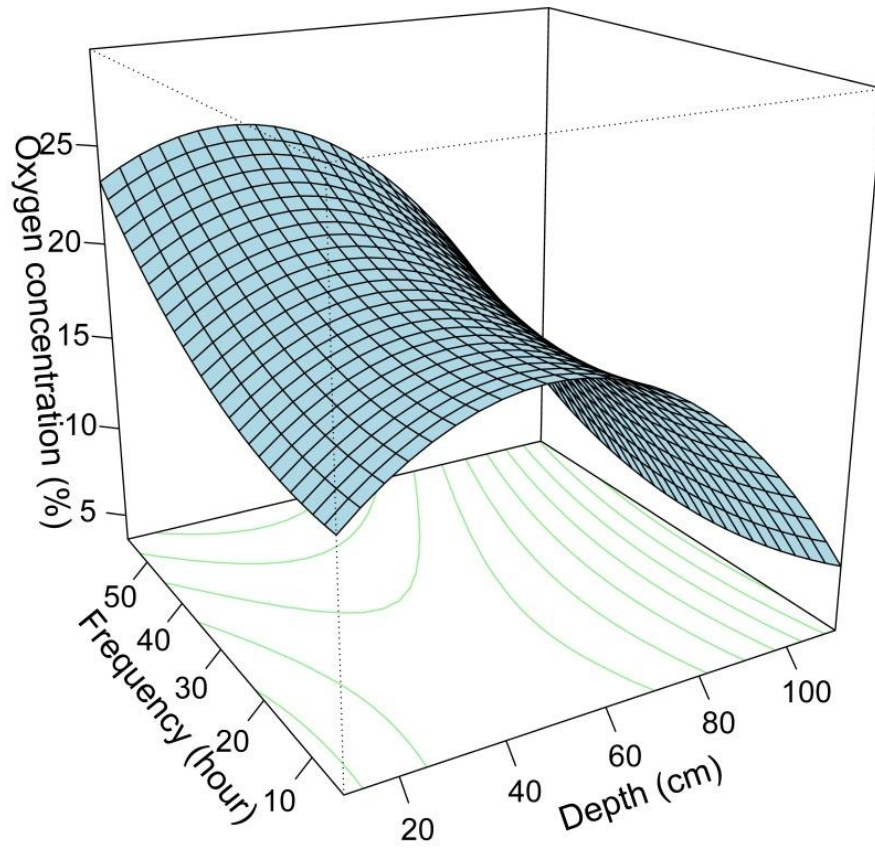


Figure 20: Oxygen concentration response surface plot under the submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)

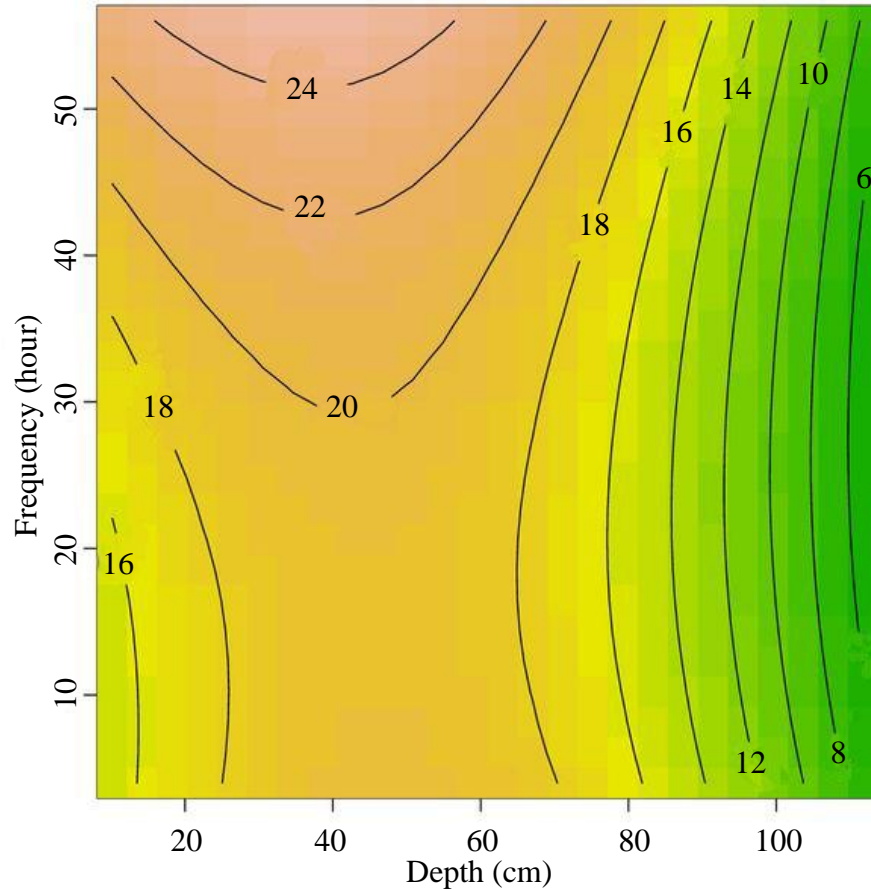


Figure 21: Oxygen concentration contour plot under the submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)

In Figure 22 and 23, the loading time and loading frequency were 70296.9 min and 15.9 hour at their mean levels. At the constant soil depth, the oxygen concentration decreased with the increased loading concentration. When the loading concentration was up to a certain level over $100 \text{ g BOD/m}^2/\text{day}$, the oxygen concentration was less than 10%. The oxygen concentration increased with increasing soil depth up to 40 cm and then decreased. With a soil depth less than 100 cm, the oxygen concentration was greater than 10%.

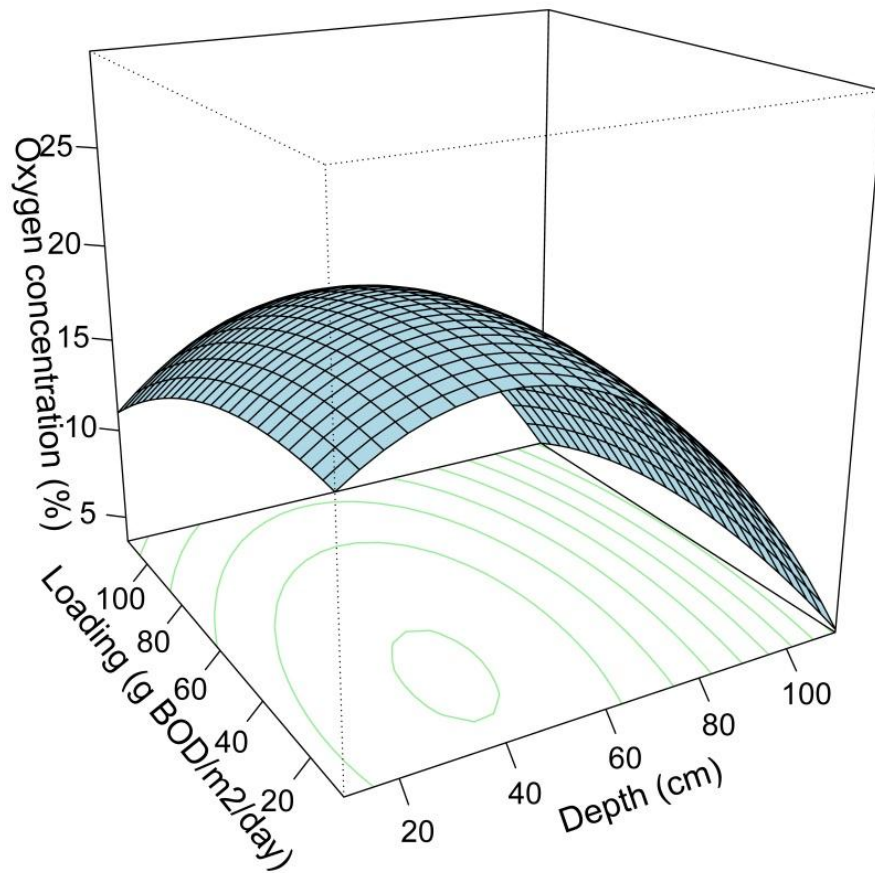


Figure 22: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)

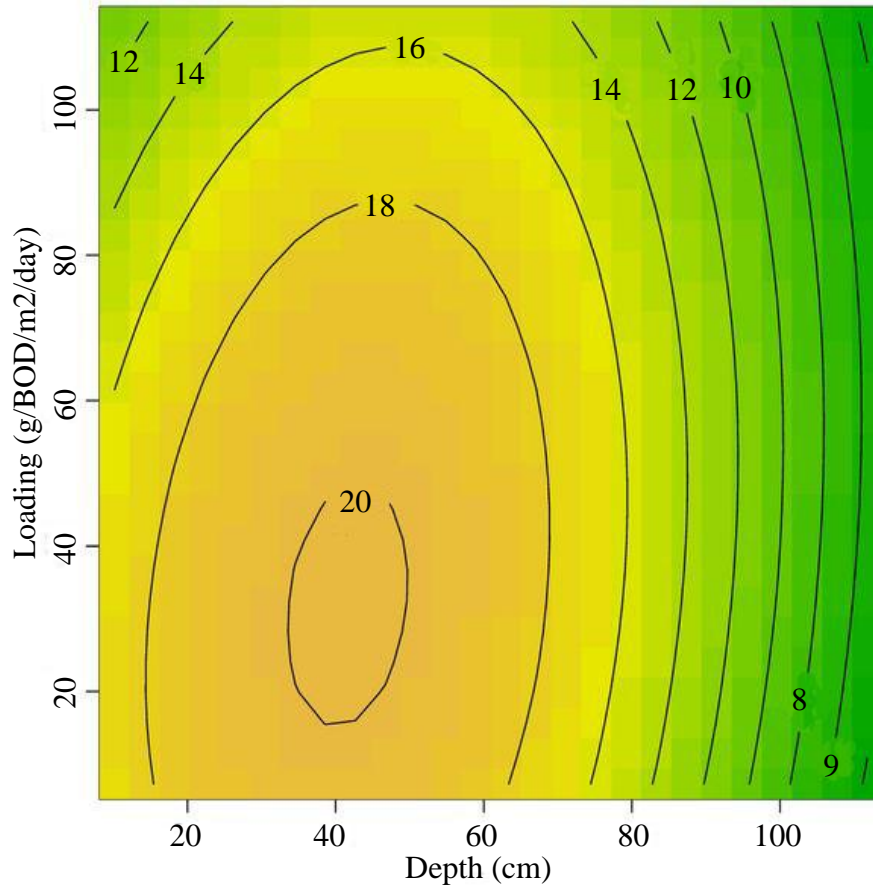


Figure 23: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)

In Figure 24 and 25, the soil depth and loading time were and 55.43 cm and 70296.9 min at their mean levels. With the constant loading frequency, the oxygen concentration decreased with increased loading concentration. When the loading concentration was constant, the oxygen concentration increased with increased loading frequency.

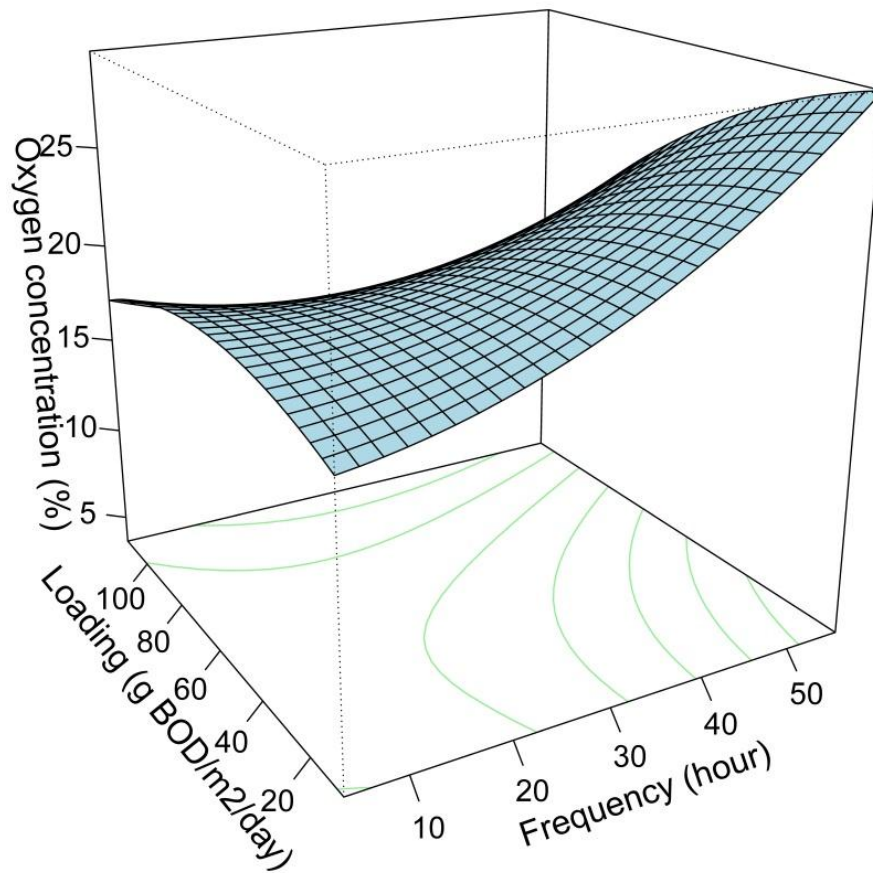


Figure 24: Oxygen concentration response surface plot under the submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)

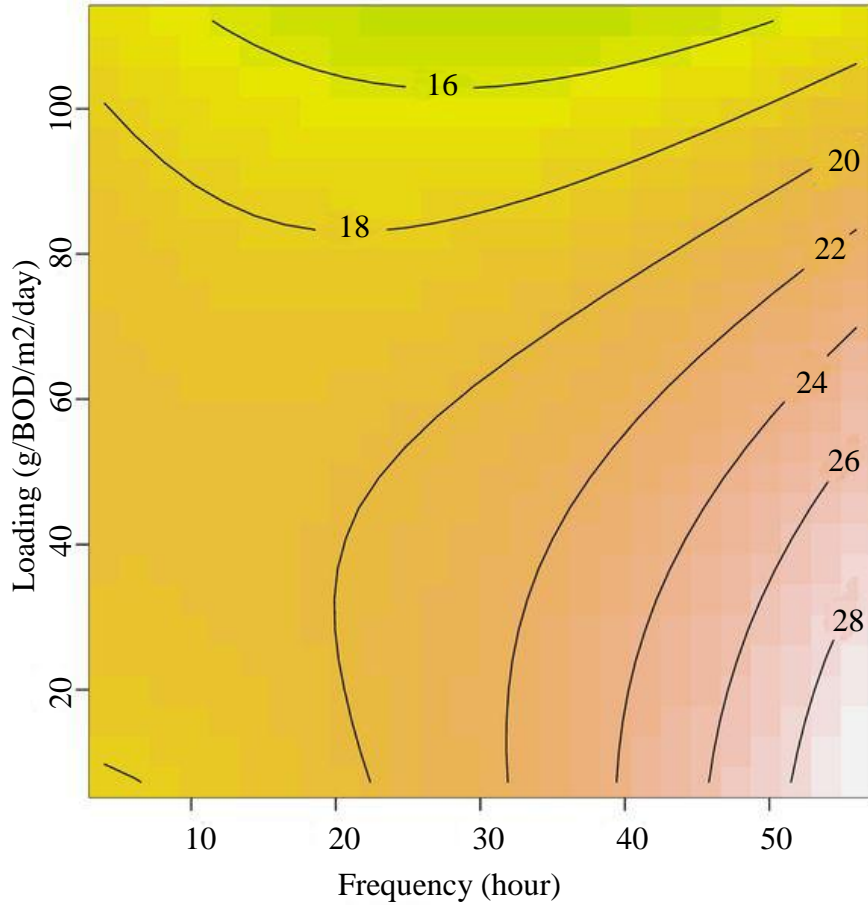


Figure 25: Oxygen concentration contour plot under the submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)

In all, under submerged conditions, the conditions to keep soil aerobic (oxygen concentration greater than 10%), was that the soil depth was less than 100 cm, and the loading concentration was less than $110 \text{ g BOD/m}^2/\text{day}$, although lower loading concentration was preferred. The loading time and frequency had little effect on the oxygen change in soil.

RSM sub-models under un-submerged conditions are presented in Table 12. The log-likelihood and R^2 of L+I+Q were the greatest, while the AIC and BIC were the

lowest. This indicates that the sub-model L+I+Q or the full RSM was the best RSM sub-model. The full RSM model under un-submerged conditions was shown in Equation 22.

Table 12: RSM sub-model comparisons under un-submerged condition

Model	Log-likelihood	AIC	BIC	R ²
L	-766499	1533010	1533074	0.4079
L+I	-750765	1501554	1501682	0.4479
L+Q	-755524	1511068	1511174	0.4643
L+I+Q	-721624	1443280	1443451	0.5551

$$\begin{aligned}
 Y = & 16.9804 + (-2.78 \times 10^{-6}x_1 + 0.0113x_2 + 0.0265x_3 + \\
 & 0.0481x_4) + (-3.52 \times 10^{-12}x_1^2 + 0.0002x_2^2 + 0.0058x_3^2 - \\
 & 0.0010x_4^2) + (-1.45 \times 10^{-8}x_1x_2 + 3.53 \times 10^{-7}x_1x_3 - 1.74 \times \\
 & 10^{-7}x_1x_4 - 0.0047x_2x_3 + 0.0015x_2x_4 - 0.0016x_3x_4) \quad (22)
 \end{aligned}$$

Where,

Y (%): the predicted concentration of oxygen;

x₁ (min): loading time;

x₂ (cm): soil depth;

x₃ (hour): loading frequency;

x₄ (g BOD/m²/day): loading concentration.

The RSM plots under un-submerged conditions were investigated. The response surface plots their corresponding contour plots under the un-submerged condition were shown below. In Figure 26 and 27, the loading frequency and loading concentration were 17.39 hour and 52.49 g BOD/m²/day at their mean levels. At a constant loading time, the oxygen concentration increased with deeper soil depth. At a constant depth, the oxygen concentration decreased over loading time.

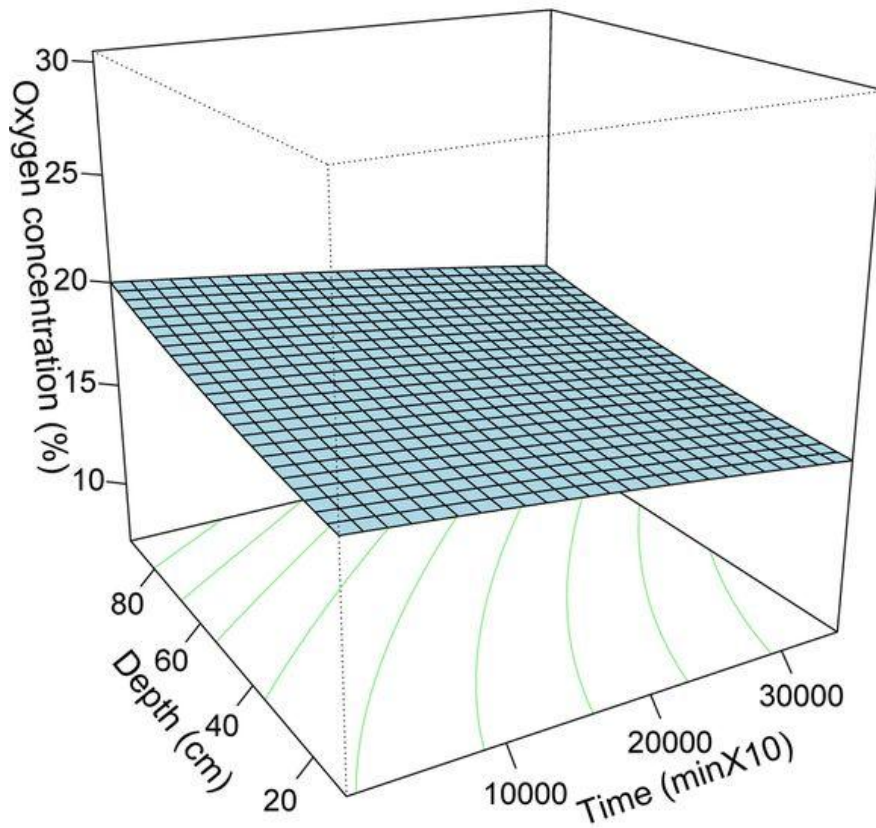


Figure 26: Oxygen concentration response surface plot under the un-submerged condition showing the influence of the soil depth and loading time (constant loading frequency and loading concentration)

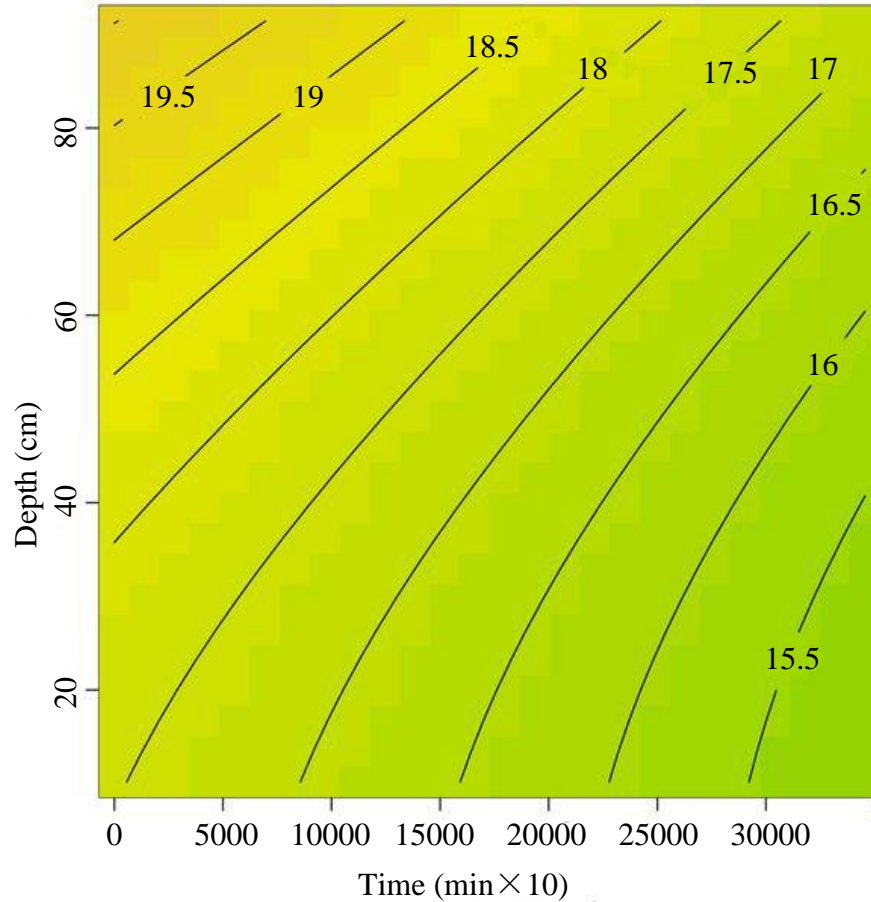


Figure 27: Oxygen concentration contour plot under the un-submerged condition showing the influence of the soil depth and loading time (constant loading frequency and loading concentration)

In Figure 28 and 29, the soil depth and loading concentration were 42.96 cm and 52.49 g BOD/m²/day at their mean levels. At the constant loading time, the oxygen concentration decreased and then increased with the increased loading frequency. When the loading time was long enough, the oxygen concentration was less than 10%. With a loading frequency less than 40 hour, the oxygen concentration decreased over time. With a loading frequency greater than 40 hour, the oxygen concentration increased over time.

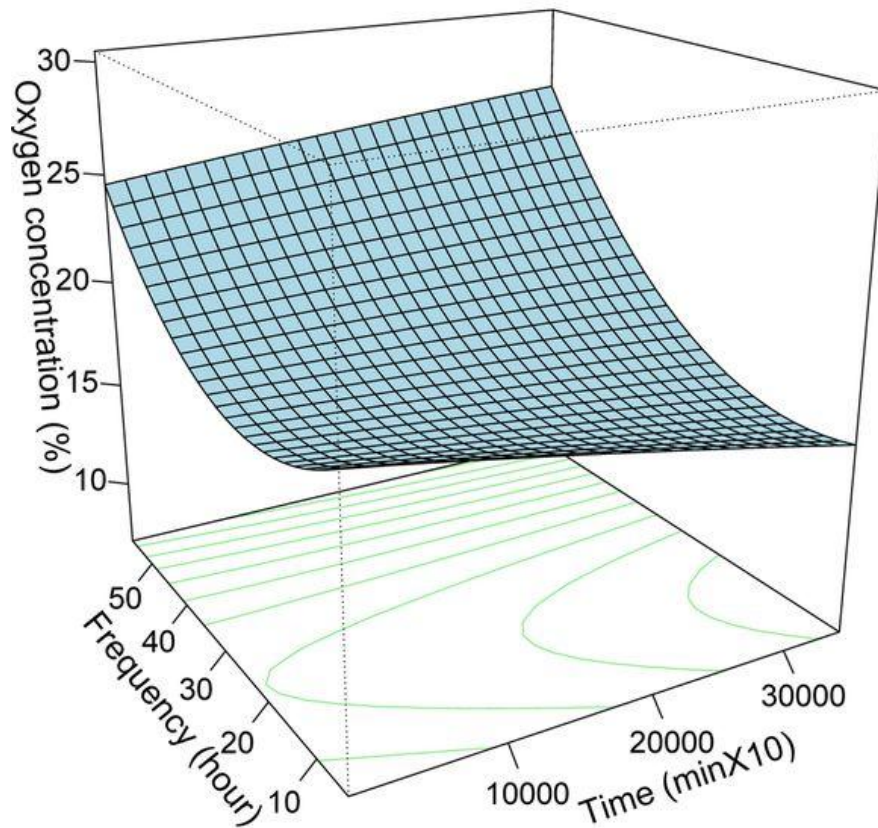


Figure 28: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading frequency and loading time (soil depth and loading concentration)

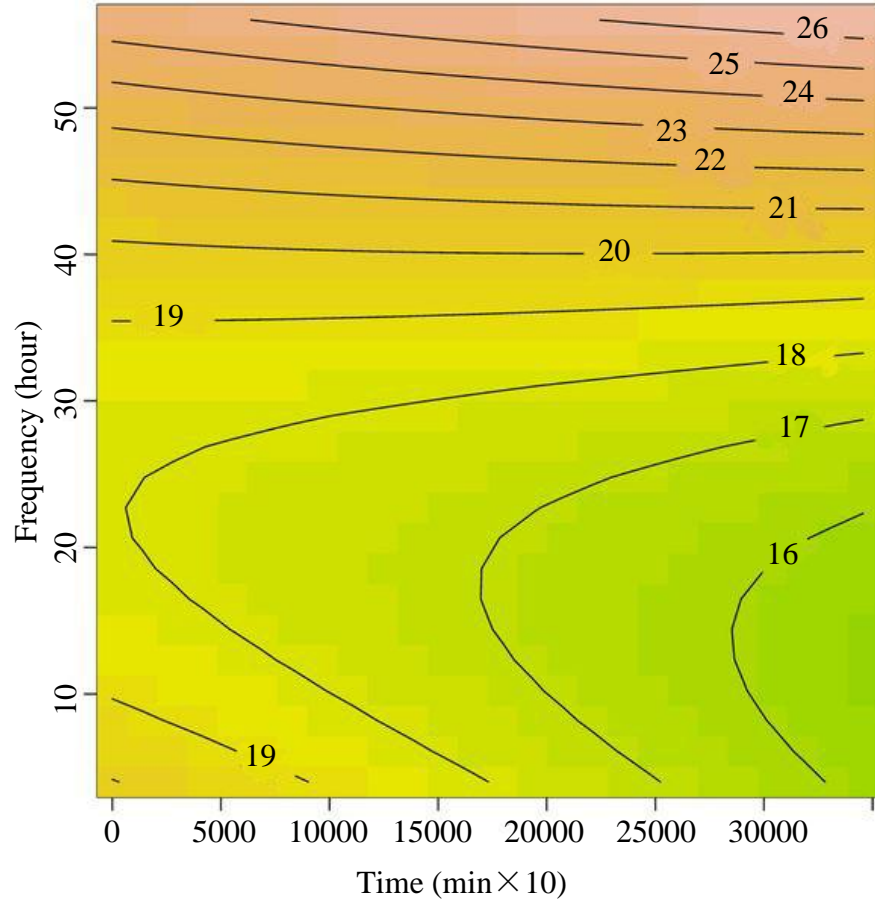


Figure 29: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading frequency and loading time (constant soil depth and loading concentration)

In Figure 30 and 31, the soil depth and loading frequency were 42.96 cm and 17.39 hour at their mean levels. At the constant loading time, the oxygen concentration decreased with increasing oxygen concentration. With the constant loading concentration, the oxygen concentration decreased over time. When the loading was less than 90 g BOD/m²/day and the loading time was controlled in 350000 min, the oxygen concentration was greater than 10%.

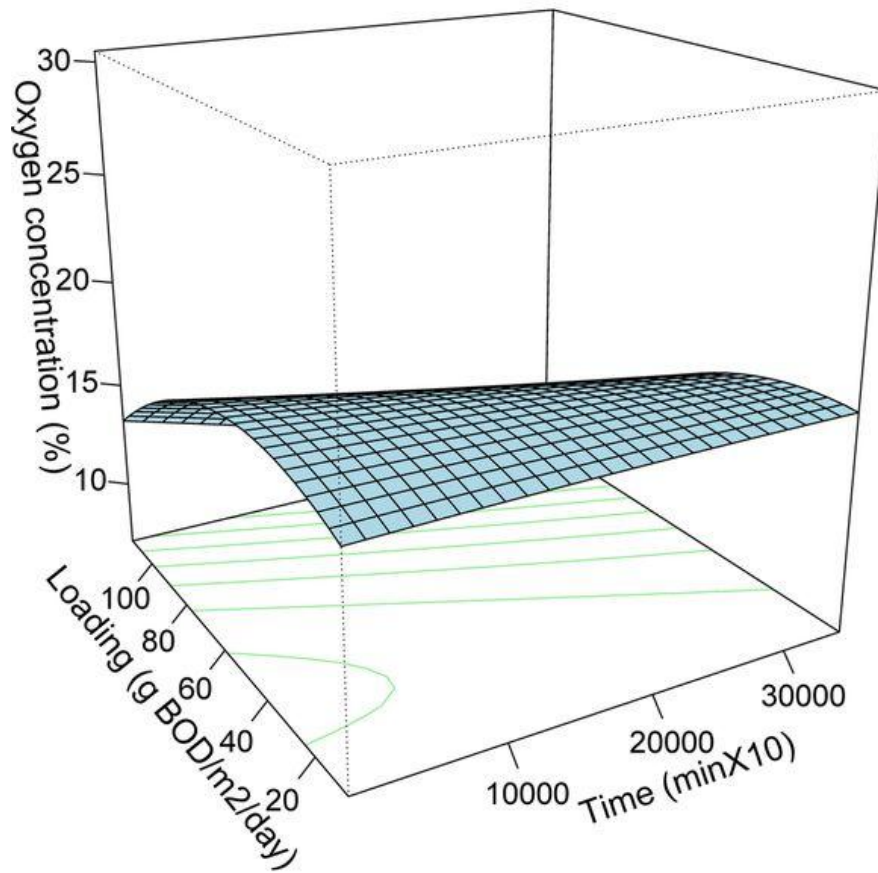


Figure 30: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)

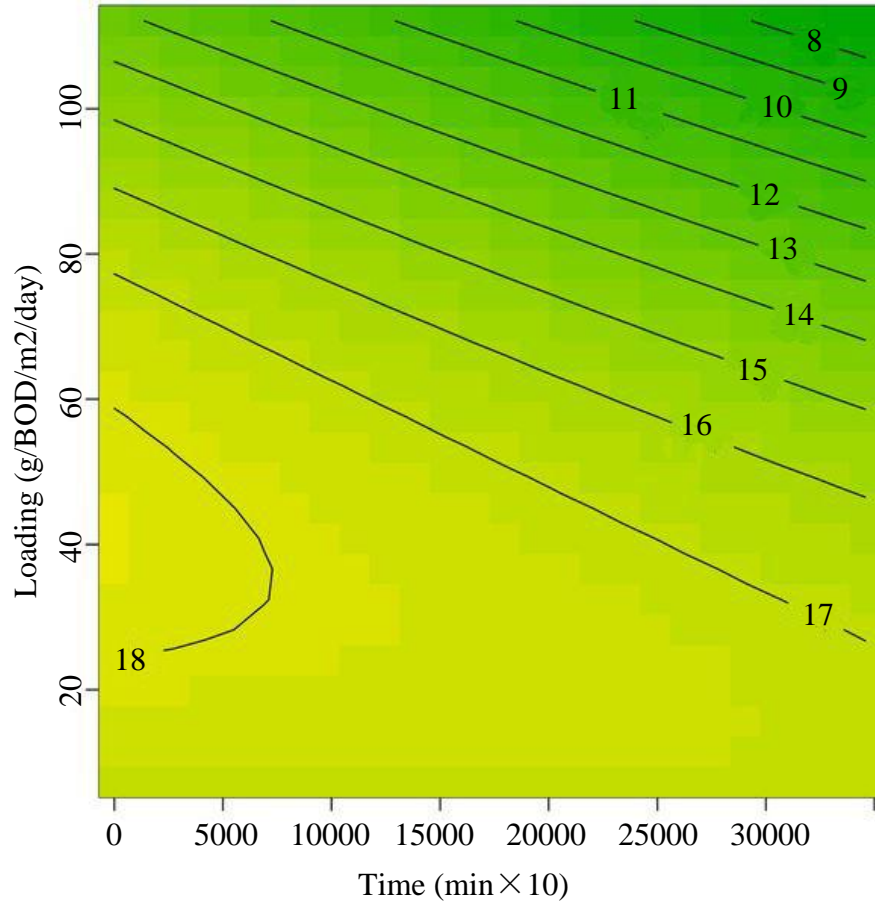


Figure 31: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and loading time (constant soil depth and loading frequency)

In Figure 32 and 33, the loading time and loading concentration were 77188.2 min and 52.49 g BOD/m²/day at their mean levels. When the soil depth was less than 30 cm, the oxygen concentration increased with increased loading frequency at the constant soil depth. When the soil depth was greater than 30 cm, the oxygen concentration decreased and then increased with the increased loading frequency. When the loading frequency was less than 20 hour, the oxygen concentration increased with deeper soil depth and the constant loading frequency. When the loading frequency was greater than 20 hour, the oxygen concentration change was opposite. The lowest oxygen concentration was 18%.

However, this happened at the loading time of 77188.2 min. From the previous analysis, at the constant soil depth, the oxygen concentration decreased over loading time according to Figure 26. Combined with the above analysis, for each soil depth in this condition, the oxygen concentration decreased over time when the loading frequency was constant. Until the loading time was 77188.2 min, the minimum oxygen concentration was 18%, still greater than 10%. With increasing time, the oxygen concentration may be less than 10% at any soil depth.

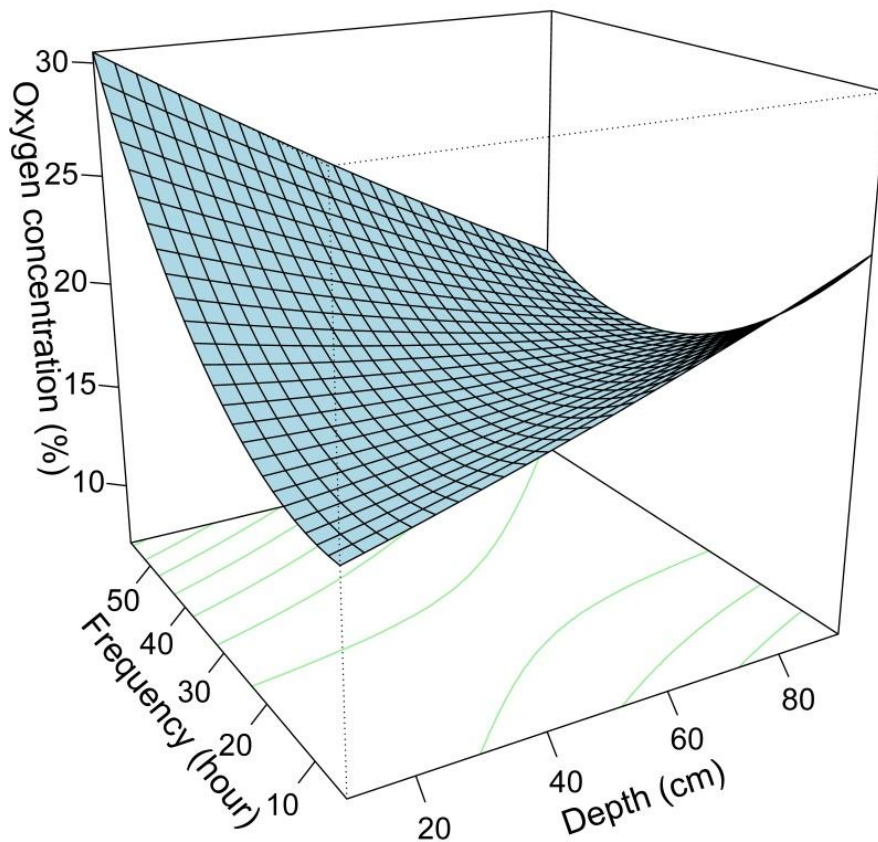


Figure 32: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)

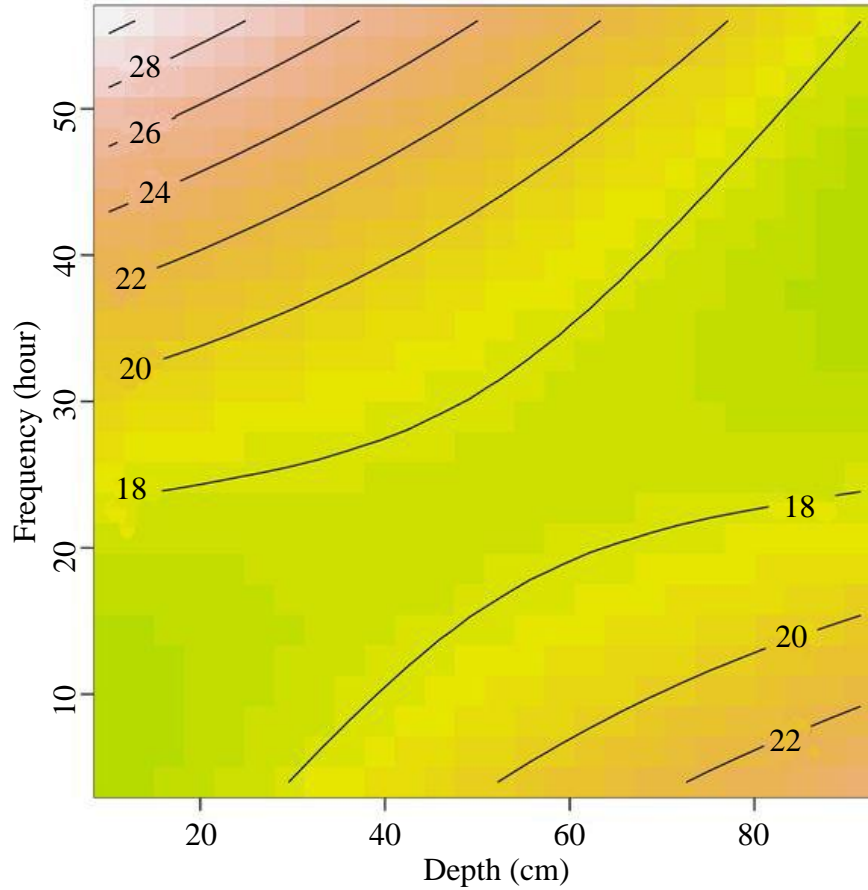


Figure 33: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading frequency and soil depth (constant loading time and loading concentration)

In Figure 34 and 35, the loading time and loading frequency were 77188.2 min and 17.39 hour at their mean levels. At a soil depth less than 40 cm, the oxygen concentration decreased when the loading concentration increased. With loading concentration higher than a certain value over $100 \text{ BOD/m}^2/\text{day}$, the oxygen concentration may be less than 10%. When the soil depth was greater than 40 cm, the oxygen concentration change was opposite. With the constant loading concentration less than $40 \text{ BOD/m}^2/\text{day}$, the oxygen concentration decreased in deeper soil. When the loading concentration was greater than

40 BOD/m²/day, the situation was opposite. The oxygen concentration was likely to be less than 10% with the deeper soil than 100 cm.

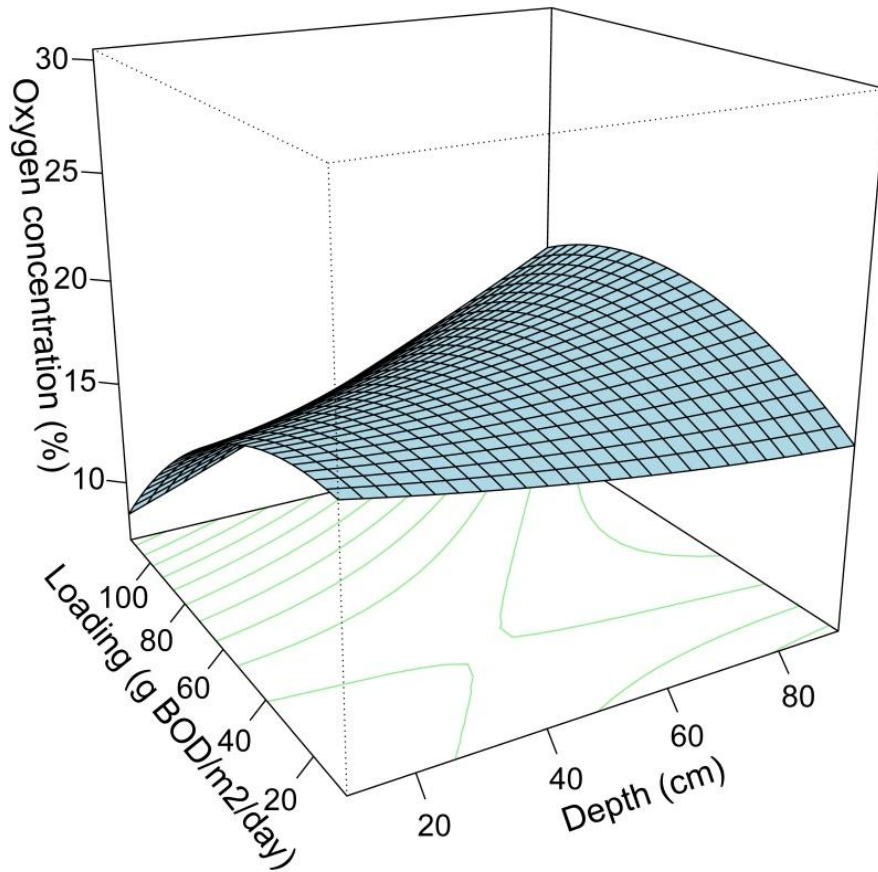


Figure 34: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)

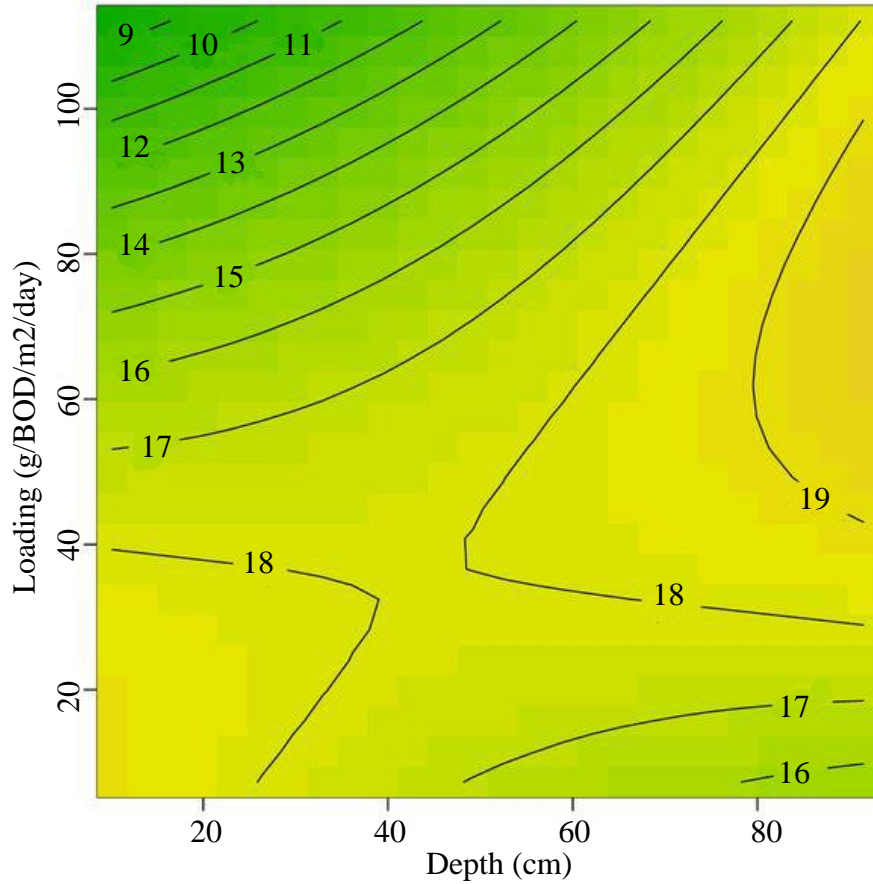


Figure 35: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and soil depth (constant loading time and loading frequency)

In Figure 36 and 37, the soil depth and loading time were 42.96 cm and 77188.2 min at their mean levels. With the constant loading frequency less than 30 hour, the oxygen concentration increased and then decreased as loading concentration increased. When the loading frequency was greater than 30 hour the oxygen concentration changed in the opposite way. If the loading concentration was larger than a certain value over 100 BOD/m²/day, the oxygen concentration was less than 10%. With the constant loading concentration, the oxygen concentration decreased first and then increased over increasing loading frequency.

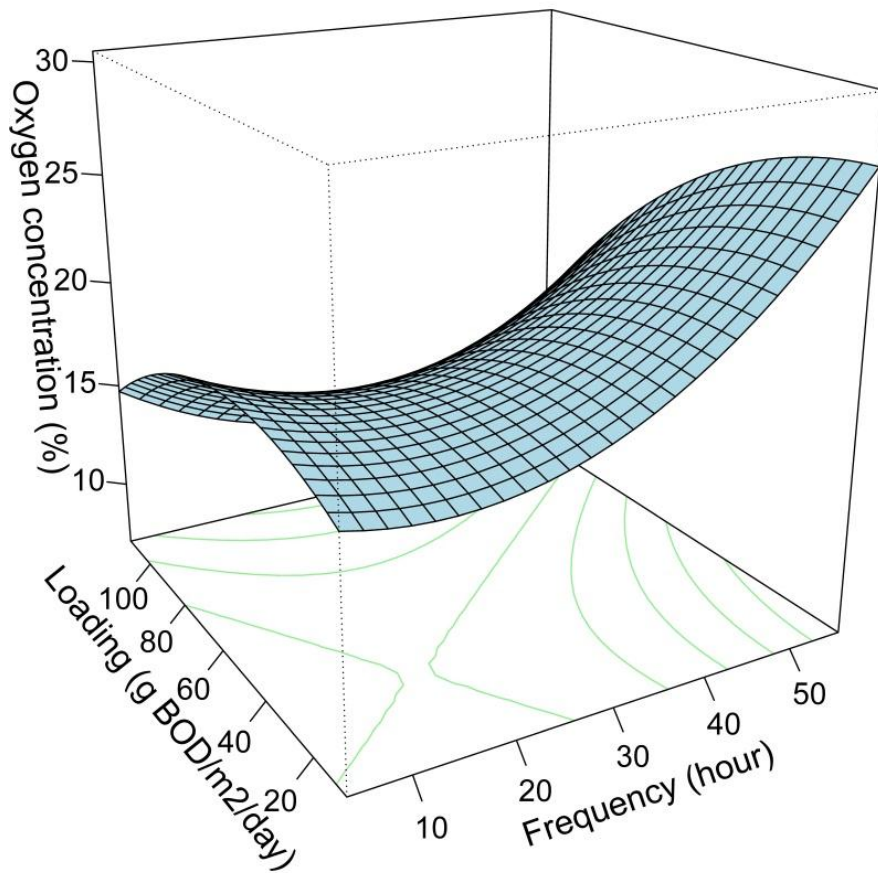


Figure 36: Oxygen concentration response surface plot under the un-submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)

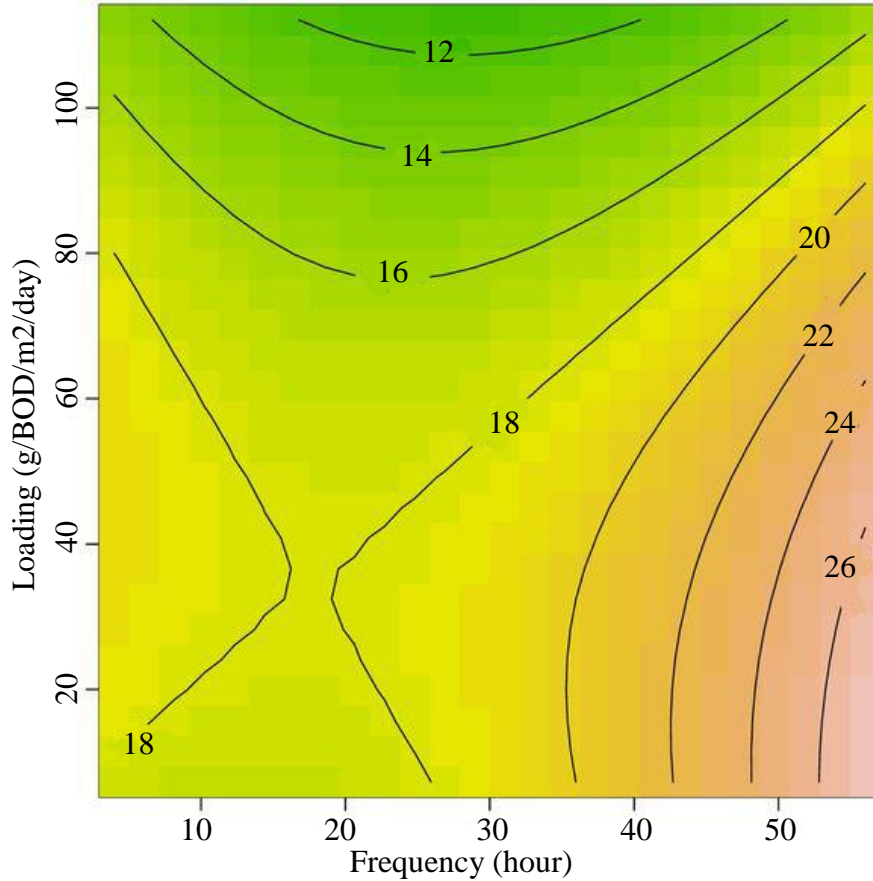


Figure 37: Oxygen concentration contour plot under the un-submerged condition showing the influence of loading concentration and loading frequency (constant soil depth and loading time)

Under un-submerged conditions, the condition to maintain aerobic soil was controlling the loading time to 350000 min, and the loading concentration was less than 90 g BOD/m²/day, although lower loading concentration was better. As the oxygen was able to diffuse from the column bottom, the soil depth was not a significant factor. In addition, the loading frequency did not show a significant effect on oxygen change.

The oxygen change is compared for submerged and un-submerged conditions. First, the oxygen change over soil depth is compared in Figures 13 and 26. The oxygen concentration increased and then decreased over depth of soil at the constant time under

the submerged conditions. This corresponds with a higher oxygen concentration for some lower sensors than that of some upper sensors. But for un-submerged condition, the oxygen increased at the constant depth. This indicates that the oxygen diffusion speed was greater than the oxygen depletion speed, as air entered through the bottom holes in the column.

Next, the effect of loading frequency on oxygen change is compared. From Figure 17 and 18, the effect of loading frequency on oxygen concentration was small, because the oxygen concentration changed from 18.5% to 24% when the loading frequency changed from 0 to 50 hour. From Figure 27 and 28, the oxygen concentration changed from 16% to 23% when the loading frequency changes from 0 to 50 hour. The effect of loading frequency on oxygen change was more obvious under un-submerged conditions.

The effect of loading concentration on oxygen change was compared in Figure 17, 18, 29 and 30. In both submerged and un-submerged conditions low loading concentration was better than the high, and the loading concentration should be limited to a certain level, which varied between submerged and un-submerged conditions. For un-submerged condition, the loading time should also be controlled.

The oxygen change with loading time was also compared. In Figure 21, 22, 33 and 34, when the loading concentration was high, the oxygen was consumed more quickly at the soil surface. Under the submerged condition with the constant loading concentration, the oxygen concentration increased and then decreased close to zero over loading time. This indicated that the nutrients were nearly exhausted in the deep soil, while oxygen diffused from the top of the column and likely did not reach the bottom of the column.

Under the un-submerged condition, the oxygen concentration changed a little when the loading concentration was less than 40 g BOD/m²/day, while increased considerably when the loading concentration was greater than 40 g BOD/m²/day. This implied that after consuming most of the oxygen the surface of soil, the oxygen in the soil column was mainly supplied by the diffusion from the column bottom.

3.2.5 Mixed-Effects Model

Similar to the RSM, the mixed-effects model did not require additional physical parameters. The mixed-effects model also simulated the oxygen concentration based on the four factors mentioned previously. However, the input data was not required to be continuous over time, and there was a random effect included in the mixed-effects model.

3.2.5.1 Governing Equation

The linear mixed-effects model was used in this study. The response and covariates of mixed-effects model were typically presented as a matrix. The governing equation of linear mix-effects model in matrix form is stated in Equation 23 and simplified as Equation 24. β was the coefficient presenting fixed effect, and ε was the error presenting random effect (Fox, 2002).

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ x_{21}, x_{22}, \dots, x_{2p} \\ \vdots \\ x_{n1}, x_{n2}, \dots, x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (23)$$

$$y = X\beta + \varepsilon \quad (24)$$

Where,

y : vector of responses in size $n \times 1$;

X : vector of covariates in size $n \times p$;

β : vector of coefficients in size $p \times 1$;

ε : vector of random errors (fixed effects) in size $n \times 1$.

By deriving Equation 23 with the y vector of one response, Equation 25 was produced.

$$(y_1) = (x_{11}, x_{12}, \dots, x_{1p}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + (\varepsilon_1)$$
$$= (x_{11}\beta_1 + x_{12}\beta_2 + \dots + x_{1p}\beta_p) + (\varepsilon_1) \quad (25)$$

The values of the first column in vector X were usually 1 (Fox, 2002). After the calculation, Equation 25 was derived as Equation 26:

$$y_1 = \beta_1 + x_{12}\beta_2 + \dots + x_{1p}\beta_p + \varepsilon_1 \quad (26)$$

In our study, the four affecting factors, loading time, soil depth, loading frequency, and loading concentration, were identified as x_1 , x_2 , x_3 and x_4 , respectively. In addition, the mutual effects between each two factors (x_1x_2 , x_1x_3 , x_1x_4 , x_2x_3 , x_2x_4 , and x_3x_4), or the factor itself were considered (x_1^2 , x_2^2 , x_3^2 , and x_4^2). In order to test whether these

mutual effects affecting the oxygen concentration in soil, four sub-models of mixed-effects named L, L+I, L+Q, L+I+Q were developed according to Equation 26 as follows (Equation 27 to 30). L+I+Q was the full mixed-effects model. The governing equations for the submerged and un-submerged conditions were the same.

$$L: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \alpha \quad (27)$$

$$L+I: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \alpha \quad (28)$$

$$L+Q: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{44}x_4^2 + \alpha \quad (29)$$

$$L+I+Q: Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{44}x_4^2 + \alpha \quad (30)$$

Y (%): predicted oxygen concentration;

x_1 (min): loading time;

x_2 (cm): soil depth;

x_3 (hour): loading frequency;

x_4 (g BOD/m²/day): loading concentration;

β : coefficients (fixed effect);

α : random errors (random effect).

Every repeated experiment produced one random effect. For each sub-model, the random effects in each experimental condition were presented as a total random effect, α . In hypothesis, the distribution of random effects was normal; the mean (μ) equal to zero; and the variance (σ^2) was known. Based on the three sigma rule shown in Equation 31, there was 99.73% probability for the random effect being $\{\mu-3\sigma < \alpha < \mu+3\sigma\}$, i.e. $-3\sigma < \alpha < 3\sigma$ (Smirnov and Dunin-Barkovskii, 1969). When the oxygen concentration in soil was measured with all conditions of loading time, soil depth, loading frequency, and loading concentration under either submerged or un-submerged condition, the total random effect, α , for both submerged and un-submerged condition should equal to zero (Cheung, 2008) due to the true standard deviation, that is the standard deviation for the whole population, was zero. Therefore, when the model was used to predict the oxygen concentration in soil at any condition within the condition range in our study, the random effect in each sub-model should be zero.

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0.99730 \quad (31)$$

Where,

X: random variable;

μ : mathematical expectation of the random variable;

σ : standard deviation.

3.2.5.2 Input and Output

The input data was the same as RSM except for two differences. The input data was required to be continuous in RSM, but in the mixed-effects model this is not a requirement. Moreover, the manner in which a dataset was identified was different. In RSM, a data set must be continuous and under the same experimental conditions at the same time. In the mixed-effects model, the data must be collected under the same experimental conditions, but does not have to be continuous. For example, some of the data from column 1 in phase A with 7 g BOD/m²/day loading concentration and 6 hour frequency under submerged condition (Figure 3) was missing. Even though the data is discontinuous, the experimental conditions were the same, resulting in a single dataset.

The input data was processed by the *lme4* package in R software. The categories of output for mixed-effects model were the same as the RSM (Table 13 and 14). The fixed effects represented as Variables. Intercept was the constant coefficient, β_0 . Time, Depth, Frequency, and Loading indicated the coefficients, β_1 , β_2 , β_3 , and β_4 , respectively. Time², Depth², Frequency², and Loading² indicated the coefficients, β_{11} , β_{22} , β_{33} , and β_{44} ,

respectively. (Time: Depth), (Time: Frequency), (Time: Loading), (Depth: Frequency), (Depth: Loading), and (Frequency: Loading) were the coefficients, β_{11} , β_{12} , β_{13} , β_{14} , β_{23} , β_{24} , and β_{34} , respectively.

The random effect in the full mixed-effect model predicted by the several experimental conditions in our study was yielded separately. The output of random effect α was different in the submerged and un-submerged conditions. For each experiment, there was a random effect α . The estimated standard deviation error was 0.779 with a 95% confidence level of 0.778 to 0.781. The estimated standard deviation for the random effect, due to experiment, however, was 5.062 with a 95% confidence level of 4.146 to 6.177. Therefore, we conclude the random effect for the experiment had explained $5.062 / (5.062 + 0.7792) = 97.68\%$ of total variation of the experiments. Similarly for un-submerged columns, there were 26 combinations of experiments with 12,074 observations points on average for each experiment based on raw data. From the model, the estimated standard deviation error was 0.920 with confidential limits of 0.918 to 0.923 and that for random effect due to experiment was 2.917 with confidential limits of 2.21 to 3.85. Hence 90.95% of the total variation can be explained by experiment effect.

Table 13: Output of the full mixed-effects model under submerged condition

Variables	Estimate	Std. Error	t statistic	p value	Lower 95% limit	Upper 95% limit
Intercept	12.2743	5.984	2.05	0.0403	0.5456	24.0030
Time	-5.57E-05	1E-6	-52.95	<.0001	-5E-5	-5E-5
Depth	0.2044	0.108	1.88	0.0674	-0.0154	0.4242
Frequency	0.2908	0.608	0.48	0.6354	-0.9426	1.5243
Loading	0.0618	0.156	0.39	0.6950	-0.2551	0.3786
Time ²	4.63E-10	3E-11	15.96	<.0001	4E-10	5E-10
Depth ²	-0.0027	0.001	-3.63	0.0009	-0.0043	-0.0012
Frequency ²	0.0044	0.004	0.96	0.3432	-0.0049	0.0137
Loading ²	-0.0008	0.001	-0.65	0.5143	-0.0033	0.0017
Time: Depth	5.90E-07	9E-9	68.74	<.0001	6E-7	6E-07
Time: Frequency	7.44E-08	4E-8	2.06	0.0391	4E-9	1E-07
Time: Loading	-7.33E-07	8E-9	-94.99	<.0001	-7E-7	-7E-07
Depth: Frequency	-0.0017	0.002	-1.06	0.2957	-0.0049	0.0015
Depth: Loading	0.0009	0.001	1.31	0.1995	-0.0005	0.0023
Frequency: Loading	-0.0040	0.006	-0.64	0.5245	-0.0168	0.0087

Table 14: Output of the full mixed-effects model under un-submerged condition

Variables	Estimate	Std. Error	t-value	p-value	Lower 95% limit	Upper 95% limit
Intercept	16.9018	4.194	4.02	5E-5	8.6805	25.1232
Time	-9.73E-06	1E-6	-6.65	2E-11	-1E-5	-7E-6
Depth	-0.0138	0.103	-0.13	0.8948	-0.2323	0.2046
Frequency	0.0618	0.450	0.13	0.8926	-0.8933	1.0170
Loading	0.0624	0.111	0.55	0.5841	-0.1744	0.2992
Time ²	-1.04E-09	4E-11	-26.43	8E-15	-1E-9	-10E-10
Depth ²	0.0004	0.001	0.47	0.6423	-0.0014	0.0023
Frequency ²	0.0073	0.004	1.92	0.0725	-0.0008	0.0154
Loading ²	-0.0010	0.001	-1.14	0.2701	-0.0030	0.0009
Time: Depth	1.97E-07	2E-8	12.57	3E-36	1E-7	2E-7
Time: Frequency	6.46E-06	6E-8	100.59	<.0001	6E-6	6E-6
Time: Loading	-3.98E-06	2E-8	-174.19	<.0001	-4E-6	-4E-6
Depth: Frequency	-0.0051	0.002	-2.74	0.0145	-0.0091	-0.0012
Depth: Loading	0.0017	0.001	2.11	0.0504	-3E-6	0.0035
Frequency: Loading	-0.0028	0.005	-0.58	0.5666	-0.0130	0.0073

3.2.5.3 Result

The best sub-model of mixed-effects was found by the comparisons of AIC and BIC for each sub-model, similar to the approach employed for RSM. For both submerged and un-submerged conditions (Table 15 and 16), L+I+Q had the highest log-likelihood, and the lowest AIC and BIC, indicating the sub-model (the full mixed-effects model) was the best model. The best mixed-effects models under submerged and un-submerged conditions are shown in Equation 32 and Equation 33, respectively..

Table 15: Mixed-effects sub-model comparisons under submerged condition

Model	Log-likelihood	AIC	BIC
L	-640266	1280545	1280624
L+Q	-638908	1277837	1277960
L+I	-632742	1265511	1265657
L+I+Q	-632609	1265251	1265442

Table 16: Mixed-effects sub-model comparisons under un-submerged condition

Model	Log-likelihood	AIC	BIC
L	-436761	873535	873609
L+Q	-434823	869667	869784
L+I	-419866	839758	839897
L+I+Q	-419514	839062	839243

$$\begin{aligned}
 Y = & 12.2743 + (-5.57 \times 10^{-6}x_1 + 0.2044x_2 + 0.2908x_3 + \\
 & 0.0618x_4) + (4.63 \times 10^{-12}x_1^2 - 0.0027x_2^2 + 0.0044x_3^2 - \\
 & 0.0008x_4^2) + (5.90 \times 10^{-8}x_1x_2 + 7.44 \times 10^{-9}x_1x_3 - 7.33 \times \\
 & 10^{-8}x_1x_4 - 0.0017x_2x_3 + 0.0009x_2x_4 - 0.0040x_3x_4) \quad (32)
 \end{aligned}$$

$$\begin{aligned}
 Y = & 16.9018 + (-9.73 \times 10^{-7}x_1 - 0.0138x_2 + 0.0618x_3 + \\
 & 0.0624x_4) + (-1.04 \times 10^{-11}x_1^2 - 0.0004x_2^2 + 0.0073x_3^2 - \\
 & 0.0010x_4^2) + (1.97 \times 10^{-8}x_1x_2 + 6.46 \times 10^{-7}x_1x_3 - 3.98 \times \\
 & 10^{-7}x_1x_4 - 0.0051x_2x_3 + 0.0017x_2x_4 - 0.0028x_3x_4) \quad (33)
 \end{aligned}$$

Where,

Y (%): the predicted concentration of oxygen;

x_1 (min): loading time;

x_2 (cm): soil depth;

x_3 (hour): loading frequency;

x_4 (g BOD/m²/day): loading concentration.

3.2.6 Time Series Analysis

In this study, the periodogram and seasonal decomposition were used to analyze the oxygen concentration series. The periodogram was used to predict the frequency of the time series, and the seasonal decomposition was used to forecast the amplitude and slope of the time series. The amplitude, slope, and frequency were hypothesized to be related to the oxygen concentration greater than 10%, (aerobic soil environment) for further investigation. The details of the analysis are explained in this section.

3.2.6.1 Governing Equation

First, the frequency was predicted by the periodogram. There were both time domains and frequency domains. The original time series was in time domain, which meant the signal was changed over time. The frequency domain indicated that the analysis was according to the frequency rather than time. The time series was transformed into a frequency domain to predict the frequency. This transformation was

based on Discrete Fourier Transform as shown in Equation 34 (Broughton and Bryan, 2008).

$$d(\omega_j) = n^{-\frac{1}{2}} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t} \quad (34)$$

Where,

ω_j (radians/s): angular frequency, recall $\omega_j = j-1/n$ where $j = 1, 2, \dots, N$;

x_t : original data.

Because several estimated frequencies were fit for the original time series, the best fit frequency was selected. In the frequency domain, the periodogram was created to get the best fit frequency. The periodogram was produced by the package “*periodogram*” in R software. Periodogram was also the name of a function calculated based on the discrete Fourier transform (Equation 35) (Vaughan and Uttley, 2006).

$$I(\omega_j) = |d(\omega_j)|^2 \quad (35)$$

On the periodogram here, the x-axis represented the estimated loading frequency. The y-axis represented the spectral density. The spectral density represents the power of every herz, which indicates the number of data points located in the estimated frequency band in our research. The highest and sharpest peak on the periodogram indicated the largest and concentrated power. Therefore, the frequency with the highest and sharpest peak on the periodogram was the best fit frequency. However, if the highest peak corresponded to zero or was not sharp, the absence of periodicity was suspicious.

According to the reciprocal of best fit frequency, the best fit period was calculated (Schuster, 1898).

Based on the predicted best fit period, the time series was broken down into sub-series by the seasonal decomposition. The seasonal decomposition used an additive model as shown in Equation 36.

$$Y_t = S_t + T_t + R_t \quad (36)$$

Where,

Y_t : data point value of time series at period t ;

S_t : seasonal component at period t ;

T_t : trend cycle component at period t ;

R_t : remainder component at period t .

The seasonal component was calculated based on the mean of the sub-series. After remove the seasonal component, the remainder was left. The remainder was fitted with a linear regression as the trend. The procedure to extract the seasonal, trend, and the remainder components was implemented by *stl* package in the statistical software R.

3.2.6.2 Input and Output

Among the studied scenarios, the time series of long column experiments with the loading rate of 112 g BOD/m²/day and the frequency of 24 and 56 hour exhibited a clearer pattern. Therefore, they were studied in more detail. The periodogram was used to

estimate the frequency of oxygen changes in soil. Then, the period was predicted based on the reciprocal of frequency. As demonstrated in Figure 38, the periodogram of the 56 hour at a 30.48 cm soil depth (Position2) displays possible periodicity as manifested in the serial plot. There was a highest and sharpest peak, which indicated predicted frequency corresponding to the dominated frequency. For the other conditions without significantly possible periodicity, the predicted frequencies were not fit adequately.

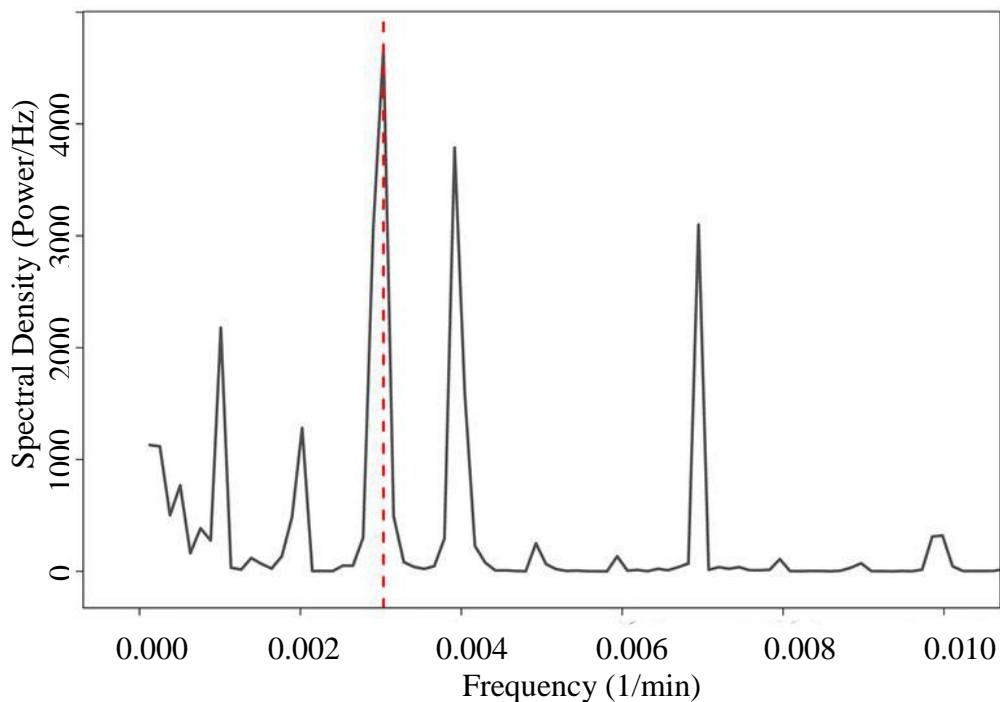


Figure 38: Periodogram of 112 g BOD/m²/day loading concentration, 56 hour loading frequency, and 30.48 cm soil depth

There was another case for no frequency as demonstrated by the periodogram of the 24 hour loading frequency at a 111.76 cm soil depth (Position 6) shown in Figure 39. There were only low and smooth peaks in the middle. So there was no frequency for this data set. For the curves measured by the bottom sensor, most of them were flat, indicating no frequency for these data sets.

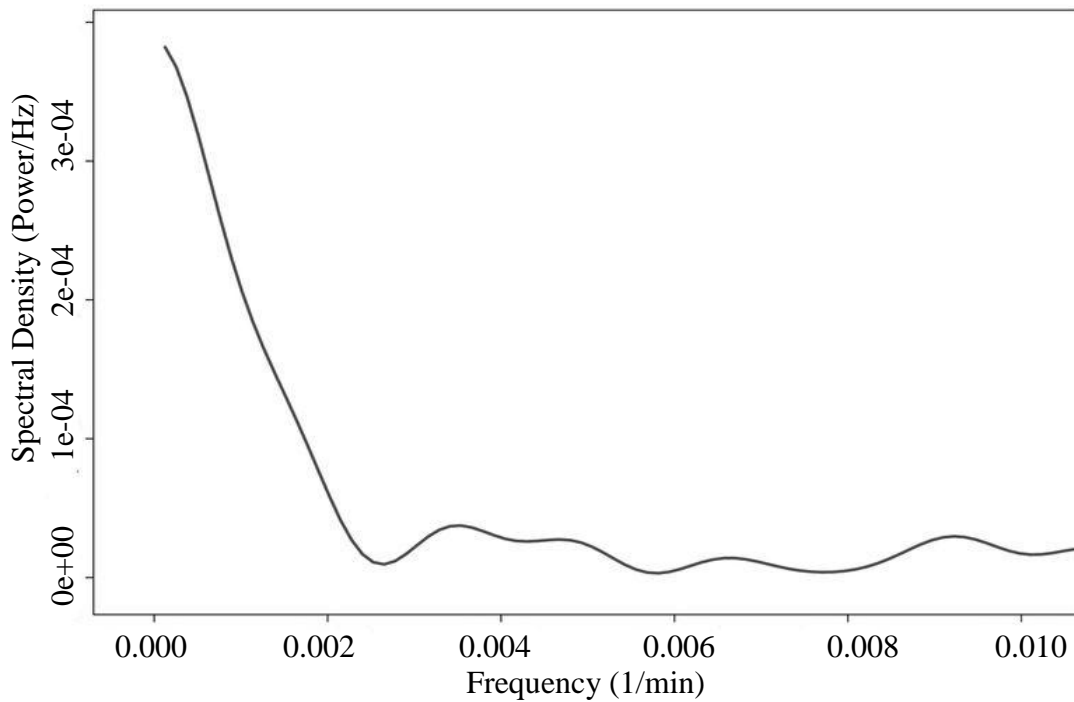


Figure 39: Periodogram of 112 g BOD/m²/day loading concentration, 24 hour loading frequency, and 111.76 cm soil depth

The seasonal decomposition for time series at the condition of 112 g BOD/m²/day organic loading concentration, 24 hour frequency, and 10.16 cm depth (Figure 40) was examined according to the periodogram. The seasonal component of oxygen concentration was smoothed but it still contained a pattern similar to the original time sequence. Beside the seasonal component of oxygen concentration, other decomposed components were found to be reasonable from the graphic results because the ranges of the seasonal term (-2 to 2) and residual terms (-4 to 4) were small, and the trend was negative and linear. Thus, the overall trend matched the pattern presented by the raw data. Furthermore, the remainders after the seasonal component and trend were extracted had a relatively small value compared to the data. The seasonal decomposition of other loading conditions is presented in the appendix.

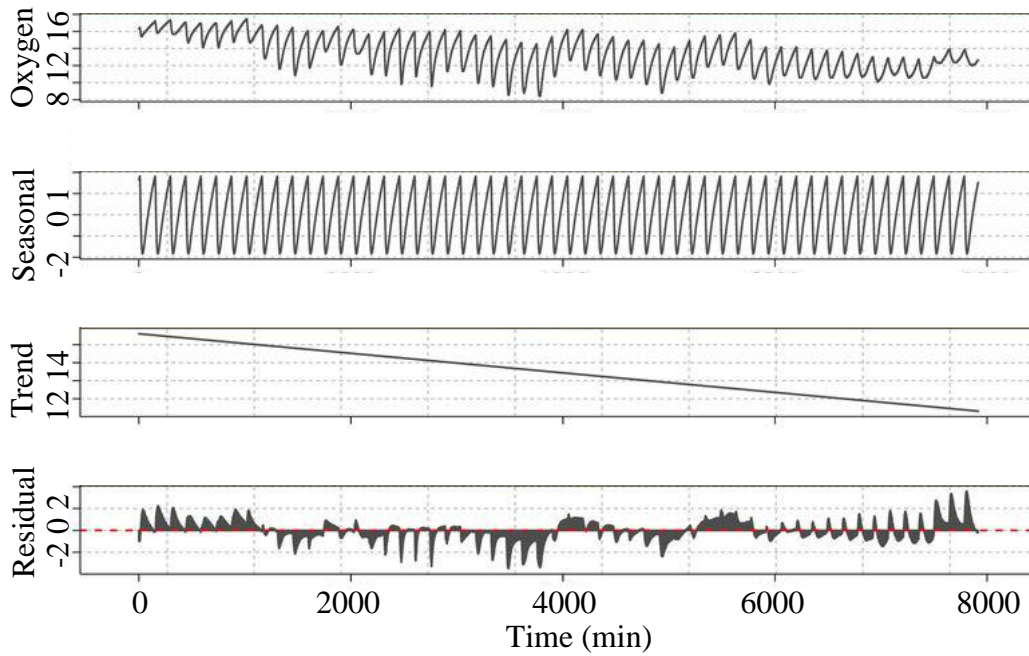


Figure 40: Illustration of seasonal decomposition for time series

The amplitude, slope, period of oxygen change in soil were predicted. Mean absolute percentage error (MAPE) (Equation 36) for the residual was used to evaluate the fitness of the seasonal component to the original time series produced by the measured oxygen concentration (Table 17). Smaller values indicate the better fit.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (36)$$

Where,

A_t : actual value;

F_t : forecast value.

Table 17: Data used for model calibration

Experimental Conditions			Predicted Graph Characteristics				Observed
Loading (g BOD /m ² /day)	Freq. (hour)	Depth (cm)	Amp. (%)	Slope	Freq. (hour)	MAPE	Oxygen Con. (%)
112.08	24	10.16	3.64	-0.0005	24	0.06	13.04
112.08	24	30.48	2.11	-0.0004	24	0.04	13.58
112.08	24	50.80	1.52	-0.0004	24	0.04	13.47
112.08	24	71.12	1.13	-0.0003	24	0.03	15.48
112.08	24	91.44	1.10	-0.0003	24	0.03	14.13
112.08	24	111.76	-	-	-	-	-
112.08	56	10.16	0.28	3E-5	188	0.005	20.33
112.08	56	30.48	3.31	-0.0002	55	0.17	13.18
112.08	56	50.8	2.68	-0.0001	55	0.13	14.31
112.08	56	71.12	2.32	-0.0001	55	0.11	14.60
112.08	56	91.44	2.05	-6E-5	55	0.10	14.40
112.08	56	111.76	0.99	7E-5	188	2E+8	0.67

Note: Because there was no obvious pattern for the oxygen at the condition of 112.08 g BOD/m²/day, 24 hour frequency and depth of 111.76 cm, the depth (cm), amplitude, slope, predicted frequency and MAPE were not calculated.

A linear regression model was created using loading frequency and soil depth from the experimental conditions and amplitude, slope and frequency estimated through time series analysis (Table 17). The model only included the data from columns 7 and 8 in phase C, which were in complete oscillation. The regression model was fitted with the ‘Data Analysis’ package in Microsoft Excel.

Table 18 shows the results of the linear regression and Figure 41 shows observed values and fitted values plotted against soil depth. The adjusted coefficient of determination, R^2 , of the predicted model was 0.77. The fitting produced an unadjusted R^2 value of 0.88, which indicated a good fit of linear regression model to the data.

Moreover, the ANOVA table shows that the model is significant at $\alpha=0.05$ ($F=0.02$). The individual parameter p-values show that depth, frequency and amplitude were the most useful parameters in estimating oxygen concentration, while loading frequency and slope were the least.

Table 18: Regression statistics for the model created using two experimental conditions and three time series-estimated parameters

Multiple R		0.94				
R Square		0.88				
Adjusted R Square		0.77				
Standard Error		2.25				
Observations		11				
ANOVA						
	df	SS	MS	F	Significance F	
Regression	5	193.18	38.64	7.65	0.02	
Residual	5	25.24	5.05			
Total	10	218.42				
Parameter	Coefficient	Stand. Error	t-stat.	P-value	Lower 95%	Upper 95%
Intercept	34.89	14.98	2.33	0.067	-3.61	73.40
Loading Freq.	0.11	0.26	0.42	0.691	-0.56	0.78
Depth	-0.15	0.04	-4.15	0.009	-0.24	-0.06
Amp.	-3.95	2.18	-1.81	0.130	-9.55	1.66
Slope	15580	27610	0.56	0.60	-55393	86553
Freq.	-0.11	0.02	-4.50	0.006	-0.17	-0.05

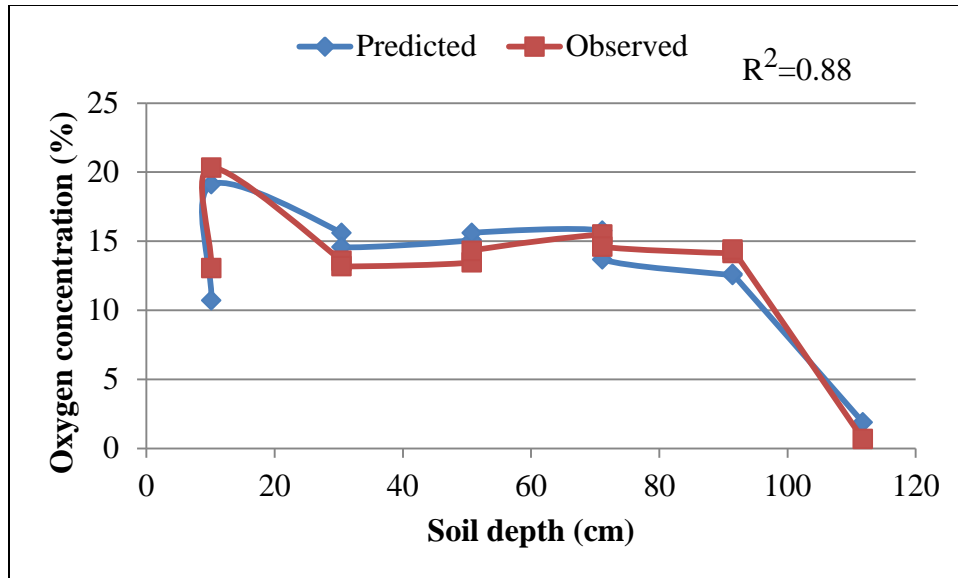


Figure 41: Linear model from time series analysis

Next, the model was validated using submerged and un-submerged data separately (Table 19). Submerged data included all the data collected under submerged condition that had complete oscillations. Un-submerged data included all data collected under submerged condition except those used to create the model.

Table 19: Data used for model validation

Column No.-Phase	Loading (g BOD /m ² /day)	Loadin g Freq. (hour)	Depth (cm)	Amp. (%)	Slope	Freq. (hour)	Mean O ₂ (%)
Submerged Condition							
1-A	7	6	10.16	0.25	9E-6	17	17.08
1-A	7	6	30.48	0.13	2E-5	13	17.16
1-A	7	6	50.80	0.09	-4E-6	17	16.19
7-A	28	12	50.80	0.36	5E-7	50	19.82
7-A	28	12	71.12	0.39	2E-5	56	19.48
7-A	28	12	91.44	0.41	-1E-4	57	19.99
7-B	112	12	30.48	0.87	-1E-4	13	15.79
7-B	112	12	50.80	0.78	-2E-4	55	17.6
7-B	112	12	71.12	0.69	-1E-4	67	17.47
7-B	112	12	91.44	0.63	-1E-4	76	17.11
Un-submerged Condition							
4-A	28	12	30.48	0.26	-1E-5	20	16.73
4-A	28	12	50.80	0.23	-4E-5	17	16.94
8-A	28	12	10.16	0.70	-5E-5	46	18.33
8-A	28	12	30.48	0.47	-2E-5	56	17.52
8-A	28	12	50.80	0.48	-2E-5	70	17.14
8-A	28	12	71.12	0.47	-6E-5	70	19.6
8-A	28	12	91.44	0.45	-9E-5	77	17.47
6-B	112	6	10.16	0.41	-0.001	20	4.18
6-B	112	6	30.48	0.20	6E-5	20	17.19
8-B	112	12	10.16	0.89	-3E-5	10	16.78
8-B	112	12	30.48	0.47	-4E-6	13	16.37
8-B	112	12	50.80	0.35	4E-6	17	16.11
8-B	112	12	71.12	0.27	-2E-5	15	17.44
8-B	112	12	91.44	0.51	-7E-6	59	16.13
6-C	112	24	10.16	0.14	3E-5	15	20.33

Figure 42 and 43 show the observed data and the validation results plotted against soil depth. The fitting resulted in R^2 values of 0.14 and 0.02 for submerged and un-submerged conditions, respectively. Low R^2 values indicate that the model is not useful in predicting oxygen concentration in both submerged and un-submerged conditions.

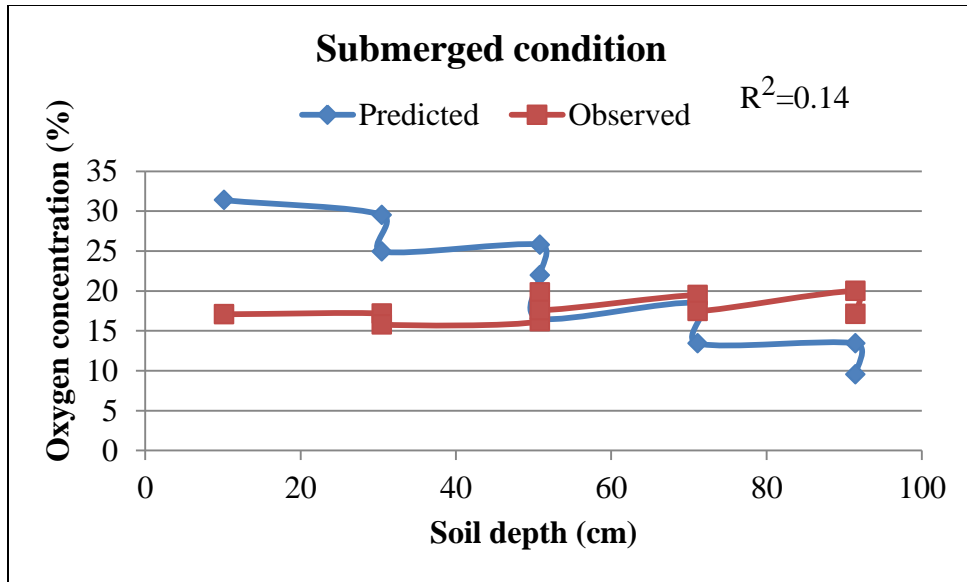


Figure 42: Validation for linear model from time series analysis under submerged condition

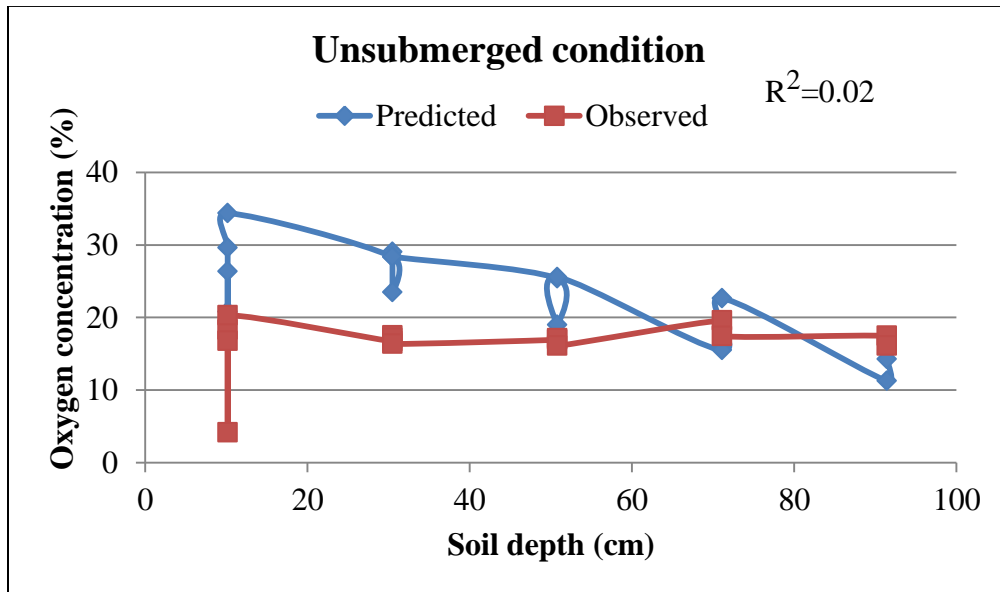


Figure 43: Validation for linear model from time series analysis under un-submerged condition

3.2.6.3 Discussion

Although the amplitude, slope, and frequency of time series were predicted, the time series analysis method was only applied when the experimental data displayed clear cycles. Also, the relationship between the loading condition, oxygen concentration, and

predicted parameters was not found. Finally, this method was not practical to discover conditions for aerobic soil.

3.3 MODEL COMPARISON

The physical models (TOUGHREACT, HYDRUS, and PHREEQC) and time series analysis were incapable of predicting oxygen concentration based on the data obtained in the soil column experiments. Only RSM and the mixed-effects model could predict the oxygen concentration using the data. These two models were compared using AIC and BIC to find the best fit model after calculating the coefficients of each model. The best fit models were evaluated by RSR (Ratio of the root mean square error to the standard deviation), NSE (Nash-Sutcliffe model efficiency coefficient), and PBIAS (percent bias). The details of the comparisons between RSM and mixed-effects model are discussed below.

3.3.1 Criterion for Model Fitness

Three criteria (RSR, NSE, and PBIAS) were selected to evaluate the fitness of the best models of RSM and mixed-effects models. RSR was calculated by the root mean square error (RMSE) dividing the standard deviation (STDEV) of observations, which was calculated by Equation 37. RSR measures residual variation of the predicted value given the observations. The optimal value 0 indicated an accurate model simulation (Moriassi et al., 2007).

$$\text{RSR} = \frac{\text{RMSE}}{\text{STDEV}} = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (37)$$

Where,

n: sample size;

y_i : observed value;

\hat{y}_i : predicted value;

\bar{y} : average of observed value.

NSE compares the residual variance to the variance from observations (Nash and Sutcliffe, 1970). The optimal value of NSE to evaluate the model fitness was 1. The model with NSE value between 0 and 1 was regarded as acceptable. The calculation of NSE was showed in Equation 38 (Moriasi et al., 2007).

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (38)$$

Where,

n: sample size;

y_i : observed value;

\hat{y}_i : predicted value;

\bar{y} : average of observed value.

PBIAS represented the the average tendency of the simulated oxygen concentration to be larger or smaller than the observed oxygen concentration (Gupta et al., 1999). The accurate model simulation was indicated by PBIAS equal to 0. PBIAS was solved by the Equation 39 (Moriasi et al., 2007).

$$\text{PBIAS} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i) \times 100}{\sum_{i=1}^n (y_i)} \quad (39)$$

Where,

n: sample size;

y_i : observed value;

\hat{y}_i : predicted value;

\bar{y} : average of observed value;

$\hat{\bar{y}}$: average of predicted value.

3.3.2 Comparative Results under Submerged Condition

Comparing the best model, L+I+Q of RSM and mixed-effects model under submerged conditions, the value of log-likelihood of RSM was lower than that of mixed-effects model; the AIC and BIC of the RSM were greater than the mixed-effects model (Table 18). This indicates that by incorporating the random effect for each experiment, the predictive power improved. The sub-model L+I+Q of mixed-effects (Equation 32), namely the full mixed-effects model, was the best fit model of all under submerged conditions. According to the values of RSR, NSE, and PBIAS from Table 18, the best fit RSM and mixed-effects table were all acceptable. In addition, the RSR and PBIAS of the best fit mixed-effects model were close to the optimal value 0, and the NSE was close to the optimal value 1. Therefore, the best mixed-effects model had a good fit to the observations under the submerged condition.

Table 20: Model comparison under submerged condition

Model	AIC	BIC	RSR	NSE	PBIAS
RSM (L+I+Q)	3174413	3174592	0.69	0.52	0%
Mixed-effects (L+I+Q)	1265251	1265442	0.12	0.99	0%

3.3.3 Comparative Result under Submerged Condition

Comparing the best model, L+I+Q, of RSM and mixed-effects model under un-submerged conditions, the log-likelihood of the sub-model L+I+Q of the mixed-effects model was higher than that of RSM; the AIC and BIC of the sub-model L+I+Q of the mixed-effects model was less than those of RSM (Table 19). This indicates that the sub-model L+I+Q of mixed-effects (Equation 33), namely the full mixed-effects model, was the best fit model of all under un-submerged condition. After evaluating values of RSR, NSE, and PBIAS of both best fit RSM and mixed-effects models from Table 19, these two models were acceptable. For the best fit mixed-effects model, the RSR and PBIAS were close to the optimal value 0, and the NSE was close to the optimal value 1. As a result, the best mixed-effects model fit the observed values very well under un-submerged condition. However, comparing Tables 20 and 21, the submerged models performed better than that of the un-submerged.

Table 21: Model comparison under un-submerged condition

Model	AIC	BIC	RSR	NSE	PBIAS
RSM (L+I+Q)	1443280	1443451	0.67	0.56	0%
Mixed-effects (L+I+Q)	839062	839243	0.25	0.94	0%

4. CONCLUSION

This research studied models to simulate the impact of wastewater irrigation on the oxygen concentration in soil. Factors affecting oxygen changes in soil considered were loading time, soil depth, loading frequency, and loading concentration. Physical models such as TOUGHREACT, HYDRUS, and PHREEQC were investigated first but their use was found to be not feasible. Due to the limitations in these physical models, statistical models were utilized.

The RSM and linear mixed-effects model were applied to simulate the oxygen concentration in sand columns. Based on the RSM analysis with response surface plots and contour plots, the conditions to maintain aerobic conditions in the soil (greater than 10% oxygen concentration) were less than 100 cm soil depth, 68.86 g BOD/m²/day loading concentration, and 15.90 hour loading frequency under submerged conditions, or 77188.2 min loading time and 52.49 g BOD/m²/day loading concentration without loading frequency limitations under un-submerged conditions. The time series analysis predicted the amplitude, slope, and frequency of oxygen concentration change. However, how to use the predicted results of time series analysis to find the aerobic soil conditions needs further study.

Under submerged and un-submerged conditions the L+I+Q model for RSM and mixed-effects was the best fit sub-model because the AIC and BIC of L+I+Q were the lowest. Comparing L+I+Q of RSM with that of mixed-effects model, the L+I+Q of mixed-effects model resulted in lower AIC and BIC values. Thus, the L+I+Q for the

mixed-effects model was best fit model. The conclusions for agricultural application were summarized below according to this study:

- When the soil is saturated with groundwater, the conditions for irrigation to maintain soil aerobic conditions are less than $110 \text{ g BOD/m}^2/\text{day}$ loading concentration, and within 100 cm soil depth, while lower BOD is preferred. The loading time and frequency contribute little to oxygen concentration in the soil.
- When the groundwater table is deep, and its impact on the aerobic soil column is minimal, the conditions for maintaining soil aerobic conditions are less than $90 \text{ g BOD/m}^2/\text{day}$ loading concentration and a low loading time. The critical value of loading time depends on the specific loading concentration, loading frequency and the range of soil depth. The loading frequency and soil depth do not significantly affect the oxygen concentration.
- The reported values to maintain aerobic soil should be used with caution due to the following reasons. The number of replications and sampling depths differed between submerged and un-submerged conditions due to sensor failure in some columns, Soil depth and time of observation were limited. Therefore, average conditions were different between submerged and un-submerged conditions. In addition, the un-submerged condition did not occur in some columns due to wastewater accumulation at the bottom of the column.
- The full mixed-effects model (sub-model L+I+Q), is the best fit model. This model was able to predict oxygen concentration at any depth of soil with any loading concentration, frequency, and time for submerged and un-submerged conditions within

conditions range in this research. For the condition out of the range, this model needs further test.

5. FUTURE RESEARCH

This study offered a valuable approach to discover the conditions for maintaining aerobic conditions in the soil when irrigated with food processing wastewater. Parameters examined included soil depth, loading frequency, loading concentration, and loading time. There is still more that needs to be done to enhance understanding of the relationship between soil aerobic conditions and chemical, physical, and microbial factors. Future studies are needed to enhance the understanding of the following aspects:

- This study assumed a 10% oxygen concentration minimum as the aerobic soil environment to avoid metal mobilization. This critical value 10% was hypothesized according to the activities of microorganisms. Determining the true critical value of oxygen concentration to immobilize metals in soil is still needed because it will help avoid pollution of drinking water.
- In this study, the distribution and activities of aerobic and anaerobic microorganisms in the soil column significantly affected the experimental results. Further research is should focus on the specific species of these microorganisms. The oxygen consumption by microorganisms in different positions with different loading is still needed. This will facilitate better interpretation of experimental data, and develop rules of oxygen concentration in the soil to improve irrigation.
- The loading time should be controlled to keep the soil aerobic when the groundwater is deep under the soil. However, the optimal time interval between each loading period is not known. Moreover, the critical distance of the groundwater table to soil surface to ignore the existence of groundwater is unknown. Understanding these

issues and applying them to the models developed in this study will better emulate real agricultural activities.

- To examine the relationship between oxygen concentration and additional physical, chemical, and biological factors such as water content, organic matter, soil temperature, air pressure, and biochemical reactions will be beneficial. The studied models (TOUGHREACT, HYDRUS, and PHREEQC) have limited or no capability to estimate oxygen concentration in soils as a result of direct wastewater discharge to soil. However, these limitations can be eliminated by developing a new set of hybrid models, which combine two or more of these models to create a more comprehensive simulation tool.

APPENDIX

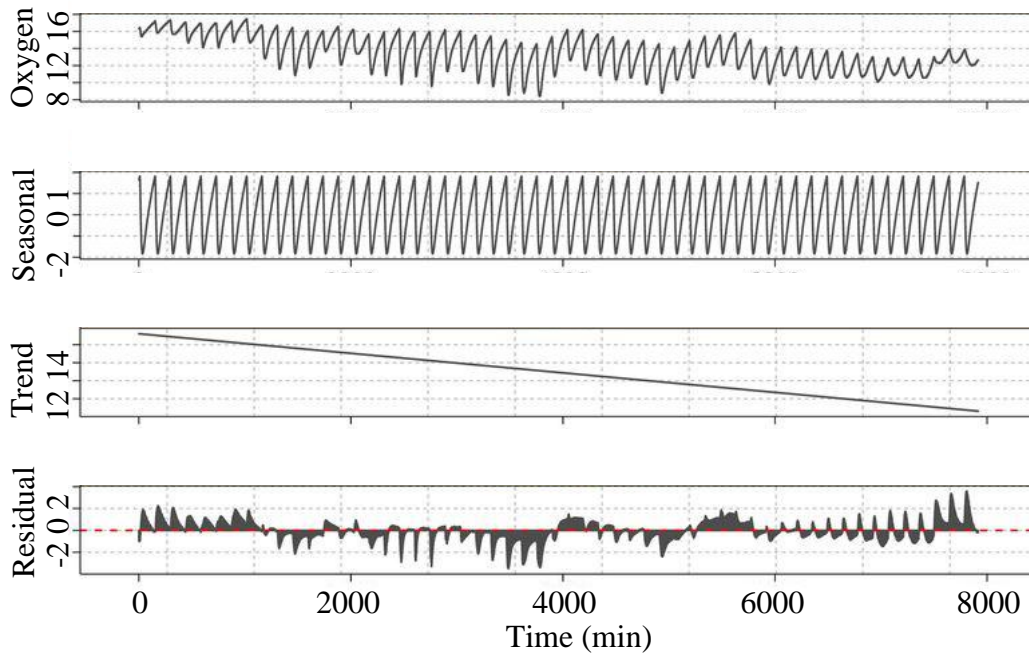


Figure 44: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 24 hour frequency, 10.16 cm soil depth

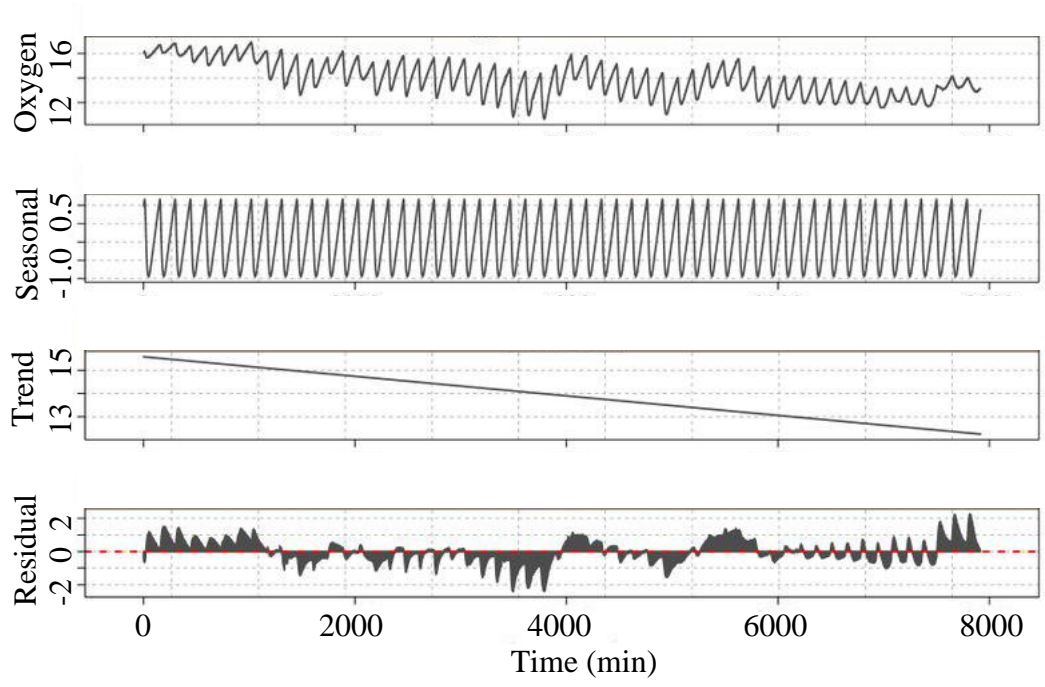


Figure 45: Seasonal decomposition for time series at 112.08 g BOD/m²/day concentration, 24 hour frequency, 30.48 cm soil depth

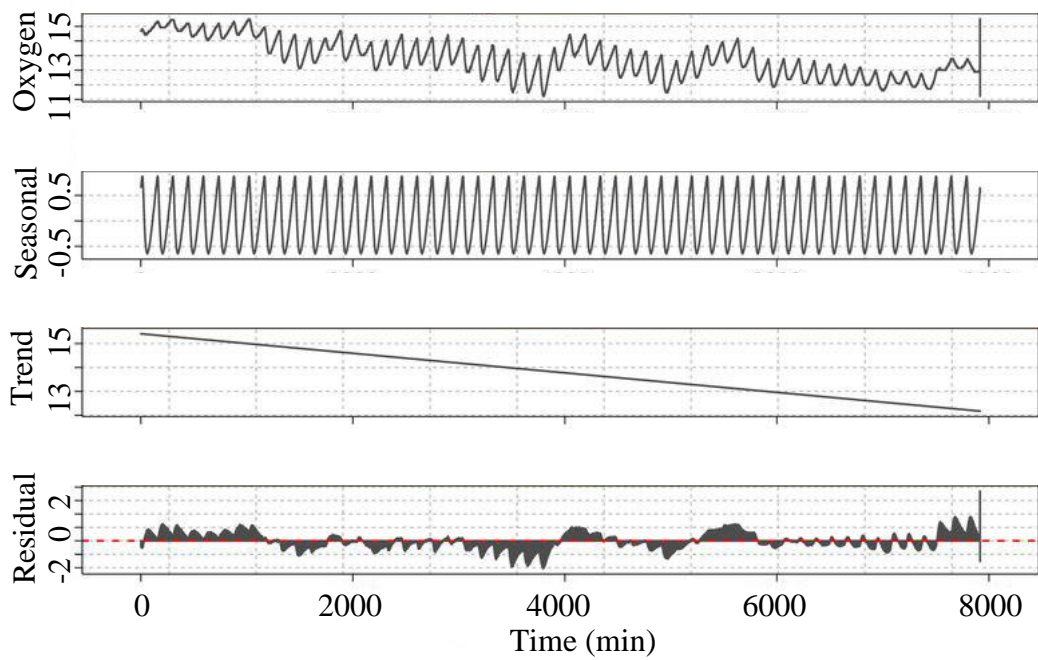


Figure 46: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 24 hour frequency, 50.80 cm soil depth

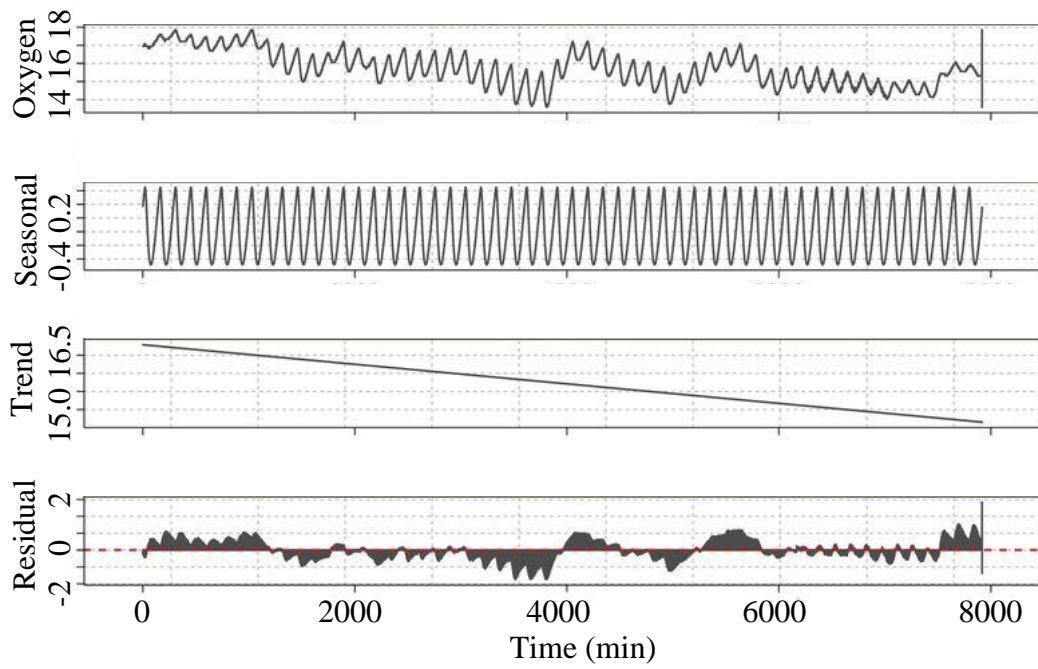


Figure 47: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 24 hour frequency, 71.12 cm soil depth

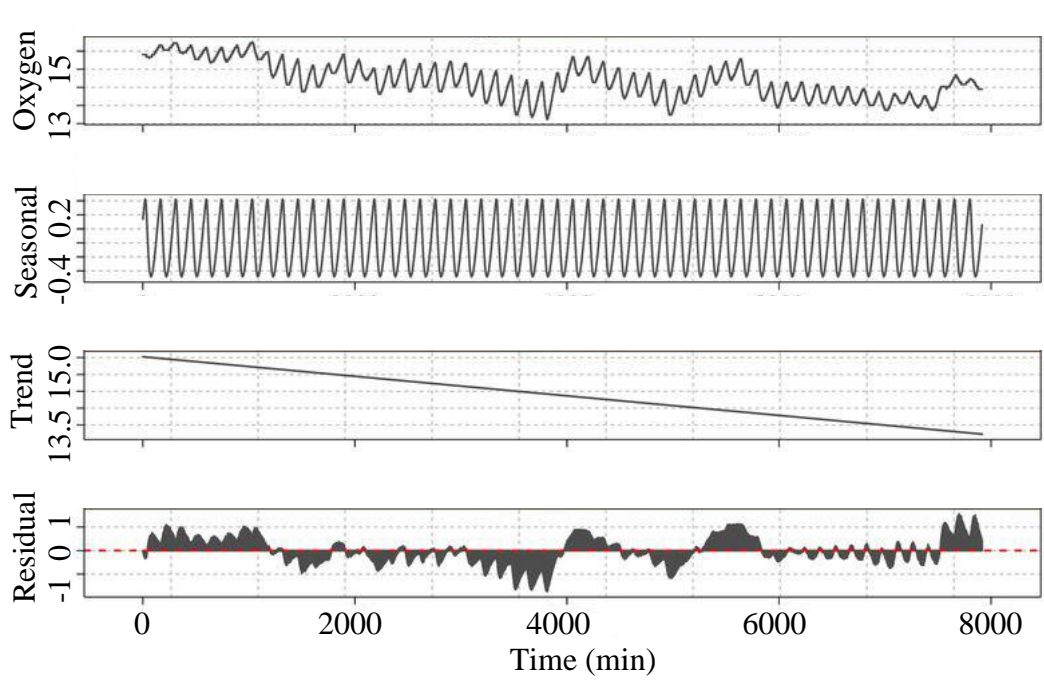


Figure 48: Seasonal decomposition for time series at 112.08 g BOD/m²/day concentration, 24 hour frequency, 91.44 cm soil depth

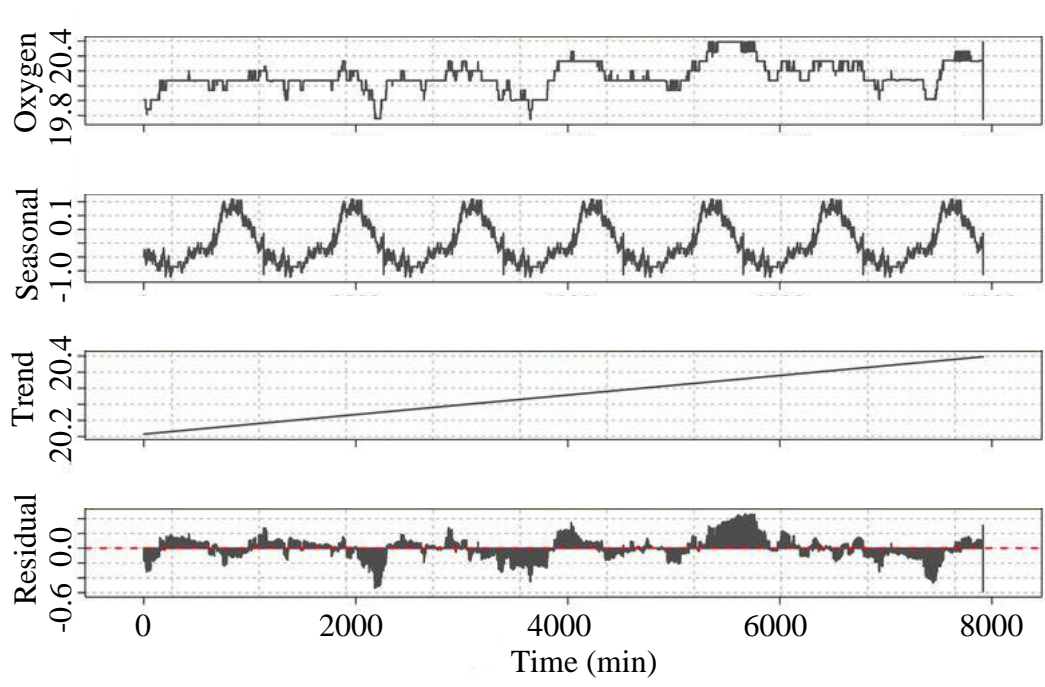


Figure 49: Seasonal decomposition for time series at 112.08 g BOD/m²/day concentration, 56 hour frequency, 10.16 cm soil depth

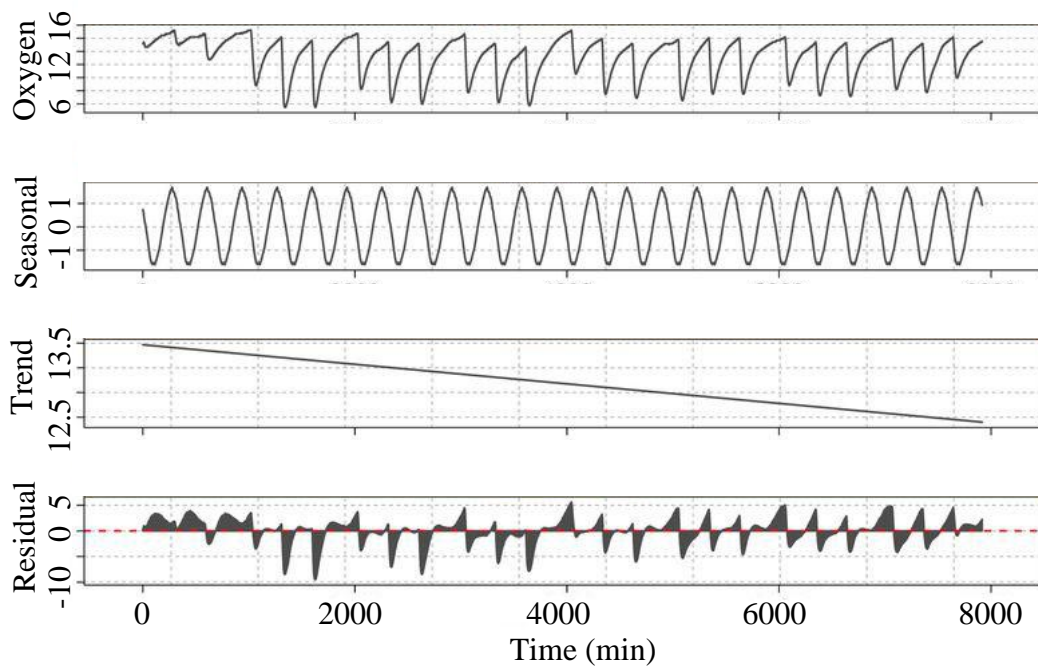


Figure 50: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 56 hour frequency, 30.48 cm soil depth

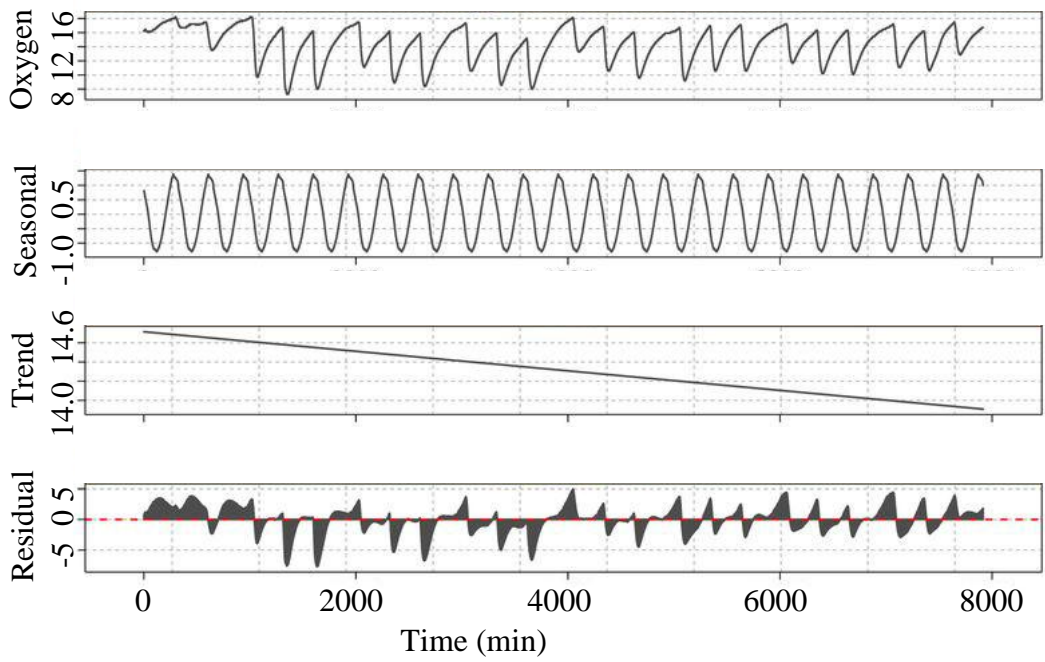


Figure 51: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 56 hour frequency, 50.80 cm soil depth

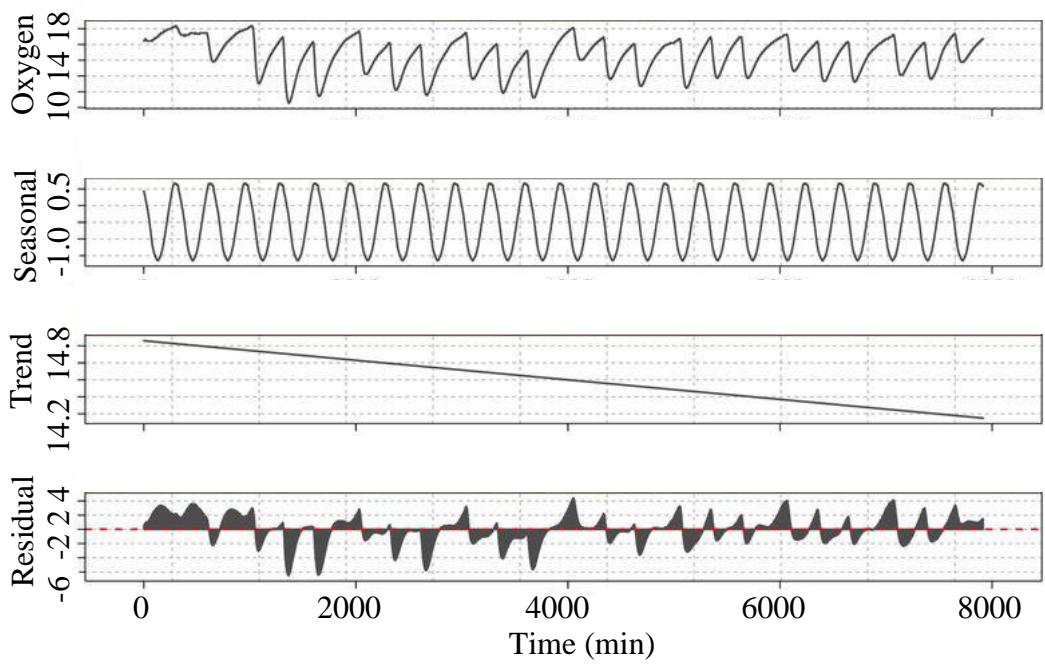


Figure 52: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 56 hour frequency, 71.12 cm soil depth

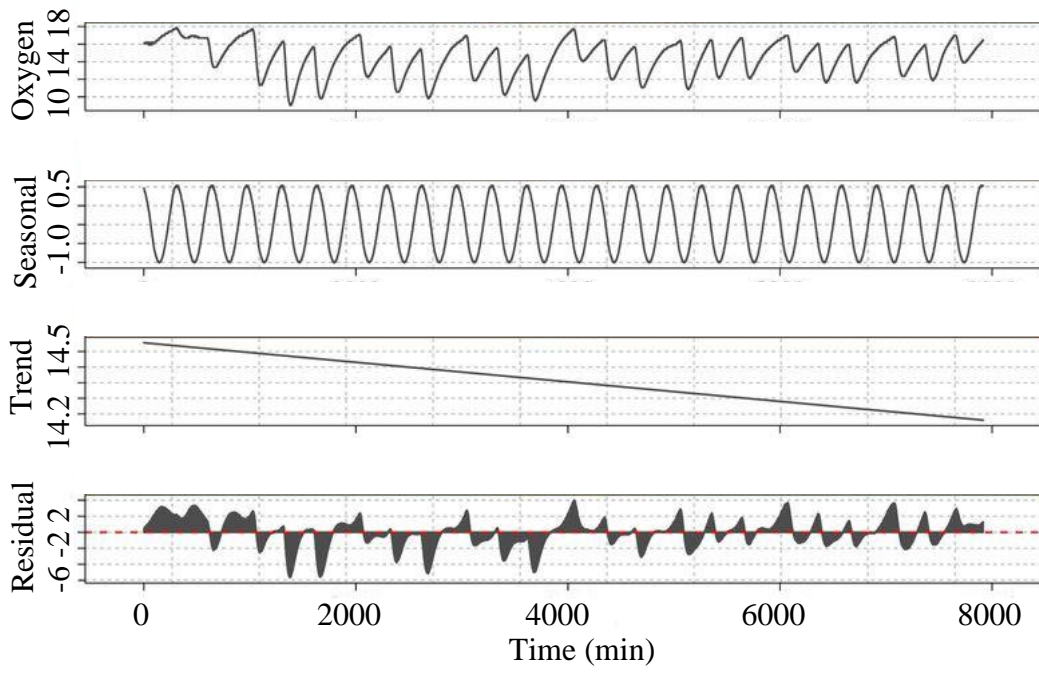


Figure 53: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 56 hour frequency, 91.44 cm soil depth

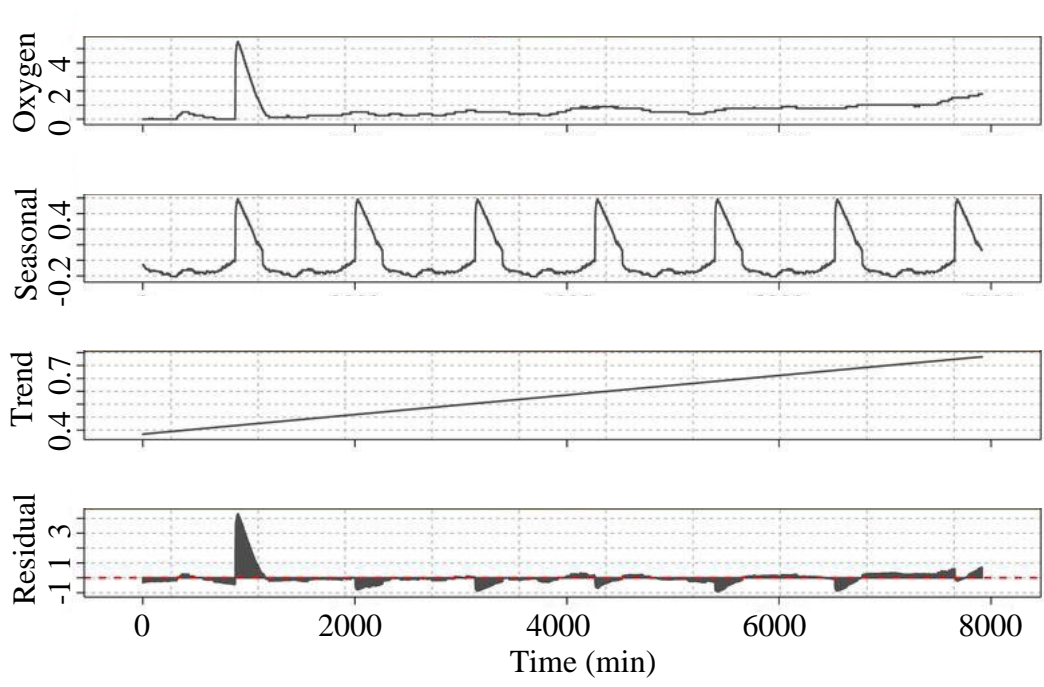


Figure 54: Seasonal decomposition for time series at $112.08 \text{ g BOD/m}^2/\text{day}$ concentration, 56 hour frequency, 111.76 cm soil depth

REFERENCES

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*: 19 (6), 716-723.

Box, GEP. and Wilson, KB. (1951). On the Experimental Attainment of Optimum Conditions (with discussion). *Journal of the Royal Statistical Society Series B*: 13(1), 1-45.

Box, G. and Jenkins, G. (1994). *Time Series Analysis: Forecasting and Control*. 3rd ed. NJ: Prentice-Hall, 1-21.

Broughton, SA. and Bryan, K. (2008). *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*. NY: 72.

Brown and Caldwell. (2007). *Manual of good practice for land application of food processing/rinse water*, California League of Food Processors, Davis, CA.

Charlton SR. and Parkhurst, DL. (2011). Modules based on the geochemical model PHREEQC for use in scripting and programming languages. *Computers and Geosciences*: 37,1653-1663.

Cheung, M. (2008). A Model for Integrating Fixed-, Random-, and Mixed-Effects Meta-Analyses Into Structural Equation Modeling. *Psychological Methods*: 13 (3), 182-202.

Claus, D, Fritze, D. and Kocur, M. (2006). Genera related to the genus *Bacillus*—*Sporolactobacillus*, *Sporosarcina*, *Planococcus*, *Filibacter* and *Caryophanon*. *The Prokaryotes*, 3rd edn. NY: Springer, 4, 631-653.

Cleveland, RB, Cleveland, WS, MaRae, JE. and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of official statistics*: 6(1), 3-73.

Degelmann, DM, Kolb, S, Dumont, M, Colin Murrell, J. and Drake, HL. (2009) *Enterobacteriaceae facilitate the anaerobic degradation of glucose by forest soil*. *FEMS Microbiology Ecology*: 68, 312-319.

Elitzak, H. (2000). Food marketing costs: a 1990s retrospective. *Food Market*: 23(3), 27-30.

Erickson, J and Tyler, EJ. (2001). A Model for Soil Oxygen Delivery to Wastewater Infiltration Surfaces. *On-site Wastewater Treatment*, K. Mancel (ed.) *Proceedings of the Ninth International Symposium on Individual and Small Community Sewage Systems in ASAE*, St. Joseph, MI: 11-17.

Farrar, J, Hawes, M, Jones, D. and Lindow, S. (2003). How roots control the flux of carbon to the rhizosphere. *Ecology*: 84, 827-837.

Fox, J. (2002). Linear Mixed Models. Appendix to An R and S-PLUS Companion to Applied Regression: 1-24.

Gelman, A. (2005). Analysis of variance-why it is more important than ever. *The Annals of Statistics*: 33 (1), 1-53.

Gu, C. and Riley, WJ. (2010). Combined effects of short term rainfall patterns and soil texture on soil nitrogen cycling — A modeling analysis *Journal of Contaminant Hydrology*: 112, 141-154.

Gupta, HV, Sorooshian, S. and Yapo, PO. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.*:4(2), 135-143.

Halim, CE, Short, SA, Scott, JA, Amal, R. and Lowc, G. (2005). Modelling the leaching of Pb, Cd, As, and Cr from cementitious waste using PHREEQC. *Journal of Hazardous Materials*: A125, 45-61.

Hoorman, J, Aziz, I, Reeder, R, Sundermeier, A. and Islam, R. (2011). Soil Terminology and Definitions. *Agriculture and Natural Resources*. OH: Ohio State University Extension.

Howell, DC. (2012). Treatment of Missing Data Part I.
http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Mixed-Models-Repeated/Mixed-Models-for-Repeated-Measures1.html.

Klute, A. and Dirksen, C. (1986). Hydraulic conductivity and diffusivity: laboratory methods. *Amer. Soc. Agron*: 47.

King, EM, Valley, JW, Davis, DW. and Edwards, GR. (1998). Oxygen isotope ratios of Archean plutonic zircons from granite-greenstone belts of the Superior Province: indicator of magmatic source. *Precambrian Res* 92:365-387.

Lee, FY, and Kittrick, JA. (1984). Electron microprobe analysis of elements associated with zinc and copper in an oxidizing and an anaerobic soil environment. *Soil Sci. Soc. Am. J.* 48:548-554.

Lenth, RV. (2009). Response-Surface Methods in R, Using rsm. *Journal of Statistical Software*: 32 (7), 1-17.

Liddle, AR. (2008). Information criteria for astrophysical model selection.
<http://arxiv.org/pdf/astro-ph/0701113v2.pdf>.

Lipson, DA, Zona, D, Raab, TK, Bozzolo, F, Mauritz, M. and Oechel, WC. (2012). Water-table height and microtopography control biogeochemical cycling in an Arctic coastal tundra ecosystem. *Biogeosciences*: 9, 577-591.

Masscheleyn, PH, Delaune, RD, and Patrick, WH. (1991). Effect of Redox Potential and pH on Arsenic Speciation and Solubility in a Contaminated Soil. *Environ. Sci. Technol* 25: 1414-1419.

Moncayo, FHO. (2003). Oxygen Transport in Waterlogged Soils, Part I. Approaches to Modelling Soil and Crop Response to Oxygen Deficiency. Lecture given at the College on Soil Physics in University of Caldas, Manizales, Colombia: LNS0418024.

Moriasi, DN, Arnold, JG, Van Liew, MW, Bingner, RL, Harmel, RD. and Veith, TL. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *American Society of Agricultural and Biological Engineers*: 50(3), 885-900.

Myung, IJ. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*: 47, 90-100.

Nash, JE. and Sutcliffe, JV. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrology*: 10(3), 282-290.

Oh, S. and Logan, BE. (2005). Hydrogen and electricity production from a food processing wastewater using fermentation and microbial fuel cell technologies. *Water Research*: 39, 4673-4682.

Parkhurst, DL. and Appelo, CAJ. (1999). User's guide to PHREEQC (VERSION 2)-a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. *Water-Resources Investigations Report*: 99-4259.

Radcliffe, DE. and Simunek, J. (2010). *Soil Physics and HYDRUS-Modeling and Applications*. CRC Press, 86-87, 122-124, 249-250.

RKB (Rice Knowledge Bank). (2009). Effects of Submergence on Soil Properties, Soil Oxygen Supply. International Rice Research Institute : <http://www.knowledgebank.irri.org/submergedsoils/mod2menu/lesson-1.html>.

Schellenberger, S, Drake, HL. and Kolb, S. (2011). Functionally Redundant Cellobiose-Degrading Soil Bacteria Respond Differentially to Oxygen. *Applied and environmental microbiology*, Sep: 6043-6048.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism and Atmospheric Electricity*: 3, 13-41.

Schwarz, GE. (1978). Estimating the dimension of a model. *Annals of Statistics*: 6 (2), 461-464.

Simunek, J, Sejna, M. and Van Genuchten, M. (1999). The HYDRUS-2D software 500 package for simulating water flow and solute transport in two-dimensional variably 501

saturated media. Version 2.0 IGWMC-TPS-53C. Golden, Colorado: International 502 Groundwater Modeling Center, Colorado School of Mines.

Simunek, J, Sejna, M, Saito, H, Sakai, M. and Van Genuchten. M. (2009). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media version 4.08. CA: University of California Riverside.

Sin, HN, Yusof a, S, Abdul Hamid, NS. and Rahman, RA. (2006). Optimization of hot water extraction for sapodilla juice using response surface methodology. *Journal of Food Engineering*:74, 352-358.

Smirnov, NV and Dunin-Barkovskii, IV. (1969). *Mathematische Statistik in der Technik*, Deutsch. Verlag Wissenschaft: 1-200.

Steel, RGD, and Torrie, JH. (1960). *Principles and Procedures of Statistics with Special Reference to the Biological Sciences*. McGraw Hill: 187-287.

Tiedje, JM, Sexstone, AJ, Parkin, TB, Revsbech, NP, and Shelton, DR. (1984). Anaerobic processes in soil. *Plant and Soil*: 76,197-212.

Trulear, MG and Characklis, WG. (1982). Dynamics of Biofilm Processes. *Water Pollution Control Federation*: 54(9), 1288-1301.

Vaughan, S. and Uttley, P. (2006). Detecting X-ray QPOs in active galaxies. *Advances in Space Research*: 38 (7), 1405-1408.

Wu, M, Li, D, Wang, LJ, Zhou, YG, Brooks, MS, Chen, XD. and Mao, ZH. (2008). *Separation and Purification Technology*: 61, 51-59.

Xu, TF, Sonnenthal, E, Spycher, N. and Pruess, K. (2004). TOUGHREACT user's guide: a simulation program for non-isothermal multiphase reactive geochemical transport in variable saturated geologic media. Berkeley, CA: Lawrence Berkeley National Laboratory Report LBNL-55460, 192.

Xu, TF, Sonnenthal, E, Spycher, N. and Pruess, K. (2006). TOUGHREACT — a simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: applications to geothermal injectivity and CO₂ geological sequestration. *Computers and Geosciences*: 32, 145-165.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*: 92(4), 937-950.

Zar, JH. (1984). *Biostatistical Analysis*. NJ: Prentice Hall International, 43-45.

Zhang, ZS, Li, D, Wang, LJ, Ozkanc, N, Chen, XD, Maa, ZH. and Yang, HZ. (2007). Optimization of ethanol–water extraction of lignans from flaxseed. *Separation and Purification Technology*: 57, 17-24.