DECENTRALIZED WASTEWATER UTILIZATION FOR SUSTAINABLE WATER AND ENERGY MANAGEMENT

By

Benjamin Thomas

A DISSERTATION

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

Biosystems Engineering – Doctor of Philosophy

2024

ABSTRACT

With growing water scarcity as a leading challenge for sustainable development, decentralized wastewater treatment and recycling strategies are emerging as viable solutions to address the water needs of a significant portion of the population. Unlike centralized wastewater treatment facilities, decentralized systems, especially those incorporating the source separation of wastewater, offer cost-effective and efficient ways to treat wastewater.

This study first conducted a comprehensive life cycle impact assessment and techno-economic analysis to compare five treatment scenarios for two types of source-separated wastewater: blackwater (from toilets and kitchens) and greywater (from showers and laundry). These scenarios utilized different combinations of three scalable technologies: activated sludge, anaerobic digestion (AD), and membrane filtration. Activated sludge was employed to treat source-separated wastewater, while anaerobic digestion processes sludge into biogas for energy generation. Membrane filtration, including ultrafiltration and reverse osmosis, further purified the treated wastewater for discharge or recycling. The study revealed that using activated sludge and membrane filtration to treat blackwater and greywater separately, followed by anaerobic digestion to reduce the sludge and generate methane energy, offered superior environmental and techno-economic performance among the evaluated scenarios. The study highlighted the importance of biological treatments in removing pharmaceutical and personal care products (PPCPs) from wastewater, thus reducing their environmental impact.

A baffled bioreactor (BBR) was utilized for blackwater treatment, showing high removal rates of organic content and inorganic nitrogen, which increased with higher feed amounts. The microbial diversity within the BBR system was also greater at higher feed amounts, facilitating the removal of total solids, total nitrogen, and nitrates. An economic analysis examined the

treatment costs under different energy scenarios, including electricity from the grid, propane gas engines for remote communities, and diesel engines for military and extreme environments.

Greywater, which can be separated from blackwater due to its lower contaminant concentration, is an excellent candidate for recycling. To optimize greywater treatment, the study evaluated three ultrafiltration membranes: Pittsburgh Plate Glass (PPG), Polyvinylidene Fluoride (PVDF), and Polyethersulfone (PES), using greywater from showers, laundry, and a combination of both as feed water. The PPG membrane demonstrated the fastest flux and least fouling across all water types, while PVDF and PES were more efficient at nutrient removal. The study concluded that a multiple objective optimization (MOO) approach is effective for selecting membranes and designing treatment processes tailored to different greywater sources.

Addressing the inherent trade-offs in wastewater treatment of balancing water quality, energy consumption, and cost, the study employed a MOO approach to optimize treatment combinations. The system studied included electrocoagulation (EC) for blackwater treatment, AD for food waste and EC sludge, electrodialysis (ED) for final water treatment, and electricity generation from biogas and photovoltaic (PV) solar energy. The combination of PV, AD, EC, and ED achieved the best performance in terms of water quality, meeting EPA discharge standards, and demonstrated a low global warming potential (GWP) and high energy output. The Pareto frontier analysis highlighted AD+EC+ED and PV+AD+EC+ED as the preferred treatment combinations, prioritizing water quality and overall environmental performance. This integrated approach to decentralized wastewater treatment and recycling not only addresses water scarcity but also offers sustainable and economically viable solutions for various applications, from domestic to industrial and agricultural settings.

ACKNOWLEDGEMENTS

The author would like to thank the financial support from the U.S. Department of Defense, this research was supported by the U.S. Department of Defense (W56HZV-17). The author would also like to thank Delhi Township Wastewater Treatment Plant in Holt, Michigan for providing the site for this research, the Research Technology Support Facility (RTSF) of Michigan State University for their support on gene sequencing, and Dr. Jianmin Wang at Missouri University of Science and Technology for his technical support on operation of the baffled bioreactor. Thanks to the Army Public Health Center (APHC) for obtaining and analyzing the greywater and blackwater samples used in this study and Dr. Sibel Uludag-Demirer for the lab support and assistance throughout the program. The Scanning Electron Microscopy (SEM) data was analyzed by the Center for Advanced Microscopy at Michigan State University. The FT-IR analysis was carried out under the guidance of Dr. Mojgan Nejad at the Green Bioproducts Science and Engineering lab at Michigan State University.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS vi

INTRODUCTION 1

CHAPTER 1: LIFE CYCLE IMPACT AND ECONOMIC ASSESSMENT OF DECENTRALIZED STRATEGIES TO TREAT SOURCE-SEPARATED WASTEWATER........ 9

CHAPTER 2: DECENTRALIZED HIGH-STRENGTH WASTEWATER TREATMENT USING A COMPACT AEROBIC BAFFLED BIOREACTOR 40

CHAPTER 3: EVALUATION OF ULTRAFILTRATION MEMBRANE FOULING FOR GREYWATER RECYCLING USING A MULTIPLE-OBJECTIVE OPTIMIZATION APPROACH 79

CHAPTER 4: MULTI-OBJECTIVE OPTIMIZATION OF A MODULAR BASED TREATMENT SYSTEM TOWARDS SUSTAINABLE WASTE AND WASTEWATER MANAGEMENT.. 104

CONCLUSIONS AND FUTURE WORK 133

REFERENCES 139

APPENDIX A: ORIGINAL R SOFTWARE CODE 149

APPENDIX B : SUPPLEMENTAL TABLES AND FIGURES.. 274

LIST OF ABBREVIATIONS

AD	Anaerobic Digestion
ADREC	Anaerobic Digestion Research and Education Center
AFB	Aerobic Fluidized Bed
AnMBR	Anaerobic membrane bioreactor
ANOVA	Analysis of Variance
APHC	Army Public Health Center
BBR	Baffled bioreactor
BOD	Biochemical Oxygen Demand
C	Carbon
CapEx	Capital Expenditure
CdTe	Cadmium telluride
CHP	Combined Heat and Power
CH_4	Methane
CO_2	Carbon dioxide
COD	Chemical Oxygen Demand
CRD	Completely Randomized Experimental Design
CSTR	Continuously Stirred Tank Reactor
DEET	N, N-Diethyl-meta-toluamide
DoD	Department of Defense
EC	Electrocoagulation
EDR	Electrodialysis Reversal
EPA	Environmental Protection Agency

FFB	Fixed film bioreactors
FTIR	Fourier-transform infrared spectroscopy
FOG	Fats/Oils/Greases
GWP	Global warming potential
GPM	Gallons per minute
GVSC	Ground Vehicle Systems Center
HDPE	High Density Polyethylene
HRT	Hydraulic retention time
IC	Internal Combustion
ICD	Initial Capabilities Document
I/O	Inputs/Outputs
IIET	Indirect interspecies electron transfer
kWh	Kilowatt-hour
kV	Kilovolt
L	Liter
LCIA	Life Cycle Impact Assessment
LPD	Liters per day
MACRS	Modified Accelerated Cost Recovery System
MBR	Membrane bioreactor
MLSS	Mixed Liquor Suspended Solids
MOO	Multiple objective optimization
MSU	Michigan State University
N	Nitrogen

N_2O	Nitrous Oxide
NH_3	Ammonia
NMDS	Non-metric multidimensional scaling
NO_3	Nitrate
NO_2	Nitrite
NPDES	National Pollution Discharge Elimination System
NSF	National Sanitation Foundation
NTU	Nephelometric Turbidity Unit
OpEx	Operational Expenditure
O_3	Ozone
P	Phosphorus
PCR	Polymerase chain reaction
PES	Polyethersulfone
PI	Principal Investigator
P&ID	Process and Instrument Drawing
PLC	Programmable Logic Controller
PPCP	Pharmaceuticals and Personal Care Products
PPE	Personal Protection Equipment
PPG	Pittsburgh Plate Glass
PSI	Pounds per square inch
PV	Photovoltaic
PVDF	Polyvinylidene Fluoride
RBC	Rotating biological contactors

RO	Reverse Osmosis
rRNA	Ribosomal RNA
RTSF	Research Technology Support Facility
SEM	Scanning electron microscope
TDS	Total Dissolved Solids
TEA	Techno-economic Analysis
TEG	Thermoelectric Generator
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
TOC	Total Organic Carbon
TP	Total Phosphorus
TRACI	Tool for reduction and assessment of chemicals and other impacts
TS	Total Solids
TSS	Total Suspended Solids
UASB	Upflow Anaerobic Sludge Blanket
UF	Ultrafiltration
VS	Volatile Solids
W	Watts
WEP	Water eutrophication potential

INTRODUCTION

1. Problem background

Untreated wastewater being discharged into the environment is a global problem that has a direct correlation to growing water scarcity and accessibility issues. According to the 2017 United Nations World Water Development Report, over 80% of global wastewater is discharged into the environment without any treatment, creating a public health and environmental liability. High-income countries are able to treat around 70% of their municipal and industrial wastewaters, while low-income countries are only able to treat 8% of their wastewater [1]. Untreated wastewater can pollute freshwater resources that are a valuable and diminishing source of potable water for many communities. According to Avalon Global Research, the majority of the pollution contaminating clean water resources is from untreated city sewage and industrial waste discharged into rivers [2].

Decentralized wastewater treatment is an option to treat currently untreated wastewater and it is also a solution for existing wastewater treatment infrastructures. Decentralized wastewater treatment technologies can be a potential solution to the costly burden for refurbishing or upgrading systems facing a large percentage of the wastewater infrastructure. It is estimated that in centralized wastewater management, 80-90% of the total cost is attributed to the transportation of wastewater, with only 10-20% attributed to the treatment process [3]. The increasing demand from small rural/suburban communities and military bases requires a decentralized solution tailored to treat source-separated wastewaters.

The lower flows that are seen in small-scale communities allow for a wider range of technical options. Methods that may not be feasible to use in centralized systems have the potential to be utilized in decentralized operations. Some methods that can be investigated for

decentralized systems are separated into the following categories: physical separation, biological, electrochemical, membrane filtration, and energy co-generation. Decentralized systems also allow for easier source-separation of the wastewaters. Greywater (shower and laundry wastewater) can be separated from blackwater (kitchen and latrine wastewater), which allows for the unique utilization of each water stream. Greywater is a great candidate for recycling due to its low contaminant concentration, and blackwater can be utilized for energy generation. The activated sludge process is the conventional approach to wastewater treatment with its widespread usage for the biological treatment of municipal and industrial wastewaters. Predecessors to the modern activated sludge process date back to the 1880s in England [4]. Anaerobic digestion (AD) allows for the inherent energy in wastewater to be utilized with an energy generating component. Including an energy generation process can determine if waste utilization is feasible in a decentralized scenario. In order to achieve a water quality that can be utilized for recycling and other potable purposes, membrane treatment needs to be adopted. The membrane treatment serves as a selective barrier that can filter out a range of contaminants including particles and dissolved constituents [4]. Membrane treatment is a great option for decentralized treatment due to its scalability and ability to operate at smaller scales.

The proposed project will research and develop strategies for utilizing and treating source-separated wastewaters in decentralized scenarios, thereby removing the environmental liability of wastewaters and turning them into valuable assets.

2. Literature review

2.1. The conventional centralized wastewater treatment approach

The Environmental Protection Agency (EPA) estimates that $271 billion will be required for the wastewater infrastructure over the next 25 years [5]. This massive cost burden is required

to replace and repair old and failing infrastructure, and it is estimated that 95% of the money spent for water infrastructure is paid for at the local level [6]. Decentralized wastewater treatment can be a potential solution to reduce the costly burden facing a large percentage of the wastewater infrastructure by serving rural and distributed regions or reducing the growing burden on existing infrastructure. It is estimated that in centralized wastewater management, 80-90% of the total cost is attributed to the transportation of wastewater, with only 10-20% attributed to the treatment process [3]. The current centralized municipal wastewater system and corresponding treatment technologies have been intensively investigated in the past decades [7]. However, decentralized, less typical wastewater treatment operations (rural and suburban communities, small industrial/agricultural operations, and military bases) have not been investigated as deeply as municipal wastewater treatment plants and are therefore not as well understood and conventionalized. The wastewater produced from small-scale operations often has a much higher pollution concentration than typical municipal wastewaters due to the mixing of some concentrated waste streams (e.g., food waste, latrine waste) with less dilution [8].

Activated sludge processes as a biological treatment system are widely used to treat wastewater [7]. They are highly effective at removing organic matter, suspended solids, and nutrients from wastewater due to the synergy of a variety of aerobic microorganisms in the activated sludge. The major groups of microorganisms found in activated sludge are bacteria, protozoa, metazoa, filamentous bacteria, and algae/fungi. Among them, bacteria are the largest group comprising approximately 95% of the total microorganisms in activated sludge [9]. They are the primary microbes in charge of metabolizing a wide range of organic compounds as well as removing inorganic nitrogen and phosphorus. The key physiological groups of bacteria in activated sludge include: chemoorganohetorotrophs (e.g., Proteobacteria and Desulfovibrio) that

use fermentation and respiration to degrade and utilize organic compounds in wastewater, chemolithoautotrophs (e.g., Candidatus, Nitrosomonas, Nitrobacter, and Ferroplasma) that oxidize a range of inorganic compounds to obtain energy, and photoorganoheterotrophs and photolithoautotrophs that use light as an energy source but utilize organic and inorganic carbon and nutrient sources, respectively [10]. Several variables influence the effectiveness of the activated sludge process, including the concentration of organic matter in the wastewater, the concentration and type of microorganisms in the activated sludge, the aeration rate, and the hydraulic retention time. Activated sludge processes have flexibility to treat a wide range of wastewater streams and produce a high-quality effluent that can be discharged into the environment or reused for irrigation or other purposes. Activated sludge processes have advantages including high treatment efficiency, modular design, and relatively low energy demand. Scaling them down and using them for decentralized wastewater treatment presents challenges of operational instability (flow or composition changes and environmental conditions), microbial health, sludge management, etc.

2.2. Source separation of wastewaters: greywater and blackwater

Decentralized wastewater and water management allows for easier separation of wastewaters, giving more options for wastewater treatment to reduce energy costs and allow for water recycling. Greywater and blackwater are the two main sources of municipal wastewater, which can be separated at the source for further treatment.

Greywater refers to wastewater that is generated from sinks, showers, and laundries [11]. It contains soaps, detergents, and other household cleaning products, but does not contain fecal matter. Greywater has a lower contaminant concentration and is a prime candidate for recycling as it is easier to treat, and accounts for a large percentage (approximately 75%) of the total

wastewater produced from a household [12]. Therefore, reusing the greywater can reduce the potable water burden of a community by a large amount. If a water reuse system has a recovery (% of water treated for potable use) of 75%, then the total potable water demand can be reduced by around 56%. This can have a major impact on communities that experience water scarcity and communities that have high costs for potable water.

Blackwater is generated from toilets and kitchens, containing fecal matter and urine. It typically has elevated concentrations of biochemical oxygen demand (BOD) (2,000 mg/L), chemical oxygen demand (COD) (3,000 mg/L), total suspended solids (TSS) (1,000 mg/L), and ammonia (300 mg/L) [8]. Since blackwater is a highly contaminated wastewater, it requires specialized treatment to ensure that it is safe for disposal or reuse. Proper management and treatment of blackwater are important to protect public health and the environment. Current treatment options include sewer-based systems, septic tanks, constructive wetlands, sand filters, membrane filtration, and electrochemical treatment. Meanwhile, due to its high carbon and nitrogen contents, blackwater is also a great candidate for energy-generating technologies.

2.3. Decentralized wastewater treatment for small and remote communities

Decentralized wastewater treatment systems are a viable and preferable option for small and remote communities. Decentralized systems can provide cost-effective treatment of wastewater while also providing other benefits, such as increased water conservation, reduced energy consumption, and increased local control over wastewater management. One of the main advantages of decentralized systems is that they can be technically and economically tailored to meet the specific requirements of the community, such as the size and growth rate of the community, the available land and water resources, and the end use of the treated wastewater. Besides providing custom-designed treatment of wastewater, decentralized systems can help

small communities recycle the treated water locally and conserve water resources. Decentralized systems can also be more resilient to disruptions (power outages or natural disasters) than centralized treatment facilities since they are often designed to operate independently of external power sources and can continue to provide treatment even in the event of a loss of grid power. In addition, an emerging circular economy approach of wastes/wastewater management has gained traction in recent years [13]. Decentralized wastewater treatment fits into the concept of a circular economy. The treated water can be recycled locally for non-potable uses, and the nutrient-rich sludge can be used as a fertilizer in nearby farms or gardens. Such an approach will not only benefit the environment but also create jobs and help the local economy.

2.4. Multi-objective optimization to select and configure preferred decentralized wastewater treatment system

During waste and wastewater treatment, key factors such as water quality, energy consumption, and treatment cost are often conflicted with each other. For example, high water quality typically demands more energy and requires more sophisticated and expensive equipment to achieve it. To optimize such a multiple objective system, trade-off(s) between these conflicting factors need to be considered. Multi-objective optimization (MOO) is a tool to consider the trade-off(s) and develop solutions. Through the synthesis of diverse objectives such as cost-effectiveness, energy efficiency, pollutant removal efficiency, and environmental sustainability, MOO facilitates the design, operation, and management of wastewater treatment systems tailored to the specific needs and constraints of decentralized settings. This approach enables decision-makers to explore trade-offs and identify Pareto-optimal solutions that balance conflicting objectives, thus maximizing overall system performance while minimizing environmental impact and resource consumption. By integrating advanced optimization algorithms, lifecycle cost

analysis, and stakeholder engagement processes, recent research endeavors have yielded significant insights into optimizing wastewater treatment for enhanced resilience, resource recovery, and water quality improvement. Therefore, a multi-objective optimization (MOO) approach was adopted in this study to carry out the optimization and selection of suitable treatment combinations.

3. Goal, scope, and objectives

The overall goal of the proposed study is to research and develop scalable systems to utilize wastewaters for decentralized communities. Going beyond just treating the wastewaters to remove them as an environmental hazard, this study focuses on utilizing the wastewaters as resources to reduce energy consumption and create a more sustainable method for wastewater treatment. The scope of this study is shown in Figure 1. A decentralized community is able to source-separate their wastewaters into two streams: greywater and blackwater. Once the wastewaters are separated, they can each be treated with different treatment technologies that are tailored to the water quality parameters of each water source. Greywater is recycled at a 75% recovery rate, with the 25% concentrate waste stream being sent to the blackwater treatment system. The recycled water is returned to the decentralized community for utilization thereby reducing the water supply requirements of the community. The blackwater is treated in order to discharge safely into the environment or can be returned to the input of the greywater recycling system. Blackwater utilization will include an energy generation component, and the energy generated from the blackwater can be utilized on-site at the community for energy demand needs. The treated wastewater can be discharged into the environment and satisfy NPDES discharge requirements.

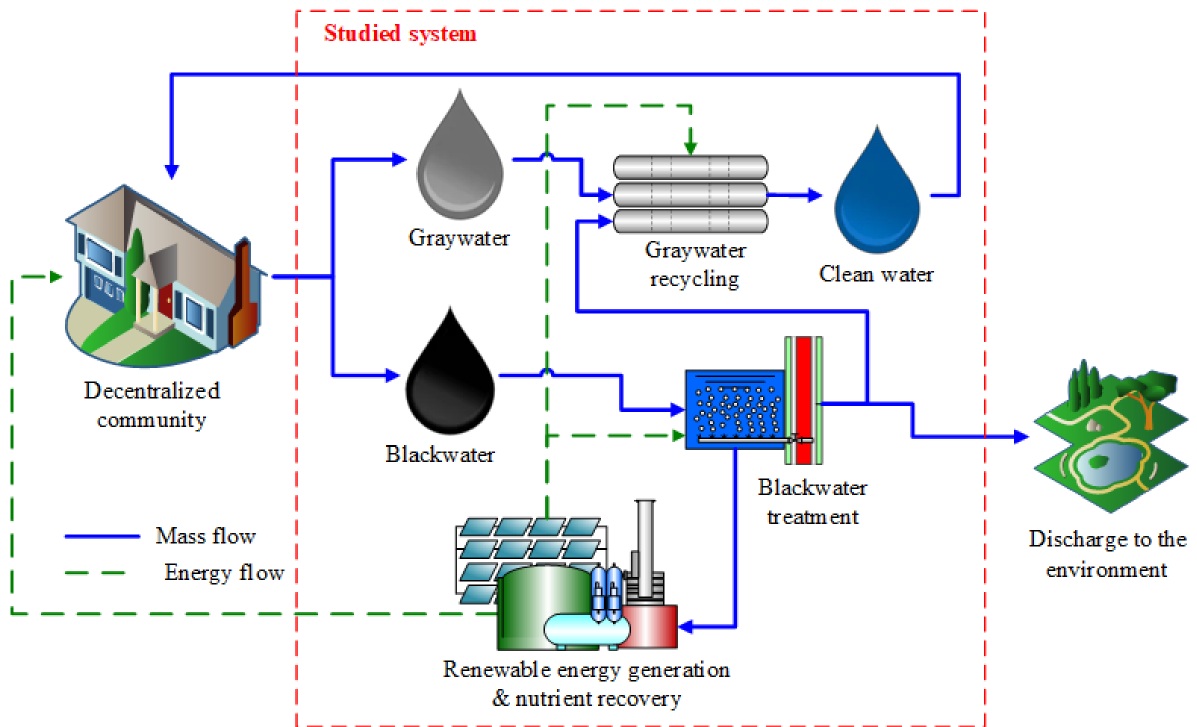


Figure 1. Decentralized wastewater utilization flow chart.

The specific objectives of the proposed study are: 1) Conduct a life cycle and economic assessment on the source-separation of wastewaters in a decentralized scenario, 2) Analyze the treatment capabilities of a small-scale baffled bioreactor for the treatment of blackwater, 3) Characterize the fouling characteristics on ultrafilters from the direct recycling of greywater, and 4) Conduct a multi-objective optimization to develop technically sound, environmentally friendly, and economically feasible decentralized wastewater treatment systems in remote environments.

CHAPTER 1: LIFE CYCLE IMPACT AND ECONOMIC ASSESSMENT OF DECENTRALIZED STRATEGIES TO TREAT SOURCE-SEPARATED WASTEWATER

1. Introduction

Untreated wastewater being discharged into the environment is a global problem that has a direct correlation to growing water scarcity and accessibility issues. According to the 2017 United Nations World Water Development Report, over 80% of the global wastewater is discharged into the environment without any treatment, creating a public health and environmental liability. High-income countries are able to treat around 70% of their municipal and industrial wastewaters, while low-income countries are only able to treat 8% of their wastewater [1]. Untreated wastewater can pollute freshwater resources that are a valuable and diminishing source of potable water for many communities. According to Avalon Global Research, the majority of the pollution contaminating clean water resources is from untreated city sewage and industrial waste discharged into rivers [2].

Decentralized wastewater treatment is an option to treat currently untreated wastewaters and it is also a solution for existing wastewater treatment infrastructures. Decentralized wastewater treatment technologies can be a potential solution to the costly burden for refurbishing or upgrading systems facing a large percentage of the wastewater infrastructure. It is estimated that in centralized wastewater management, 80-90% of the total cost is attributed to the transportation of wastewater, with only 10-20% attributed to the treatment process [3]. Increased average transportation distance for small rural/suburban communities and military bases may exacerbate costs compared to urban areas. Using source-separated wastewater management for these areas as part of a decentralized solution would improve transportation costs.

The lower flows that are seen in small-scale communities allow for a wider range of technical options. Methods that may not be feasible to use in centralized systems have the potential to be utilized in decentralized operations. Some methods that can be investigated for decentralized systems are separated into the following categories: physical separation, biological, electrochemical, membrane filtration, and energy co-generation [4]. An activated sludge process was selected to be analyzed in this study due to its widespread usage for the biological treatment of municipal and industrial wastewaters. Activated sludge has been practiced for over a century, with predecessors to the modern activated sludge process dating back to the 1880s in England [4]. Anaerobic digestion (AD) was selected for this study as an energy generating component to be utilized on sludge wasted from the activated sludge process. Including an energy generating treatment process will allow this study to determine if waste utilization is feasible in a decentralized scenario. In order to achieve water quality that can be utilized for recycling and other potable purposes, membrane treatment was selected. The membrane treatment serves as a selective barrier that can filter out a range of contaminants including particles and dissolved constituents [4]. Membrane treatment is a great option for decentralized treatment due to its scalability and ability to operate at smaller scales.

Decentralized water treatment is a potential solution to address some of the arising water scarcity issues. Decentralized wastewater and water management allows for easier separation of wastewaters, giving more options for the treatment of the wastewater to reduce energy costs and allow for water reuse. For example, greywater (wastewater without any contribution from latrine water [11] can be separated and sent to a different treatment system than blackwater (latrine wastewater). Greywater has a lower contaminant concentration and is a prime candidate for recycling as it is easier to treat, and accounts for a large percentage (approximately 75%) of the

total wastewater produced from a household [12]. Therefore, reusing the greywater can reduce the potable water burden of a community by a large amount. If a water reuse system has a recovery (% of water treated for potable use) of 75%, then the total potable water demand can be reduced by around 56%. This can have a major impact on communities that experience water scarcity and communities that have high costs for potable water.

To comprehensively understand the environmental performance of decentralized source-separated wastewater treatment systems, life cycle assessment (LCA) has been applied because of its unique capabilities of providing holistic view of the technologies, identifying critical points for improvement, enabling technology comparison, assessing resource consumption and emissions, supporting environmental policies, etc. [14,15]. Kobayashi et al. studied LCA of decentralized greywater treatment systems in cold regions and concluded that system scale, wastewater quantity, and mix of power technologies are the key factors to determine environmental performance of the treatment systems [16]. LCA has also been used to compare environmental performance of decentralized wastewater treatment systems with centralized ones [17]. Sharvini et al. investigated environmental impacts of three technologies of extended aeration, Imhoff, and activated sludge on decentralized sewage treatment [18]. However, there are no comprehensive LCAs to date on integrated treatment systems of source-separated wastewater – greywater and blackwater.

Therefore, this study focuses on investigating combinations of three currently available technologies: activated sludge, anaerobic digestion, and membrane filtration to treat source-separated wastewater (greywater and blackwater). Detailed techno-economic analysis and life cycle impact assessment were conducted on five different treatment scenarios to conclude the

most environmentally friendly and cost-effective decentralized wastewater treatment operation and process configuration.

2. Materials and methods

2.1. Source-separated wastewaters and their characterization

The source separated wastewater data used for this study were obtained from a military basecamp located in the United States. A military base camp is a good representative of source separation of wastewaters for decentralized treatment. The basecamp had separate shower, laundry, latrine, and kitchen wastewater collection systems. The greywater sample was the combined shower and laundry water taken from a tank that the shower and laundry waters get pumped into. The blackwater sampling point was from a tank that kitchen and latrine wastewater was pumped into. The kitchen wastewater includes water from food preparation including garbage disposal wastewater. Both kitchen and latrine wastewater on an expeditionary base are relatively concentrated because usage of fresh water is minimized, therefore less dilution occurs. Samples were collected using pre-preserved bottles, placed into coolers with ice directly after collection, and delivered overnight to the laboratory performing the analyses.

All parameters used for the characterization of wastewater were completed immediately after their transfer to the laboratory. Total solids (TS) and total suspended solids (TSS) concentrations were measured using the standard gravimetric method (Method 2540 B &D) from Standard Methods for the Examination of Water and Wastewater [19]. Turbidity was measured using the nephelometric method (Method 2130) [19] with a portable turbidimeter (HACH, 2100Q). The concentration of chemical oxygen demand (COD) and total organic carbon (TOC) was analyzed using a wet oxidation-colorimetric method based on standard Methods 5520-D and 5310 respectively [19] and kits (HACH) were used for the measurement. All nutrients (TN,

TKN, TP, NH_3-N, NO_3-N, NO_2-N) were measured using colorimetric methods using HACH kits prepared based on Standard Methods for the Examination of Water and Wastewater analyses [19]. Five-day BOD tests were carried out based on a respirometric technique using BOD TrakII Respirometric BOD apparatus (HACH) using a fresh seed capsule (HACH) for every measurement. Total coliforms and E-coli were detected using a membrane filter technique (Method 9222) [19] in a biosafety cabinet with laminar flow. All wet oxidation reactions were carried out in a digester (HACH DRB200) and colorimetric measurements were fulfilled by a spectrophotometer (HACH DR3900). Pharmaceuticals & Personal Care Products (PPCPs) analyses were conducted by a contract laboratory using the methods listed in Table S1.

2.2. Treatment scenarios

Three commercial wastewater treatment technologies of activated sludge, anaerobic digestion, and membrane filtration were selected to form five different treatment scenarios for this study. A containerized baffled bioreactor (BBR) from a previous study was used as the base for the decentralized activated sludge treatment [20]. The removal of TSS, TN, TP, COD, and BOD during the activated sludge treatment are 96, 91, 94, 94, and 90%, respectively, based on our previous study [20]. The activated sludge production was calculated based on the characteristics (BOD, COD, TSS, and TN) of wastewater using the calculation of a complete-mix activated sludge process for BOD removal with nitrification [4]. The pharmaceuticals and personal care products (PPCPs) are either degraded or removed by the sludge during the activated sludge treatment. According to the references, the degradation of caffeine, methylphenol, permethrin, phenol, salicylic acid, nicotine, DEET, benzyl alcohol, ibuprofen, chloroform, and acetone during the activated sludge treatment are 100 [21], 100 [22], 90 [23], 100 [24], 30 [25], 100 [26], 70 [27], 95 [28], 100 [21], 80 [29], and 100% [30], respectively. The

removal of di-2-ethylhexyl-phthalate during the activated sludge treatment is 94% [31]. A continuous stirred tank reactor (CSTR) with a combined heat and power unit of biogas utilization was used as the anaerobic digestion unit to convert discharged activated sludge into renewable energy. The removal of TSS, TN, TP, COD, and BOD during the anaerobic digestion process are 60, 50, 50, 70, and 70%, respectively according to the data collected from previous studies (unpublished). The degradation of caffeine, methylphenol, permethrin, phenol, di-2-ethylhexyl-phthalate, salicylic acid, nicotine, DEET, benzyl alcohol, ibuprofen, and acetone during the digestion based on literature results are 87.5 [32], 90 [33], 92 [34], 92 [33], 50 [31], 95 [25], 75 [35], 0 [36], 100 [37], 41 [38], and 97% [39], respectively. It has also been reported that 32% of chloroform was evaporated during anaerobic digestion [40]. The membrane filtration operation includes both ultrafiltration (UF) and reverse osmosis (RO). A spiral wound PPG ULA UF membrane and a DOW FILMTEC SW30 RO membrane were selected for the UF and RO units, respectively. The water recovery for both ultrafiltration and RO units is 85%. The removal of TSS, TN, TP, COD, and BOD in the UF permeate from the UF membrane were 100 [41], 67 [42], 30 [43], 50 [44], and 50%, respectively. The removal of TSS, TN, TP, COD, and BOD in the RO permeate from the RO membrane were 100, 92, 98, 98, and 100%, respectively [45]. The removal of permethrin, di-2-ethylhexyl-phthalate, salicylic acid, DEET, benzyl alcohol, and chloroform during the combination of UF and RO treatment were 90, 90 [46], 97 [47], 92 [48], 90, and 90% [49], respectively.

Table 1 provides information on each treatment scenario and Figure 2 shows the flow path for each treatment scenario. Treatments A and B utilize anaerobic digestion for energy generation from activated sludge Treatment A combines greywater and blackwater for input into the activated sludge process followed by the UF/RO process to treat the activated sludge effluent,

while treatment B separates greywater and then combines it with the activated sludge effluent prior to the UF/RO membrane process. Treatment C mimics treatment A without anaerobic digestion. Treatment D mimics treatment B without anaerobic digestion. Treatment E is a control scenario that has both greywater and blackwater being treated by an activated sludge process without anaerobic digestion for energy generation, or UF/RO for recycling water. Treatment F is another control scenario that has blackwater being treated by activated sludge and discharged without UF/RO treatment while the greywater is treated by UF/RO for recycling.

Table 1. Description of treatment scenarios in this study.

Treatment name	Treatment description
A	Combined greywater and blackwater recycling with membrane filtration, activated sludge, and anaerobic digestion
B	Source separated greywater recycling with membrane filtration for both, but activated sludge and anaerobic digestion for blackwater only
C	Combined greywater and blackwater recycling with membrane filtration and activated sludge, but without anaerobic digestion
D	Source separated greywater recycling with membrane filtration for both and activated sludge for blackwater only, but without anaerobic digestion
E	Control treatment with recycled greywater and discharged treated blackwater

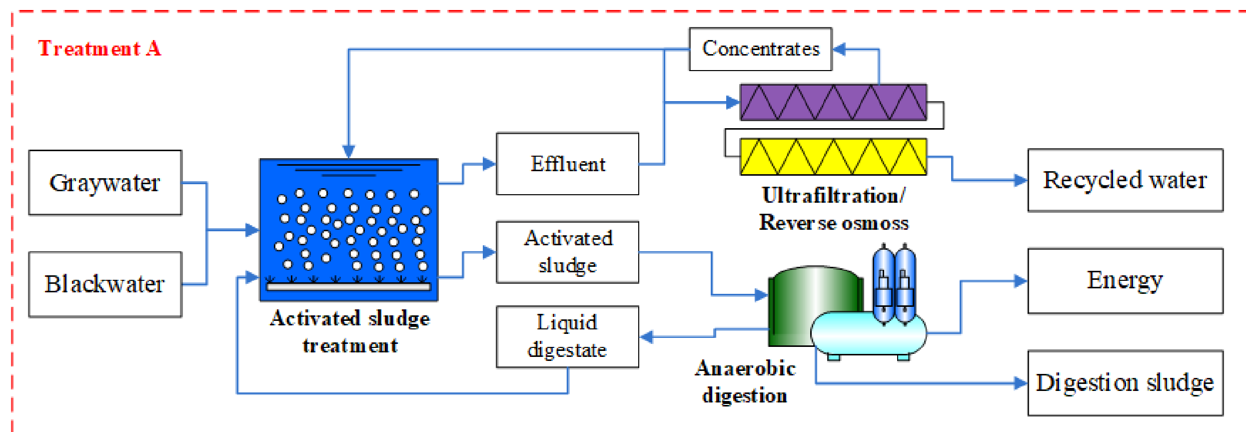


Figure 2. Five treatment scenarios for the treatment of greywater and blackwater were analyzed in this study.

Figure 2 (cont'd)

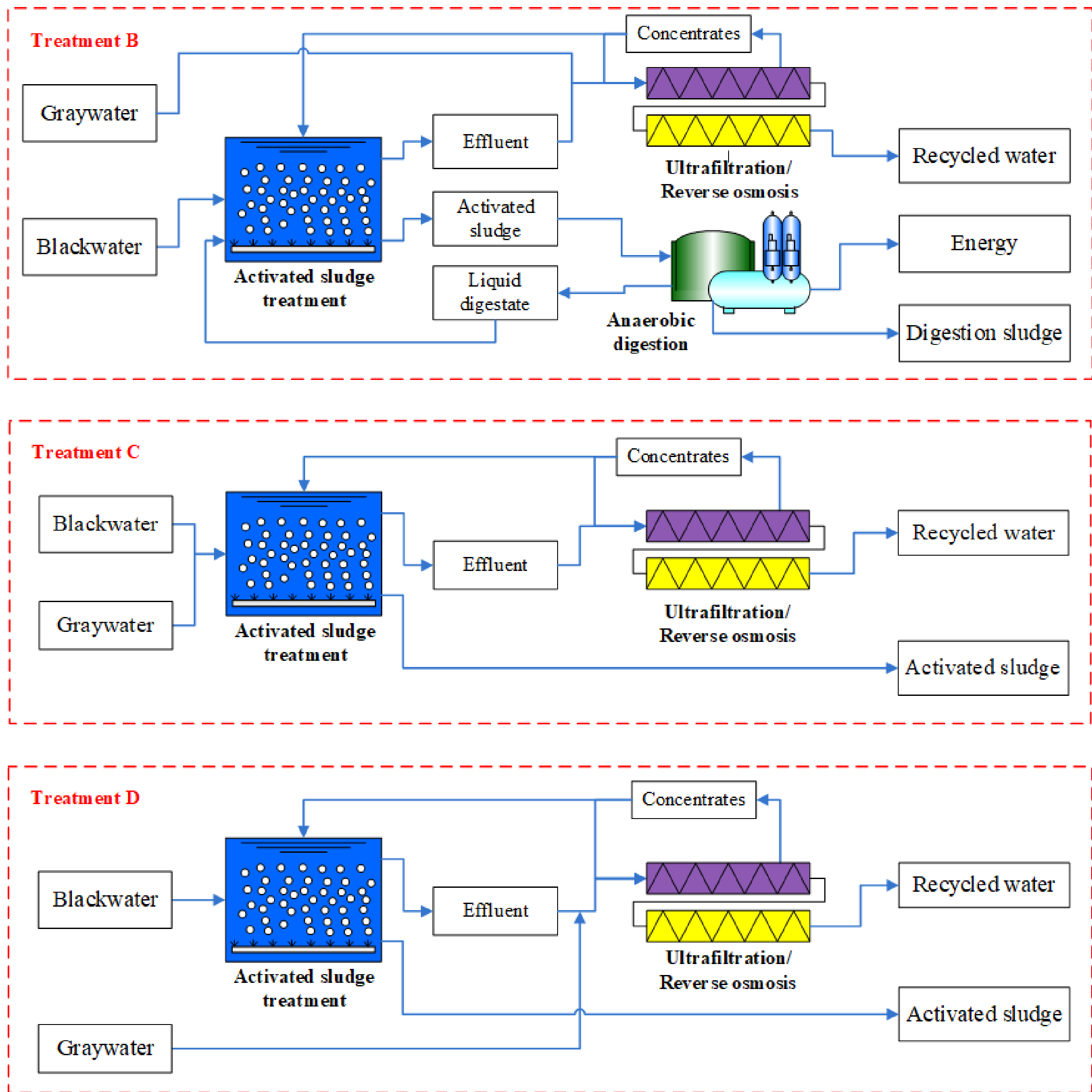
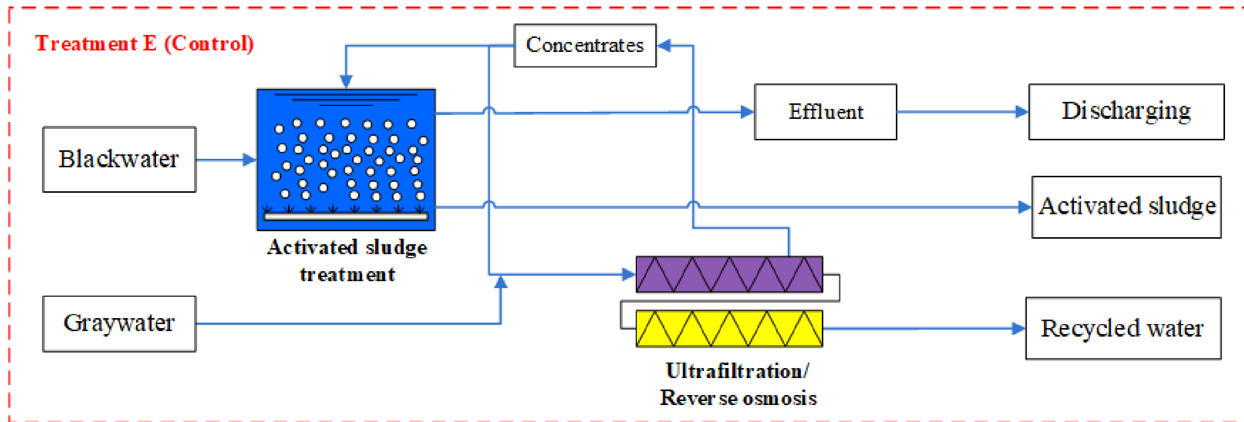


Figure 2 (cont'd)



2.3.Mass and energy balance analysis

Mass and energy balance analyses were carried out for the different treatment scenarios. The mass balance includes the following flows in (Liters/day): greywater influent, blackwater influent, activated sludge influent, activated sludge effluent, ultrafiltration permeate, reverse osmosis permeate, anaerobic digestion influent, ultrafiltration concentrate, reverse osmosis concentrate, liquid digestate, digestion sludge, activated sludge, recycled water, and discharge water. Comprehensive mass balance analyses including the mass balance on each wastewater component for individual treatments were carried out in this study. The removal and degradation of the wastewater components for individual unit operations were based on the data in the published literature as mentioned in Section 2.2. Methane generation from the anaerobic digestion operation is calculated using the stoichiometric conversion of COD to 0.395 L methane/g COD destroyed at the digestion temperature of 35°C [50].

Based on the mass balance analysis, an energy balance was conducted for each treatment scenario. Energy demand data for the BBR activated sludge operation and AD operation are obtained from pilot operations at both MSU Anaerobic Digestion Research and Education Center (ADREC) and U.S. ARMY Combat Capabilities Development Command (DEVCOM) Ground

Vehicle Systems Center (GVSC). Energy demand for the BBR-activated sludge operation includes both aeration and waste transfer (pumping) [20]. The energy demand for the aeration is 2.264 kWh/m^3 of wastewater. The pump used to transfer wastewater is a 0.37 kW unit with a flow rate of 4.56 m^3/hour. The UF/RO filtration operation needs three pumps including: UF feeding pump, RO boost pump, and RO high-pressure pump. The MP Flomax 8 pump with a power of 1.49 kW and a flow rate of 38 L/min was selected for both the UF feeding pump and RO boost pump. A high-pressure G10-E pump with 1.6 kW and a flow rate of 33 L/min and a pressure head of 16 bar was selected as the RO high-pressure pump.

As for anaerobic digestion with a power unit, the pump used to transfer the activated sludge is similar to the pump transferring the wastewater. The pump is a 0.37 kW unit with a flow rate of 4.56 m^3/hour. The heating energy demand ($Q_{AD,\ heat\ demand}$, kWh-e) to heat the activated sludge to the digestion temperature at 35°C was calculated using the heat equation as follows:

$$\boldsymbol{Q_{AD,\ heat\ demand} = m_{Sludge} \times C_{p,\ sludge} \times (T_{Digestion} - T_{Sludge}) \times 0.0002777} \quad \textbf{\textit{Eq. 1}}$$

Where m_{Sludge} is the mass amount of the activated sludge (kg), $C_{p,\ sludge}$ is the specific heat of the activated sludge (3.8 kJ/kg/C), $T_{Digestion}$ is the targeted digestion temperature of 35°C, T_{Sludge} is the average sludge temperature of 20°C, and 0.0002777 is the conversion factor of kJ to kWh-e.

The methane energy generated from biogas combustion is calculated by the methane heating value of 36 kJ/L methane. The power unit converts 30% of the methane energy for electricity generation, and 60% of the methane energy for heat generation. Both electricity and heat are used to maintain the digestion temperature and compensate for the energy demands from the activated sludge and filtration operations.

The mass and energy balance analysis determined the energy demand per unit of wastewater treated for individual treatment scenarios. This data is used for the following life cycle impact assessment and economic analysis.

2.4. Life cycle impact assessment

With the detailed mass and energy balance analysis, a life cycle impact assessment (LCIA) was carried out to evaluate the environmental impacts of individual treatments compared to the conventional wastewater treatment practices (Treatment E and F). The boundary of the life cycle impact assessment is from the source-separated wastewater to the end products of individual treatment including recycled water, renewable energy, discharging water, activated sludge or digestion sludge. Four impact categories related to carbon emission, air and water quality were chosen for the life cycle impact assessment: Global Warming Potential (GWP), Water Eutrophication Potential (WEP), Eco-Toxicity, and Smog Formation. CO_2 emissions were assumed to be biogenic for the activated sludge treatment and therefore have no impact on the treatment emissions, and N_2O emissions were analyzed from the wastewater flowrate through the system and the total nitrogen concentration in the wastewater. For AD, CO_2 emissions are assumed to be biogenic, while CH_4 and N_2O are greenhouse gases. For the land application of sludges and recycled or discharge water, CO_2 emissions are assumed biogenic. The data generated from the mass and energy balance was used to establish a life cycle inventory (Table S2). All emission factors for individual compounds are listed in Table S2. The EPA Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (TRACI) version 2.1 was used for the LCIA [51]. This tool provides characterization factors for a comprehensive list of substances. To calculate the impact for each category being considered, the substance mass from each emission source is multiplied by the listed characterization factors. Summing the total

emissions within each impact category results in the total impact score for each category. Contribution analysis was performed to elucidate the influences of different treatment scenarios on each impact category.

2.5. Economic analysis

To elucidate the viability of each treatment scenario, an economic analysis was conducted. For each treatment, Capital Expenditure (CapEx) and Operational Expenditure (OpEx) were utilized. Revenues from the recycled water and the renewable electricity generated from anaerobic digestion were also included in the analysis. The CapEx of the activated sludge treatment was calculated using the following reference equation (Services, 1978).

$$\boldsymbol{CapEx_{Activated\ sludge\ treatment} = 2.12 \times 10^{6} \times \left(\frac{Flow\ rate}{3.875 \times 10^{6}}\right)^{0.88}} \qquad \textbf{Eq. 2}$$

Where $CapEx_{Activated\ sludge\ treatment}$ is the CapEx of the activated sludge treatment unit ($/unit), 2.21 $\times 10^6$ is the conversion factor for scaling, the flow rate is the daily wastewater flow rate (L/day), 3.875×10^6 is the conversion factor of liter to million gallons, and 0.88 is the power coefficient.

The CapEx of both UF and RO units including the membranes and pumps is based on the cost ($30,000) of the commercial units with a treatment capacity of 38,000 liter/day. The CapEx of the studied UF and RO units was calculated using the linear relationship between treatment capacity and CapEx. The CapEx of the AD with CHP unit is based on the cost ($30,000) of a pilot unit with a capacity of 2,000 liter/day that MSU ADREC fabricated. The linear relationship between the treatment capacity and CapEx was used to calculate the CapEx of the AD with CHP in this study. In addition, the added direct costs (i.e., warehouse, site development, and additional piping) and indirect costs (i.e., prorateable costs, field expenses, office and construction, project contingency, and other costs) were set at 20% of the total capital investment.

The OpEx includes energy consumption of the entire treatment, replacement of the UF and RO membranes, sludge land application cost, and system maintenance. The electricity costs for the natural gas power and the diesel power are $0.1/kWh and $0.21/kWh, respectively. Twelve UF membranes and six RO membranes need to be replaced per year. The replacements of the UF and RO membranes are $200/each and $250/each, respectively. The annual maintenance cost is set at 2% of the total capital cost of the system (Activated sludge unit, AD unit, and filtration unit). Each treatment also needs a half-time operator. The salary of the operator is based on the current rate in Michigan. 50% of the labor burden is applied to include the benefits for the operator.

The revenues for Treatment A, B, C, and D are from recycled water and saved electricity from the renewable biogas electricity. The sale price of recycled water is set at $0.8/m^3$ water. The sale price of renewable electricity is set at $0.14/kWh.

The Modified Accelerated Cost Recovery System (MACRS) was used to calculate the annual depreciation of CapEx. The MACRS annual depreciation rates are 0.100, 0.188, 0.144, 0.115, 0.092, 0.074, 0.066, 0.066, 0.065, 0.065, 0.033, and 0.033 (after 10 years). Twenty years was set as the lifetime for each system in the treatment scenarios. Annual inflation of 3.2% was set for OpEx. The tax rate is 35%.

The net cash flow based on depreciated CapEx and inflated OpEx was conducted to determine the cost of each treatment scenario. A sensitivity analysis was carried out to elucidate the effects of revenue, labor, energy demand, and operational parameters of the treatment systems. Each parameter was varied by ± 25%, while all other parameters were held constant, and the subsequent change in impact was recorded and compared.

3. Results and discussion

3.1. Characteristics of source-separated wastewater

The water quality data from the greywater and blackwater obtained at a military base camp were analyzed for this study (Table 2). Since the greywater at the military base camp only contains wastewater from laundry and shower, it has the TSS, TN, TP, COD, and BOD_5 contents much lower than blackwater. As for PPCPs, the blackwater has a total PPCP content of 3,882 ug/L, which is 2.4 times higher than that in the greywater (1,611 ug/L). The major PPCP chemicals in greywater are DEET of 1,174 ug/L, methylphenol of 108 ug/L, salicylic acid of 82 ug/L, ibuprofen of 64 ug/L, and benzyl alcohol of 42 ug/L. The top five chemicals in blackwater are methylphenol of 1,126 ug/L, DEET of 872 ug/L, phenol of 518 ug/L, salicylic acid of 372 ug/L, and caffeine of 284 ug/L. The characteristics data indicates that two wastewaters are significantly different from each other. To efficiently treat them and recycle the water, the detailed techno-economic analysis and life cycle impact assessment of four different treatment approaches along with two control treatments were carried out in the following sections.

Table 2. Characteristics of blackwater and greywater.

Characteristics	Greywater	Blackwater
BOD_5 (mg/L)	188±14	1478±353
COD (mg/L)	386±21	3360±1278
TN (mg/L)	38±17	320±259
TP (mg/L)	3±0.3	37±26
TSS (mg/L)	27±5	801±544
Pharmaceuticals and personal care products (PPCPs)		
Acetone (ug/L)	31	350±218
Benzyl alcohol (ug/L)	42±13	58±19
Caffeine (ug/L)	19±11	284±278
Chloroform (ug/L)	9±3	16±3
N, N-Diethyl-Meta-Toluamide(ug/L)	1174±731	872±1287
Di(2-ethylhexyl) phthalate (ug/L)	21±10	7
Ibuprofen (ug/L)	64±116	172±72
Methylphenol (ug/L)	108±33	1126±284
Nicotine (ug/L)	20±4	104±102

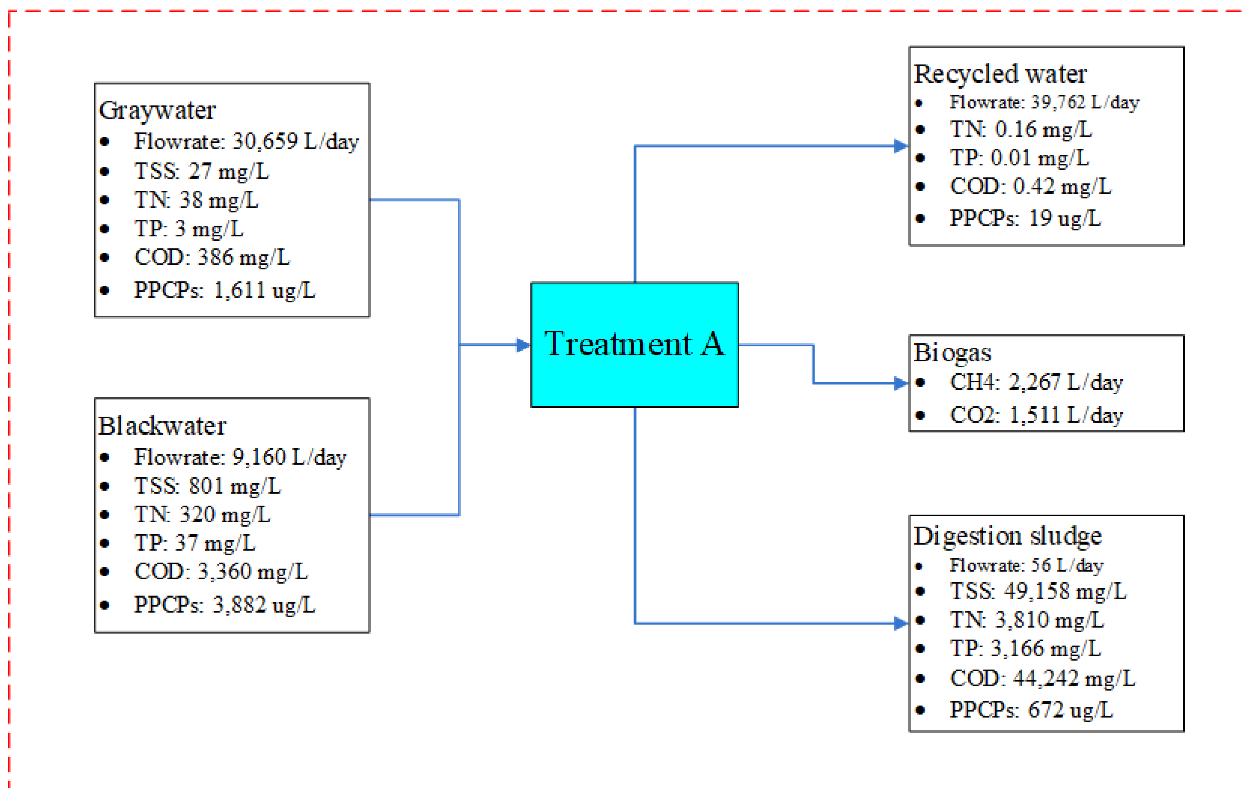
Table 2 (cont'd)

Permethrin (ug/L)	20±16	3±4
Phenol (ug/L)	21±6	518±182
Salicylic acid (ug/L)	82±78	372±251
Total PPCPs (ug/L)	**1,611**	**3,882**

*: *The data are an average of at least 3 biological replicates with standard deviation.*

3.2.Mass balance on different treatment scenarios

The mass balance was conducted on the six different treatment scenarios to evaluate their treatment performance (Figure 3 and Figure S1). Among the six treatment scenarios, Treatment A and B with AD and activated sludge have the highest recycled water daily flowrates of 39,780 and 39,794 L/day, respectively, with corresponding recovery efficiencies of 99.8 and 99.9%. Treatment D with activated sludge treatment on blackwater, membrane filtration on greywater, and without AD shows a recycled water daily flowrate of 39,575 L/day with the recovery efficiency of 99.4%, which is lower than Treatment A and B, but higher than Treatment C without source separation and AD (the recycled water flowrate of 39,269 L/day with the recovery efficiency of 98.6%). Regarding wastewater recovery, it is apparent that wastewater source separation and employment of AD significantly improves water recovery from wastewaters. All four treatment scenarios show better performance than the control (Treatment E) which includes both discharge water and activated sludge which must be discharged. Meanwhile, Treatment A and B with AD generates 56 L/day and 25 L/day of concentrated digested sludge, respectively, which are much lower than the amount of activated sludge from the treatment scenarios without AD. Treatment C and D, and E generate activated sludge of 549 L/day, 244 L/day, and 177 L/day, respectively.



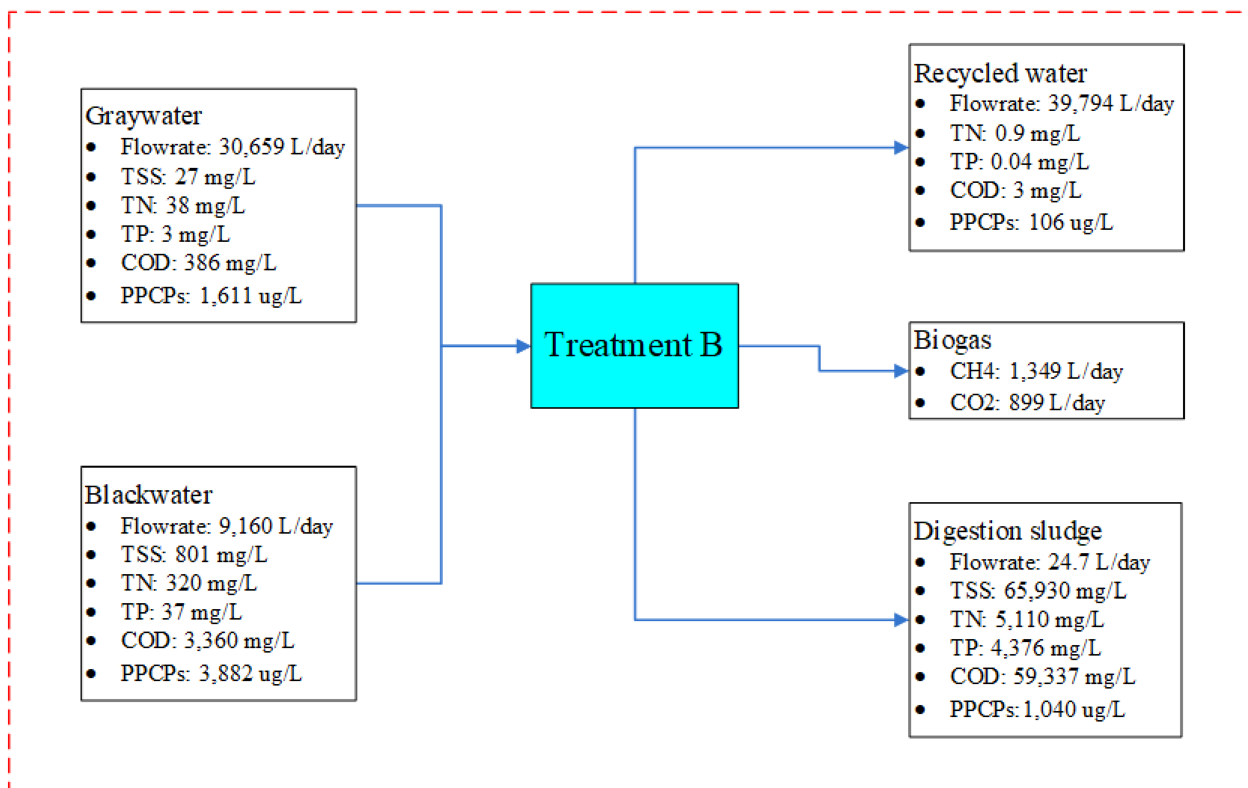
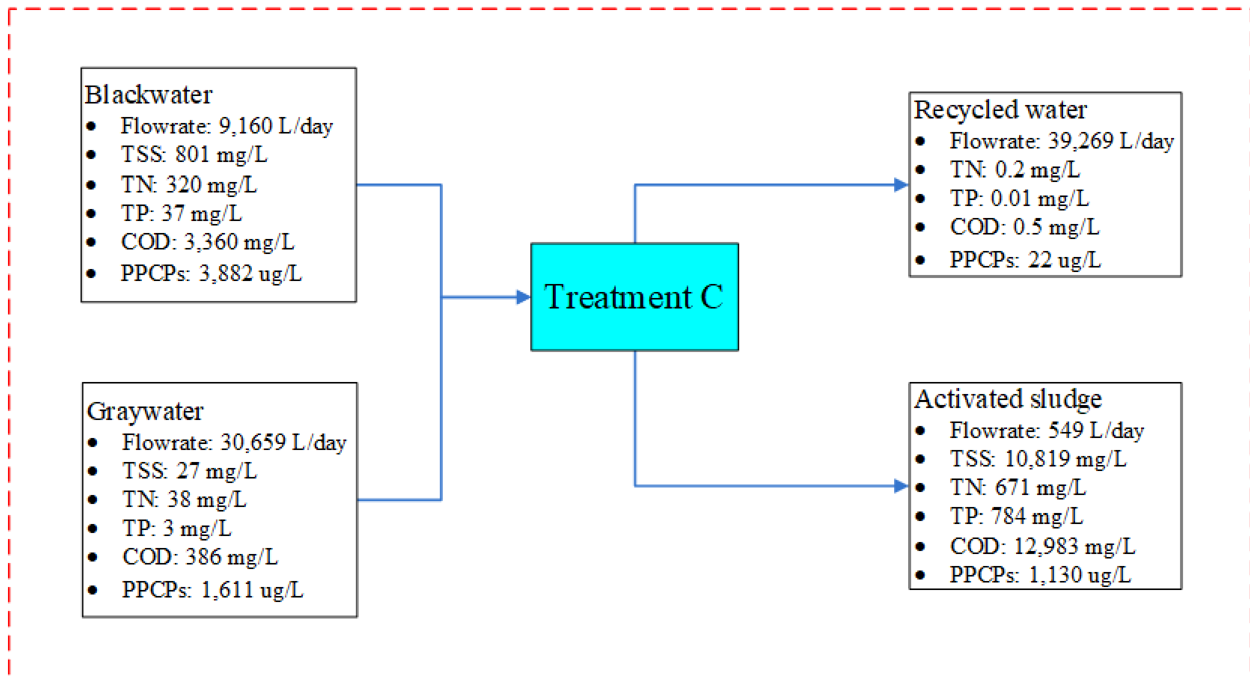


Figure 3. Mass balance of different treatment scenarios (Treatment A-F).

Figure 3 (cont'd)



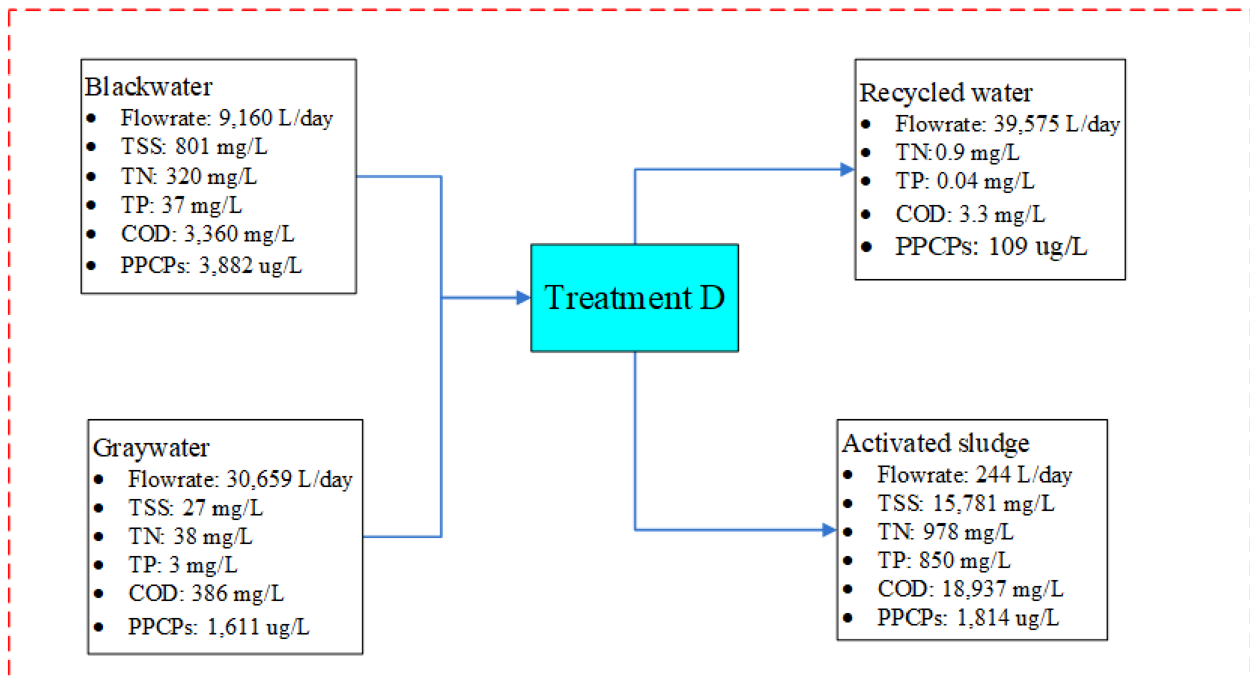
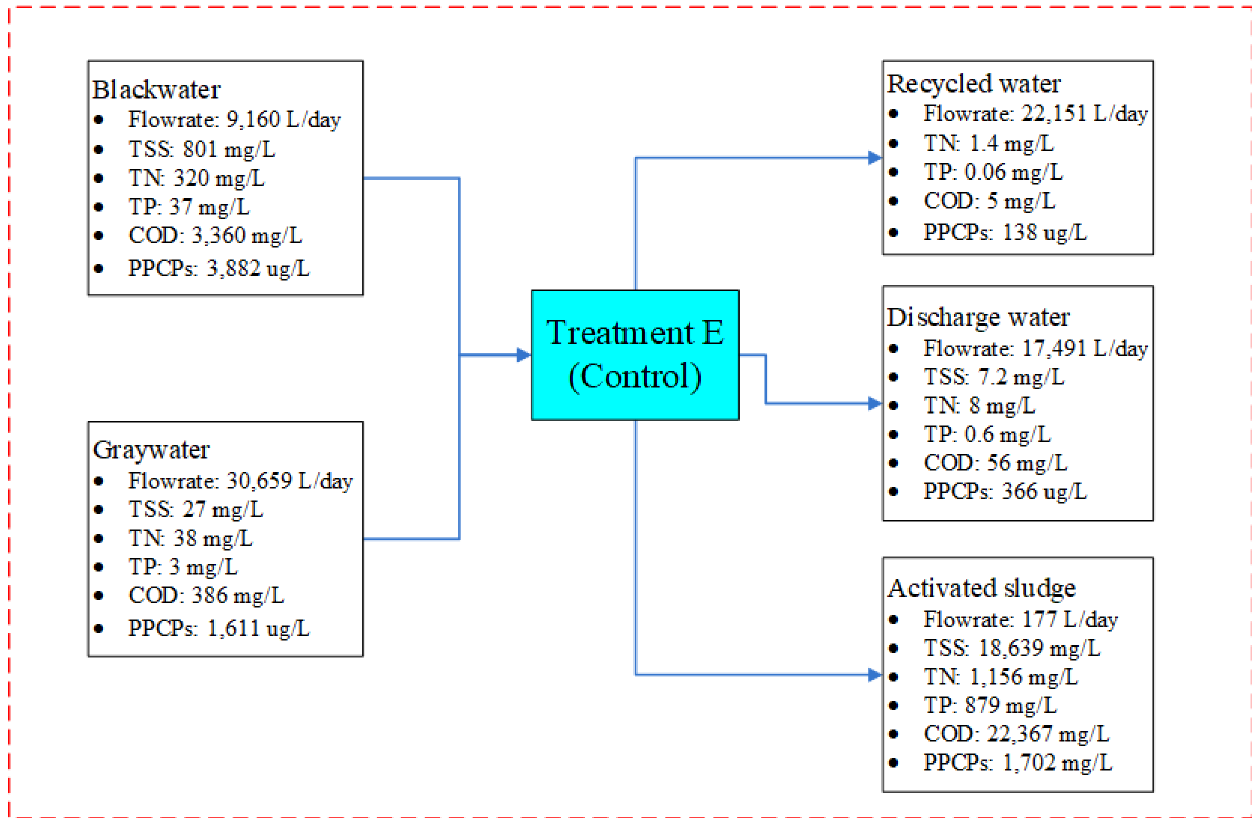


Figure 3 (cont'd)



As for the removal of TSS, TN, TP, and COD in the recycled water, there are no significant differences among treatment and control scenarios (A, B, C, D, and E) (Table 3). The concertation of TN, TP, and COD in the recycled water for the four treatment scenarios are all below 1 mg/L, 0.05 mg/L, and 3.5 mg/L, respectively, which are lower than the discharged water from the control treatment (Treatment E) (Figure 3).

The removal of PPCPs showed different performances between the treatments (Table 3). Treatment A and C have 19 and 22 ug/L of PPCPs, respectively, in the recycled water, representing 99% of the PPCP removal, which are better than other treatments and control (106, 109, and 138 ug/L in the recycled water of Treatment B, D, E, respectively). It is apparent that UF/RO operation on both blackwater and greywater significantly reduce PPCPs in the recycled water. Due to the high PCPP removal efficiency of the activated sludge treatment, Treatment A

and C that used the activated sludge to treat combined greywater and blackwater had better PCPP removal than Treatment B and D with source water separation. DEET and salicylic acid are the main PPCPs that remain in the recycled water from Treatment A and C (Table S2). The DEET in the recycled water are 16 and 18 ug/L for Treatment A and C, respectively. The salicylic acid contents in the corresponding treatments are 3 ug/L. While the main PPCPs in the recycled water of Treatment B and C are DEET (87 ug/L), salicylic acid (4 ug/L), benzyl alcohol (3 ug/L), and ibuprofen (5 ug/L) (Table S2). While PPCPs in sludges (both digestate sludge for treatment A and B as well as activated sludge for C, D, and E) are much higher than those in recycled water. The main PPCP compounds in the sludge are Di-2-ethylhexyl-phthalate, DEET, and salicylic acid (Table S2). Treatment A and B with AD have the PPCPs of 672 and 1,040 ug/L in the digestate sludge, respectively, which are lower than those in the activated sludge of Treatment C, D, and E (1,130, 1,814, and 1,702 ug/L respectively). It is due to the fact that Di-2-ethylhexyl-phthalate, the main PCPP compound in the sludge, is degraded by anaerobic digestion, but cannot be degraded by the activated sludge process. It is accumulated in the activated sludge from Treatment C, D, and E.

Table 3. Recycled water generation and removal of key compounds in blackwater and greywater from different treatment approaches. [a]

Treatment	Water recovery [b] (%)	TSS removal (%)	TN removal (%)	TP removal (%)	COD removal (%)	PPCPs removal (%)
A	99.8	100	99.84	99.92	99.96	99.11
B	99.9	100	99.14	99.66	99.69	95.02
C	98.6	100	99.83	99.94	99.96	98.99
D	99.3	100	99.15	99.66	99.69	94.94
E	55.6	100	99.25	99.70	99.72	96.41

a. *Removal is the percentage of the compound removed during the treatment processes, comparing the concentration remaining in the recycled water with the initial concentration in the blackwater and greywater.*
b. *Water recovery is the percentage of recycled water vs. the total amount of treated greywater and blackwater.*

3.3. Energy balance of different treatment scenarios

The energy balance was conducted to evaluate the energy consumption and production from each treatment scenario (Table 4). The results show that the treatment scenarios with source water separation (B and D) have lower net energy demand for the activated sludge operation (169 and 170 kWh-e/day, respectively) than the corresponding treatment scenarios without source water separation of A and C (207 and 204 kWh-e/day, respectively), which is caused by the reduced wastewater amount required to be treated by the activated sludge. The data further indicate that renewable energy generation from the activated sludge for Treatment A and B has a minimum impact to improve the energy balance of both treatment systems due to the fact that less activated sludge is produced from the small-scale decentralized operation. Since the control scenarios generate both recycled water and discharge water, net energy demand per cubic meter of recycled water is used to compare the performance between treatment and control scenarios (Table 4). The data clearly indicate that Treatment B requires less energy (4.2 kWh-e/m^3 recycled water) than other treatment and control scenarios.

According to the mass and energy balance results, Treatment B with source water separation and AD shows better performance on water recycling, sludge generation, and energy demand than the other three treatments (A, C, and D) and control (E).

Table 4. Energy balance of different treatment approaches.

Treatment	Energy input (kWh-e/day)				Energy output (kWh-e/day)	Net energy demand (kWh-e/day)	Net energy demand (kWh-e/m^3 recycled water)
	Activated sludge	Anaerobic digestion	UF	RO	Anaerobic digestion		
A	-110.2	-8.8	-36.0	-72.3	20.3	-207.0	5.2
B	-69.0	-3.9	-36.0	-72.3	12.1	-169.1	4.2
C	-97.7	-	-35.5	-71.1	-	-204.3	5.1
D	-62.6	-	-35.8	-71.9	-	-170.3	4.3

Table 4 (cont'd)

E (Control)	-53.4	-	-20	-40.2	-	-113.6	5.2

3.4. Life cycle impact assessment and comparison of different treatment scenarios

The Life Cycle Impact Assessment (LCIA) was conducted to elucidate the environmental impacts of the five different treatment and control scenarios. Global warming Potential (GWP), Water Eutrophication Potential (WEP), Smog Potential, and Eco-Toxicity are the four impact factors evaluated in this study. For GWP and smog formation, both natural gas and diesel fuels were analyzed for power generation. The life cycle inventory for the LCIA is presented in Table S2.

GWP of each treatment and control scenario was calculated based on unit operations of activated sludge treatment, AD, and UF/RO, and final products of recycled or discharge water and sludge (digested sludge or activated sludge for land application) (Figures 4a and 4b). For the activated sludge treatment, CO_2 emissions are biogenic and therefore have no impact on the treatment emissions, and N_2O emissions were analyzed from the wastewater flowrate through the system and the total nitrogen concentration in the wastewater. For AD, CO_2 emissions are biogenic, while CH_4 and N_2O are greenhouse gases. For the land application of sludges and recycled or discharged water, CO_2 emissions are biogenic. The results show that GWPs of Treatment A, B, C, and D and Control E with natural gas-based electricity are 40, 33, 39, 32, and 21 metric tons CO_2-e/year, respectively (Figure 4a). The corresponding GWPs with diesel electricity are 58, 48, 57, 47, and 32 metric ton CO_2-e/year (Figure 4b). Using diesel electricity increases GWPs of all treatments including controls. The data of GWP per m^3 recycled water further concludes that Treatment B and D have lower GWP for both power conditions of natural gas and diesel (2.21-2.29 and 3.25-3.31 kg CO_2-e/m^3 recycled water, respectively) among all treatment and control scenarios.

Smog as air pollution is caused by the reactions between sunlight, nitrogen oxides, and other volatile organic compounds. The results show that all treatments and controls powered by diesel electricity have much higher smog potential than those powered by natural gas electricity (Figures 4c and 4d). Treatment A, B, C, D and Control treatment E with natural gas electricity have smog potentials of 2.9, 2.3, 2.8, 2.3, and 1.6 metric tons O_3/year, respectively. The corresponding smog potentials with diesel electricity are 37, 30, 37, 30, and 20 metric tons O_3/year. Based on the data of smog potential per m^3 recycled water (Figure 4c and 4d), Treatment B has the lowest values of 0.16 and 2.08 kg O_3/m^3 recycled water for natural gas-based and diesel electricity, respectively, among all five treatment and control scenarios.

WEP was calculated for each scenario using the total amount of N and P discharged to the environment from the treatment. Since power sources do not influence WEP, there are no differences in all scenarios between natural gas and diesel power. WEPs of Treatments (A, B, C, and D) and control treatment (E) are 0.6, 0.4, 1.0, 0.7, and 0.7 metric ton N eq/year, respectively (Figure 4e). As for WEP per m^3 recycled water, Treatment B also has the lowest number of 24 g N eq/m^3 recycled water among all treatment and control scenarios. According to the distribution of TN and TP in the discharge water and sludge of each treatment and control scenario, the discharge of the activated sludge and the digestion sludge has a much larger impact than the recycled and discharged water (Table S2).

Eco-Toxicity potentials were calculated using compounds in the discharge water, digestion sludge, and activated sludge that has ecological impacts. Permethrin, Di-2-ethylhexyl-phthalate, Salicylic acid, DEET, Benzyl alcohol, and Chloroform are the compounds used for the calculation. Power sources again do not influence Eco-Toxicity. Eco-Toxicity potentials of Treatments (A, B, C, and D) and control treatments (E) are 0.00009, 0.00007, 0.26, 0.19, and

10.2 CTUeco/year, respectively (Figure 4f). It is apparent that all four treatment scenarios significantly reduce Eco-Toxicity potential of the wastewater. Treatment A has the lowest Eco-Toxicity potential (9×10^{-5} CTUeco/year and 6×10^{-9} CTUeco/m^3 recycled water) among all treatment and control scenarios. The Eco-Toxicity analysis elucidates that biological treatments (activated sludge and anaerobic digestion) of greywater and blackwater can effectively remove PPCPs and lead to less eco-toxicity impact on the environment. The analysis also shows that the discharge water had a much larger impact on the eco-toxicity than the digestion sludge or the activated sludge (Table S2).

The life cycle impact assessment elucidates that Treatment B has an overall less negative impact on the environment than other treatment and control scenarios.

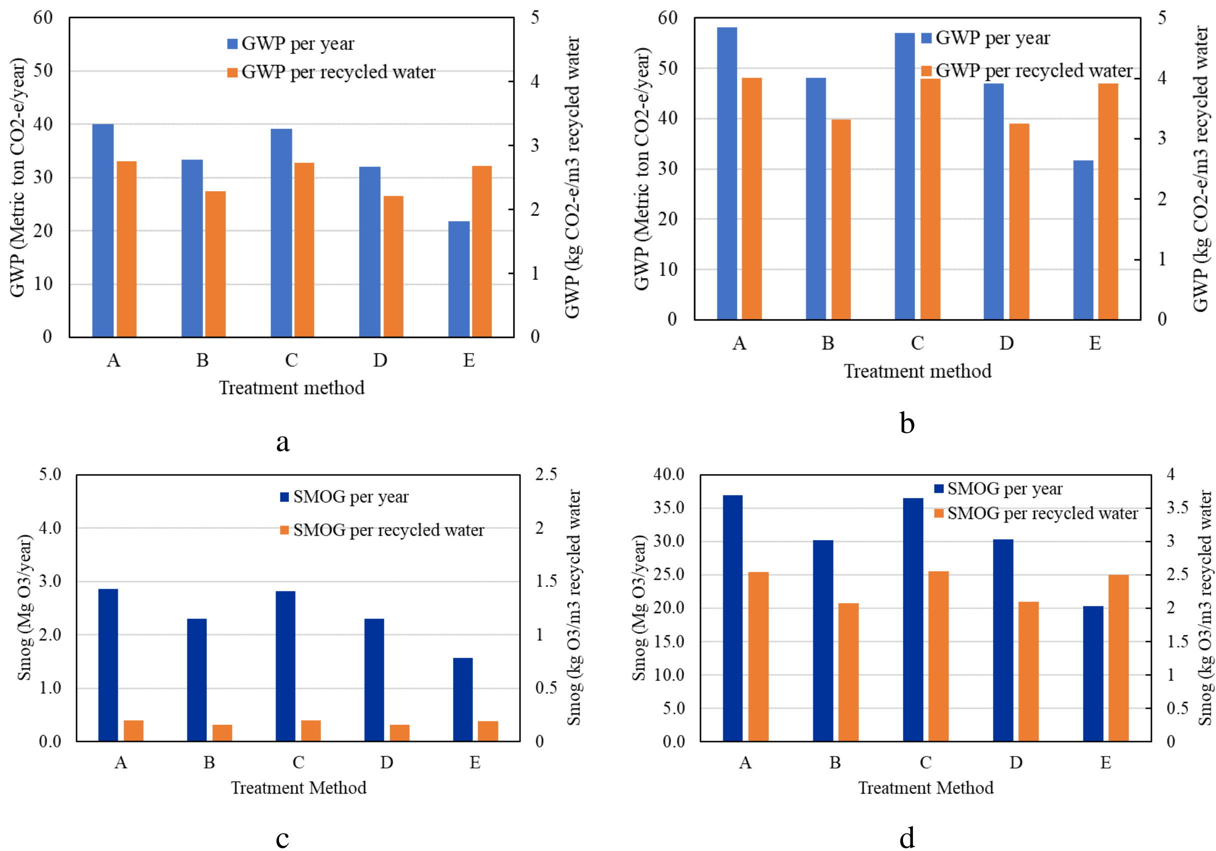
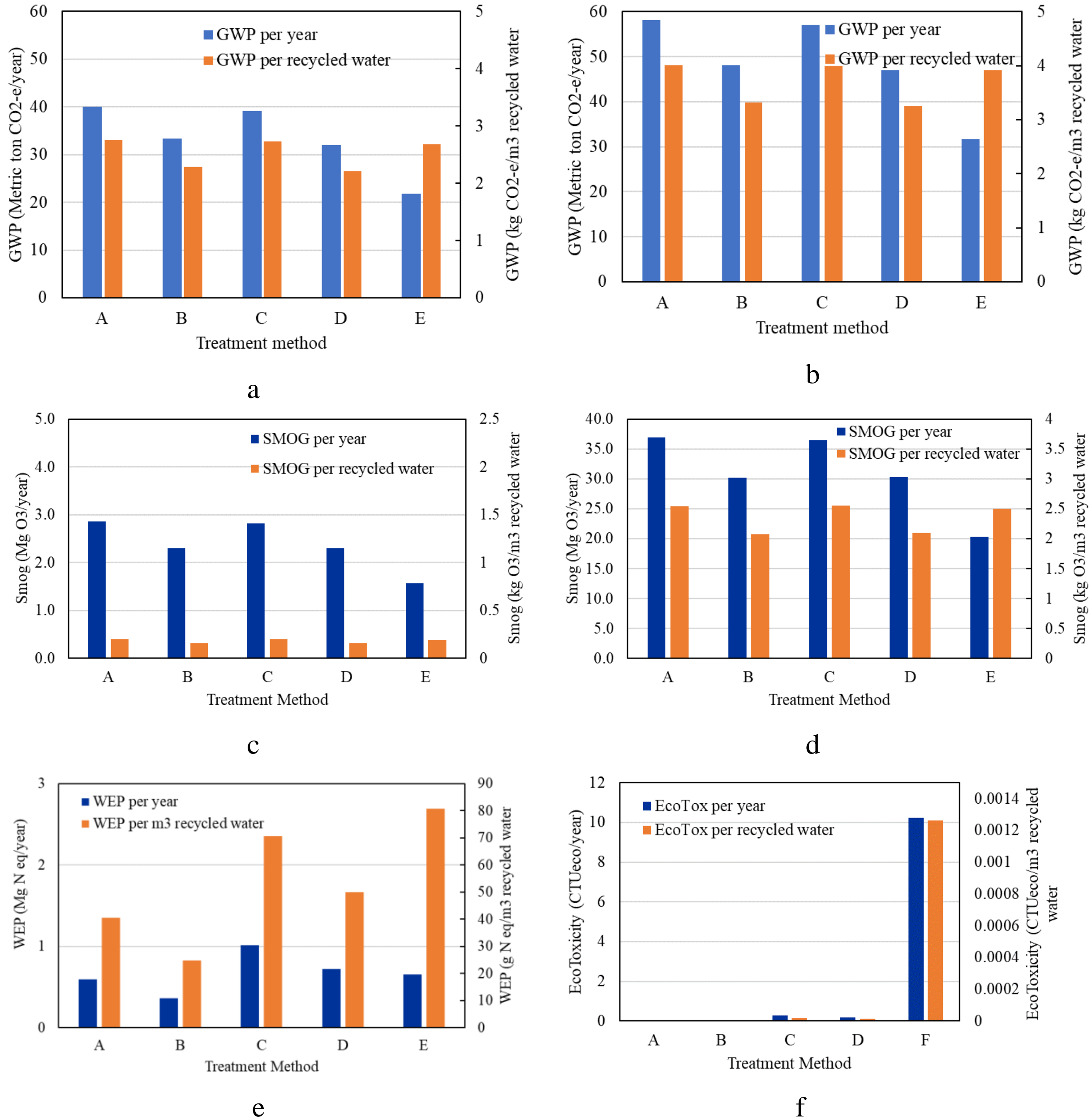


Figure 4. Contribution analysis of individual impact categories for different scenarios.

Figure 4 (cont'd)



a. Global warming potential with electricity from natural gas

b. Global warming potential with electricity from diesel fuel

c. Smog formation potential with electricity from natural gas

d. Smog formation potential with electricity from diesel fuel

e. Water eutrophication potential

f. Eco-toxicity potential

3.5. Economic assessment

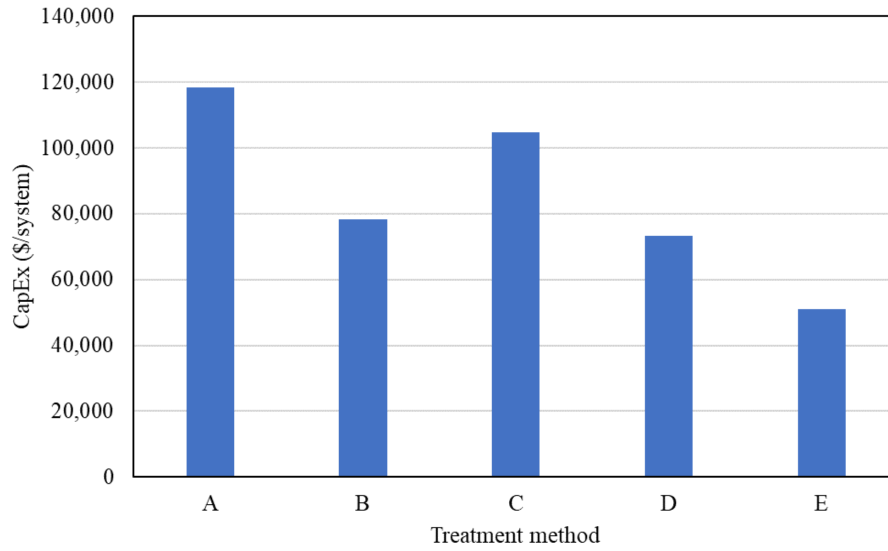
The economic assessment is important to determine the viability of the real-world application of the different treatment scenarios for decentralized wastewater treatment. CapEx, OpEx, and revenues are the parameters to assess the economic performance of the treatment and control scenarios. As presented in Table 5 and Figure 5, the CapEx of treatment scenarios (A, B, C, and D) and control scenario (E) are $118,468, $78,387, $104,820, $73,181, and $51,107, respectively. Since the four treatment scenarios have more unit operations than the control scenario, they are more expensive.

Due to the cost differences between diesel electricity and natural gas electricity, OpEx for the treatment scenarios with diesel electricity are higher than the treatment scenarios with natural gas electricity. OpEx for the treatment scenarios (A, B, C, and D) and the control scenario (E) with diesel electricity are $60,920, $56,637, $60,665, $56,479, and $51,547 per year, respectively. While corresponding OpEx with natural gas electricity are $51,790, $49,357, $52,451, $49,637, and $46,982 per year. Due to the source water separation and AD, Treatment B had the lowest OpEx among all treatment scenarios. While it is slightly higher than the control scenario.

Revenues of the treatment scenarios (A, B, C, and D) and the control scenario (E) are $11,958, $11,824, $11,467, $11,556, and $6,468 per year, respectively. Treatments A and B generate slightly more revenue than the other treatment and control scenarios since less sludge leads to more recycled water being recycled. Due to small-scale operation of the decentralized treatment, the recycled water is the key source of revenue generation compared to the energy saving of biogas electricity for Treatment A and B (Table 5).

The cash flow analysis demonstrates that considering a 20-year payback period, Treatment B and D with source water separation have lower treatment costs among the four treatment scenarios. Control E has lower treatment costs than Treatment B and D. However, the control scenario generates less recycled water, requires more energy, and has more negative environmental impacts than Treatment B and D.

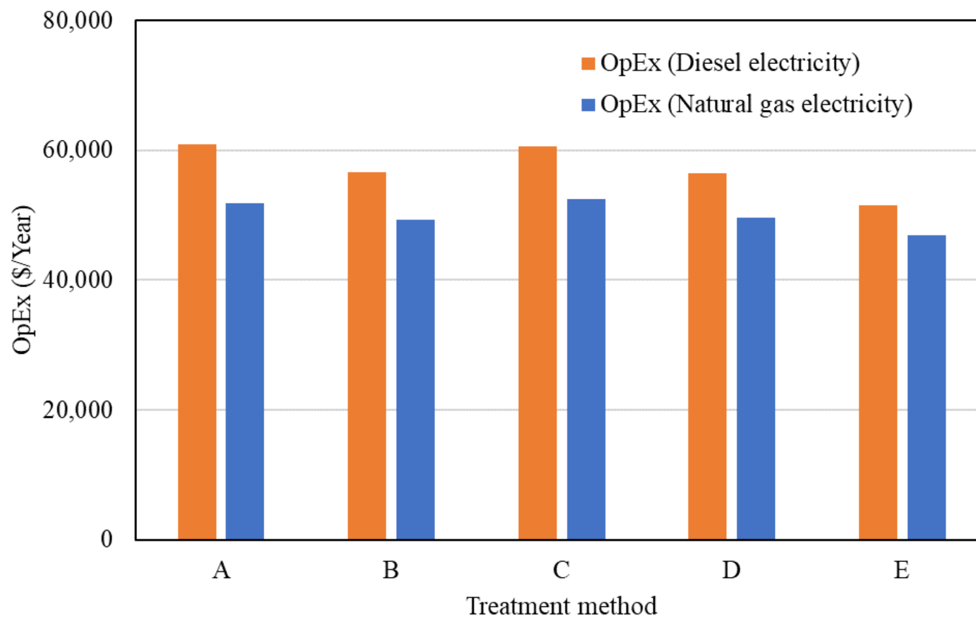
Considering both life cycle and technical aspects, Treatment B is a preferred treatment scenario. A sensitivity analysis was then conducted on three CapEx items (activated sludge, reverse osmosis, and ultrafiltration), two OpEx items (labor and energy demand), and revenue to delineate their influences on the economic performance of Treatment B (Figure 6). A decrement of 25% in the labor cost could reduce the treatment cost by $0.65/m^3$ wastewater for both natural-gas-powered treatment and diesel-powered treatment, which is the largest reduction among these six items. Meanwhile, an increment of 25% in revenue could reduce the treatment costs by $0.15/m^3$ wastewater for both cases. Besides labor cost and revenue, the other four items of activated sludge, reverse osmosis, ultrafiltration, and energy demand have much less impact on the cost of the treatment. According to the sensitivity analysis, improving the revenue and reducing labor are two key factors to further enhance the economic performance of the treatment.



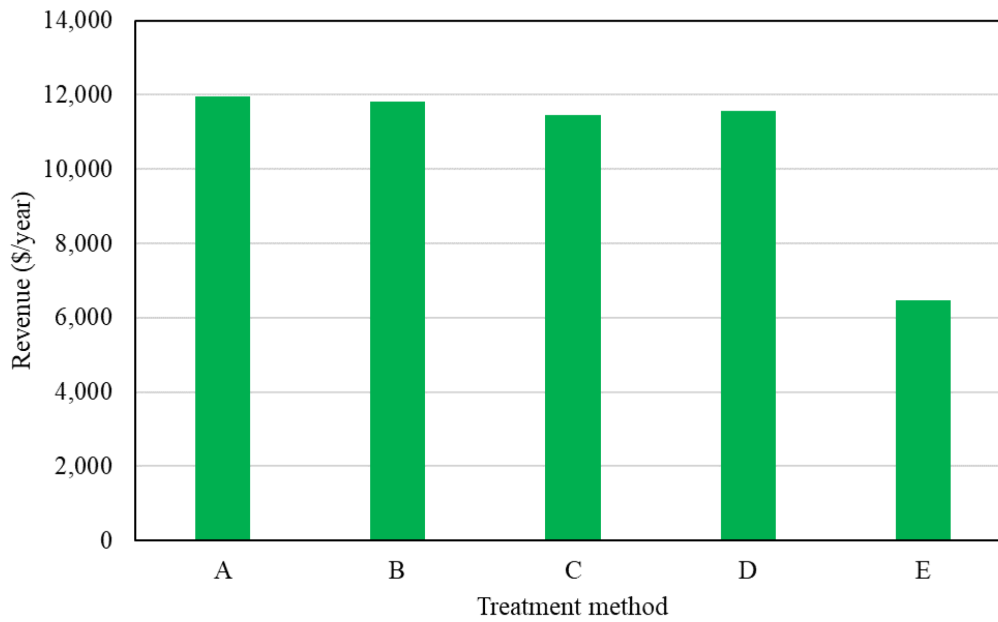
(a)

Figure 5. The CapEx, OpEx, and treatment cost of different treatment scenarios. a. CapEx; b. OpEx; c. Revenue; d. Treatment cost.

Figure 5 (cont'd)

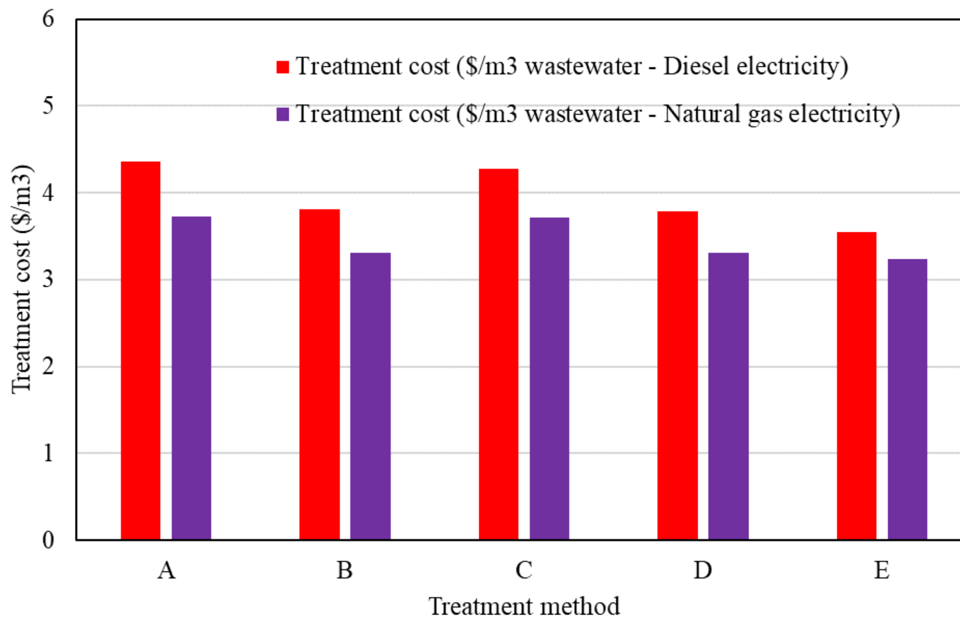


(b)



(c)

Figure 5 (cont'd)

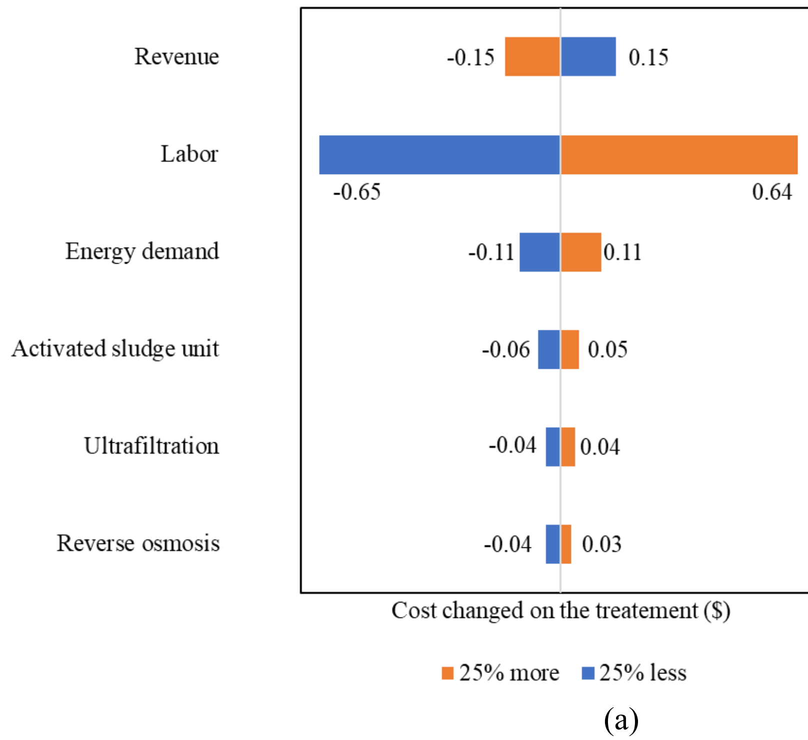


(d)

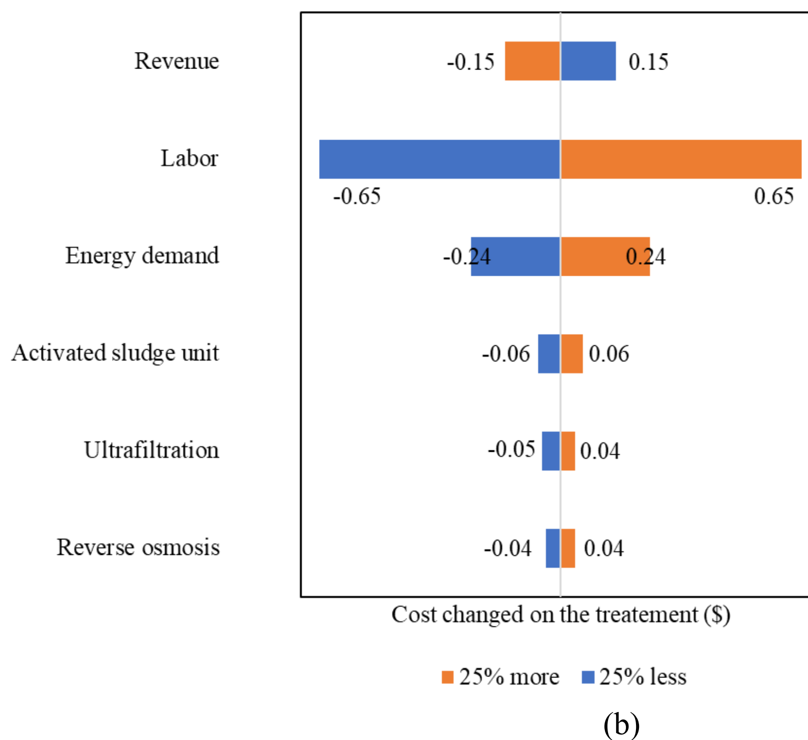
Table 5. Economic performance of different treatment scenarios.

	A	B	C	D	F
Capital expenditure (CapEx)					
Activated sludge treatment ($/unit)	52,212	24,756	50,059	24,485	18,457
UF ($/unit)	18,466	18,480	18,236	18,378	12,102
RO ($/unit)	15,756	15,769	15,561	15,682	10,327
AD ($/unit)	8,340	3,705	-	-	-
Indirect and direct CapEx cost (20% of the total capital) ($/unit)	23,694	15,677	20,964	14,636	10,222
Total CapEx ($)	118,468	78,387	104,820	73,181	51,107
Operational expenditure (OpEx)					
Energy cost ($/year)*	8,300/17,430	6,617/13,897	7,468/15,683	6,220/13,061	4,150/8,715
UF membrane replacement ($/year)	2,400	2,400	2,400	2,400	2,400
RO membrane replacement ($/year)	1,500	1,500	1,500	1,500	1,500
Sludge land application ($/year)	194	86	1,906	847	614
System maintenance ($/year)	1,895	1,254	1,677	1,171	818
Labor and labor burden ($/year)	37,500	37,500	37,500	37,500	37,500
Total OpEx*	51,790/60,920	49,357/56,637	52,451/60,665	49,637/56,479	46,982/51,547
Revenue					
Recycled water ($/year)	11,611	11,620	11,467	11,556	6,468
Renewable electricity ($/year)	347	204	-	-	-
Total revenue	11,958	11,824	11,467	11,556	6,468
Treatment cost ($/m^3 wastewater)*	3.73/4.36	3.31/3.81	3.71/4.27	3.31/3.78	3.24/3.55

**: The numbers in the front are for natural gas electricity. The numbers in the back are for diesel fuel electricity.*



(a)



(b)

Figure 6. Sensitivity analysis of key unit operations on the cost of treatment B.

a. Natural gas electricity as the power source, the baseline cost is $3.18/m3 wastewater; b. Diesel electricity as the power source, the baseline cost is $3.58/m3 wastewater.

4. Conclusions

This study comprehensively analyzed the techno-economic and environmental factors of different treatment scenarios for decentralized wastewater treatment. Among five treatment and control scenarios, Treatment B integrating activated sludge, AD, and UF/RO filtration to separately treat blackwater and greywater led to a preferred treatment process with a water recovery efficiency of 99.9% and trace nutrient and PPCP concentrations in the recycled water (106 ug/L of PPCPs, 0.9 mg/L of TN, 0.04 mg/L of TP, and 3 mg/L of COD). The treatment has a minimum net energy demand of 4.2 kWh-e/m^3 recycled water (169 kWh-e/day). The life cycle impact assessment demonstrates that Treatment B has an overall less negative impact on the environment than the other treatment and control strategies. The economic analysis concludes that Treatment B also has lower treatment costs of \$3.31/m^3 wastewater and \$3.81/m^3 wastewater, for diesel electricity and natural gas electricity, respectively. These results clearly demonstrate that the collection of source-separated wastewaters and the combination of activated sludge, AD, and membrane technologies can create a technically sound and economically feasible decentralized solution to treat wastewater.

This chapter represents published work: Thomas, Benjamin D., Marks M, Smerigan B, Aburto-Vazquez G, Uludag-Demirer S, Dusenbury JS, Liao W. Life Cycle Impact and Economic Assessment of Decentralized Strategies to Treat Source-Separated Wastewater. Journal of Cleaner Production 64(2024) 105550

CHAPTER 2: DECENTRALIZED HIGH-STRENGTH WASTEWATER TREATMENT USING A COMPACT AEROBIC BAFFLED BIOREACTOR

1. Introduction

The Environmental Protection Agency (EPA) estimates that $271 billion will be required for the wastewater infrastructure over the next 25 years [5]. This massive cost burden is required to replace and repair old and failing infrastructure, and it is estimated that 95% of the spending for water infrastructure is paid for at the local level [6]. Decentralized wastewater treatment can be a potential solution to reduce the costly burden facing a large percentage of the wastewater infrastructure by serving rural and distributed regions or reducing the growing burden on existing infrastructure. It is estimated that in centralized wastewater management, 80-90% of the total cost is attributed to the transportation of wastewater, with only 10-20% attributed to the treatment process [3]. The current centralized municipal wastewater system and corresponding treatment technologies have been intensively investigated in the past decades [4]. However, decentralized, less typical wastewater treatment operations (rural and suburban communities, small industrial/agricultural operations, and military bases) have not been investigated as deeply as municipal wastewater treatment plants and are therefore not well understood and conventionalized. The wastewater produced from small-scale operations often has a much higher pollution concentration than typical municipal wastewaters due to the mixing of some concentrated waste streams (e.g., food wastes, latrine waste) with less dilution [52,53]. The composition of such wastewater is generally high strength with the elevated concentrations of biological oxygen demand (BOD_5) (>300 mg/L), chemical oxygen demand (COD) (>900 mg/l), total suspended solids (TSS) (>600 mg/L), or fats/oils/greases (FOG) (>40 mg/L) [52]. The wastewater management for such wastewater from small-scale operations may be best treated using a decentralized solution. In addition, an emerging circular economy approach of

wastes/wastewater management has gained traction in recent years [54,55]. Decentralized wastewater treatment fits into the concept of circular economy. The treated water can be recycled locally for non-potable uses, and the nutrient rich sludge can be used as a fertilizer in nearby farms or gardens. Such an approach will not only benefit the environment but also create jobs and help the local economy.

Activated sludge processes as a biological treatment system are widely used to treat wastewater [4]. As it is well known, activated sludge is a mixture of aerobic microorganisms that oxidize biodegradable compounds (organic carbon (C) and nutrients (nitrogen (N) and phosphorus (P)) in wastewater. The excess microbial growth is controlled by recycling and wasting the active microorganisms (mixed liquor suspended solids (MLSS). The major groups of microorganisms found in activated sludge are bacteria, protozoa, metazoa, filamentous bacteria, and algae/fungi. Among them, bacteria are the largest group that comprises approximately 95% of the total microorganisms in activated sludge [56]. They are the primary microbes in charge of metabolizing a wide range of organic compounds as well as removing inorganic nitrogen and phosphorus. The key physiological groups of bacteria in activated sludge include: chemoorganohetorotrophs (e.g., *Proteobacteria* and *Desulfovibrio*) that use fermentation and respiration to degrade and utilize organic compounds in wastewater, chemolithoautotrophs (e.g., *Candidatus, Nitrosomonas, Nitrobacter,* and *Ferroplasma*) that oxidize a range of inorganic compounds to obtain energy, and photoorganoheterotrophs and photolithoautotrophs that use light as an energy source but utilize organic and inorganic carbon and nutrient sources, respectively [57].

During the activated sludge process, maintaining microbial biomass, along with their metabolic activities, is critical to achieving efficient treatment, particularly for high-strength

wastewater. Many technologies have been developed to enhance microbial biomass activities in biological wastewater treatment, such as aerobic fluidized bed (AFB), rotating biological contactors (RBC), fixed-film bioreactors (FFB), membrane bioreactor (MBR), and activated sludge [58]. Among them, the activated sludge process is the most traditional method that is adopted by municipalities since it has high treatment performance, requires minimum maintenance, and does not need supportive media and complicated process control [58]. However, high concentrations of the nutrients (greater than 300 mg N/L and 40 mg P/L) in high-strength wastewater require biological treatment with enhanced microbial activities to remove them [52]. Consequently, high concentrations of MLSS need to be maintained in the process by increasing either biological growth or the recycling ratio. In contrast to normal strength large-scale activated sludge processes, small-scale high-strength activated sludge processes require much greater (or additional) settling and pumping steps to recirculate the sludge which significantly increases capital and operational costs. This limits the implementation of the activated sludge process to treat high-strength wastewater at a small scale. It has been reported that a baffled bioreactor (BBR) configuration is able to maintain high concentrations of microbes without using biofilm growth support media, additional settling steps, or pumping to recycle activated sludge [59].

This study focused on a containerized BBR as the primary component of decentralized wastewater treatment/utilization to treat a high-strength wastewater – blackwater. In this study, the term blackwater is used to describe wastewater consisting of latrine and kitchen wastewater, which has much higher nutrient contents than normal sewage or greywater (Table 6). Chemical and amplicon sequencing analyses were conducted to elucidate the effects of microbial communities on the treatment and compare effluent water quality under different feed amounts.

Mass, energy, exergy, and economic analyses were then carried out to evaluate the performance and feasibility of the BBR to treat blackwater.

2. Materials and methods

2.1. The blackwater composition and feeding the baffled bioreactor

The blackwater was prepared at the Delhi Township Wastewater Treatment Plant in Holt, Michigan by mixing the primary clarifier sludge and raw sewage in a wet well to achieve the target blackwater composition as shown in Table 6. To achieve uniform mixing by counterflow effect in the wet well, primary clarifier sludge was fed from the bottom of the tank and raw sewage was fed from the top (Figure 7). Feeding pumps were controlled by float switches in the wet well.

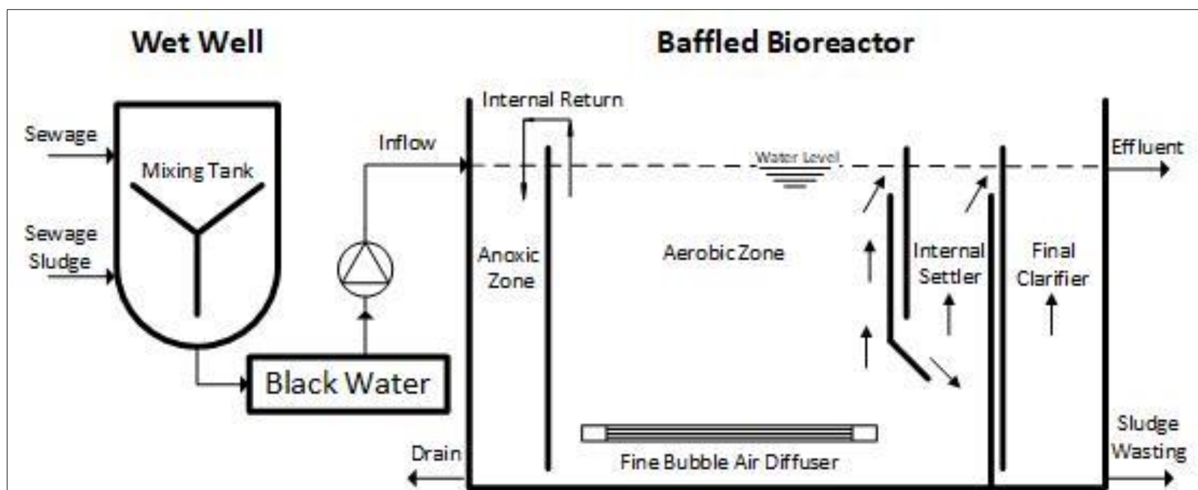


Figure 7. The blackwater feeding unit and the BBR (Liu et al., 2012).

*Table 6. Characterization of the blackwater. **

Parameter	**Blackwater**
Turbidity (NTU)	1687 ± 592
TS (mg/L)	1904 ± 466
TSS (mg/L)	1168 ± 470
COD (mg/L)	2806 ± 811
BOD_5 (mg/L)	1522 ± 432
NH_3-N (mg/L)	41 ± 8
NO_2-N (mg/L)	0.18 ± 0.08
NO_3-N(mg/L)	0.70 ± 0.21
TOC (mg/L)	702 ± 268
TN (mg N/L)	98 ± 23
TP (mg P/L)	31 ± 13
Total coli (Log/100 ml)	7.6 ± 0.3
E. coli (Log/100 ml)	6.9 ± 0.3

**: Data are average with standard deviation. Sample replications ranged between 30 and 50.*

2.2.The aerobic baffled bioreactor (BBR)

The baffled bioreactor (BBR) used for this experiment is a containerized unit that was constructed inside a Tricon shipping container [59]. A Tricon is defined as one-third of a standard 20-foot shipping container. The BBR contains five main treatment processes/operations including: anoxic, aerobic, internal settler, post-aeration, and a final clarifier (Figure 7). The BBR is designed as a pre-anoxic denitrification process and brings in considerable energy and chemical cost savings, especially by eliminating pumping via mixing using baffles [60]. The use

of nitrate (NO_3) in the oxidation of inflowing BOD_5 and production of alkalinity in the anoxic tank reduces the costs associated with aeration and bicarbonate or carbonate addition to adjust the pH in aeration tank [4].

2.3.Operational conditions

The BBR has been designed to treat wastewater with compositions ranging from greywater to blackwater. Three different feed amounts (3000, 3750, and 4500 liters per day (LPD)) were tested to evaluate the overall treatment performance of the BBR on blackwater. After stabilization of the biological process, the experiment durations were 48, 19, and 10 days for 3000, 3750, and 4500 LPD, respectively. The hydraulic retention time (HRT) varied between 1.7 and 2.6 days (the volume of the BBR = 7950 liters). The continuous aeration maintained the dissolved oxygen concentration in the aeration tank above 6 mg/L during the tests for all three feed amounts.

2.4.Chemical analysis

Wastewater samples were collected daily using 1 L Nalgene bottles from the influent and effluent streams. Samples for total coliform and *Escherichia coli* analyses were collected using sterilized sample containers (250 mL, Nalgene). All parameters used for the characterization of wastewater were completed immediately after their transfer to the laboratory. Total solids (TS) and total suspended solids (TSS) concentrations were measured using the standard gravimetric method (Method 2540 B &D) from Standard Methods for the Examination of Water and Wastewater [19]. Turbidity was measured using the nephelometric method (Method 2130) [19] with a portable turbidimeter (HACH, 2100Q). The concentration of chemical oxygen demand (COD) and total organic carbon (TOC) was analyzed using a wet oxidation-colorimetric method based on standard Method 5520-D and 5310 respectively [19] and kits (HACH) were used for

the measurement. All nutrients (TN, TKN, TP, NH_3-N, NO_3-N, NO_2-N) were measured using colorimetric methods using HACH kits prepared based on Standard Methods for the Examination of Water and Wastewater analyses [19]. Five-day BOD_5 tests were carried out based on the respirometry technique using BODTrakII Respirometric BOD apparatus and a fresh seed was collected from an activated sludge process in Delhi WWTP (Holt, MI) for every measurement. Total coliforms and E-coli were detected using the membrane filter technique (Method 9222) [19] in a biosafety cabinet with laminar flow. All wet oxidation reactions were carried out in a digester (HACH DRB200) and colorimetric measurements were fulfilled by a spectrophotometer (HACH DR3900). Samples for microbial analysis were stored at -20 ºC until they were analyzed.

2.5. Microbial community analysis

Microbial community samples (1.5 mL) were collected once per week throughout the study and stored at -20°C until DNA extraction. The samples were centrifuged using an Eppendorf 5416R centrifuge at 10,000 rpm for 5 min and the supernatant was discarded. The remaining pellet was washed by resuspension in deionized water, and the supernatant was discarded after centrifugation. The pellet was then used for DNA extraction with a DNeasy® PowerSoil® DNA Isolation Kit (Qiagen, Germany). DNA extracts were eluted with 100 µL of 10 mM Tris-HCl (pH 8.5) and the concentration and purity were determined using a NanoDrop Lite spectrophotometer (Thermo Fisher Scientific, USA). Extracted DNA samples were stored at -80°C before their use in PCR amplification and high-throughput sequencing (Illumina MiSeq flow cell).

Illumina sequencing was performed for the 16S rRNA gene region to assess the bacterial community. The PCR conditions for amplification were as follows: 1.0 µL DNA template (10x

diluted of microbial community DNA), 0.5 μL of 100 μM forward primer (IDT, Pro341F 5'-CCTACGGGNBGCASCAG-3'), 0.5 μL of 100 μM reverse primer IDT, Pro805R 3'-GACTACNVGGGTATCTAATCC-5'), 12.5 μL 2x Supermix (Invitrogen, USA), and 10.5 μL PCR grade water. The PCR program used for all assays was as follows: 96°C for 2 min, followed by 30 cycles of 95°C for 20 s, 52°C for 30 s, and 72°C for 1 min, and a final elongation period of 72°C for 10 min. Amplicons were quality-tested and size-selected using gel electrophoresis (1.0% (w/v) agarose concentration and 1× TAE run buffer). Samples were then diluted to normalize DNA concentrations within 5-10 ng μL^{-1} by measuring the DNA concentration with the PicoGreen® dsDNA quantitation assay (Invitrogen, USA) and Fluostar Optima microplate reader (BMG Labtech, Germany). The normalized PCR products were then sequenced at the Michigan State University (MSU) Research Technology Support Facility (RTSF). Illumina MiSeq (pair-end 250 bp) targeting on V3_V4 hypervariable regions was used to carry out the sequencing. Fastq files from the high-throughput sequencing were analyzed using the QIIME2 database to generate taxonomic/phylogenetic data for statistical analysis [61].

2.6. qPCR of identifying nitrifiers and denitrifiers

AOB-*amoA* (with the primers of amoA-1F and amoA-2R) and *nirK* (with the primers of F1aCu and R3Cu) are the significantly correlated genes for nitrifiers and denitrifiers, respectively [28]. They were selected for the identification of nitrifiers and denitrifiers in this study. The genes were quantified using a Real-Time PCR (Bio-rad® CFX Connect Real-Time PCR Detection System, Bio-Rad Laboratories, Inc. Hercules, California). The SYBR Green method was applied [28]. The concentration of sample template DNA was normalized to 5.0±0.1 ng/μL. The cycle threshold (C_t) as a relative measure of the target gene concentration was used

to compare relative abundances of nitrifiers and denitrifiers among three feed amounts. C_t level is inversely proportional to the concentration of the target gene.

2.7.Statistical analysis

All statistical analyses were performed using R statistical software (Version 3.6.3). The data with normal distribution and equal variance were analyzed using a one-way analysis of variance (ANOVA). When data violated the normality assumption and equal variance, the Kruskal-Wallis test was used. Tukey and Conover's pair-wise rank comparison post-hoc tests were used following ANOVA and Kruskal-Wallis tests, respectively. A significance value of $\alpha = 0.05$ was used for all tests.

Microbial analysis was performed using the R libraries Vegan, ggplot2, phyloseq, and MASS on taxonomic/phylogenetic data to graph the relative abundances of samples. Non-metric multidimensional scaling analysis (NMDS) was then used to correlate microbial communities and treatment performance at different feed amounts.

2.8.Mass, energy, and exergy analyses

Mass, energy, and exergy analyses were carried out based on the data from the tested operations at three different feed amounts of 3000, 3750, and 4500 LPD. Data of mass and energy flows were recorded daily and used to determine the amount of treated water per day and energy consumption required for the treatment.

The mass and energy balance data along with characteristics of blackwater and treated water under different feed amounts was also used to carry out the exergy analysis. The following assumptions were applied to calculate exergy flow rates [62]: 1) the processes are isothermal and isobaric; 2) the processes are at steady state; and 3) metals were not considered in the analysis.

Since the processes were isothermal and isobaric, the physical exergies of components with a similar temperature to the reference environment were negligible in comparison with chemical exergy rates. The exergy flow rates of individual compounds in the blackwater and treated water were only based on their chemical exergy:

$$\boldsymbol{B_k = \frac{m_k \cdot b_k^{ch}}{86400}} \qquad \textbf{Equation 3}$$

where B_k is the process exergy rate (W) of the kth component, k is the kth component in the process, m_k is the mass flow rate (kg/day) of the kth component, b_k^{ch} is the specific chemical exergy (kJ/g or kJ/mol) of the kth component, and 86,400 is the conversion factor of seconds in a day. The specific chemical exergy values of organic matter (based on COD), total nitrogen (TN), and total phosphorous (TP) are 13.6 kJ/g COD, 322.1 kJ/mol nitrogen (based on N in ammonia), and 134.1 kJ/mol phosphorous (based on P in phosphate), respectively [62], which will be used to calculate process exergy rates.

The universal exergy efficiency (η) was calculated as the total exergy output (B_{total}^{out}, W) divided by the total exergy input (B_{total}^{in}, W):

$$\boldsymbol{\eta = \frac{B_{total}^{out}}{B_{total}^{in}} \times 100\%} \qquad \textbf{Equation 4}$$

Where B_{total}^{out} and B_{total}^{in} are defined as follows:

$$\boldsymbol{B_{total}^{out} = B_{Organic\ matter}^{out} + B_{TN}^{out} + B_{TP}^{out} + B_{Organic\ matter\ in\ the\ sludge}^{out} + B_{TN\ in\ the\ sludge}^{out} + B_{TP\ in\ the\ sludge}^{out}} \qquad \textbf{Equation 5}$$

$$\boldsymbol{B_{total}^{in} = B_{Organic\ matter}^{in} + B_{TN}^{in} + B_{TP}^{in} + B_{Electricty\ for\ the\ feeding\ pump}^{in} + B_{Electricity\ for\ the\ treatment}^{in}} \qquad \textbf{Equation 6}$$

where $B^{out}_{Organic\,matter}$ is the exergy rate (W) of the organic content (COD) in the treated water, B^{out}_{TN} is the exergy rate (W) of the TN content in the treated water, B^{out}_{TP} is the exergy rate (W) of the TP content in the treated water, $B^{in}_{Electricity\,for\,the\,treatment}$ is the exergy rate (W) of electricity consumption of the treatment including aeration and control unit, $B^{in}_{Electricity\,for\,the\,feeding\,pump}$ is the exergy rate (W) of electricity consumption of the feeding pump and timer, $B^{out}_{Organic\,matter\,in\,the\,sludge}$ is the exergy rate (W) of the organic content in the sludge, $B^{out}_{TN\,in\,the\,sludge}$ is the exergy rate (W) of the TN content in the sludge, and $B^{out}_{TP\,in\,the\,sludge}$ is the exergy rate (W) of the TP content in the sludge.

Exergy destruction or irreversibility (*I*) during the process was defined as:

$$\boldsymbol{I = B^{in}_{total} - B^{out}_{total}} \quad \textbf{Equation 7}$$

The detailed calculation of the inputs and outputs is presented in Table S3.

2.9.Economic analysis

In addition to technical robustness, economic performance is another important factor in determining the viability of the system. An economic assessment was therefore conducted for the treatment system. The capital expenditure (CapEx) and operational expenditure (OpEx) of the operation were used for the economic assessment. A lifetime of 20 years was set for the unit. The Modified Accelerated Cost Recovery System (MACRS) was used to calculate the annual depreciation of CapEx. The MACRS annual depreciation rates are 0.100, 0.188, 0.144, 0.115, 0.092, 0.074, 0.066, 0.066, 0.065, 0.065, 0.033, 0.033 (after 10 years). Annual inflation of 3% was set for OpEx and revenues based on the five-year average inflation rate in the United States. The net cash flow based on depreciated CapEx and inflated OpEx was conducted to determine the treatment cost. A sensitivity analysis was carried out to elucidate the effects of operational

parameters on the treatment cost. Two key parameters of feed amount and energy input were investigated with 25% of their base values for the sensitivity analysis.

3. Results and discussion

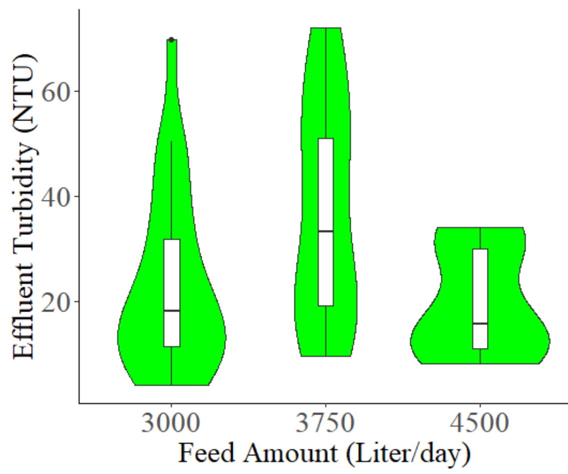
3.1. Treatment performance

3.1.1. Effluent quality from the BBR at different feed amounts

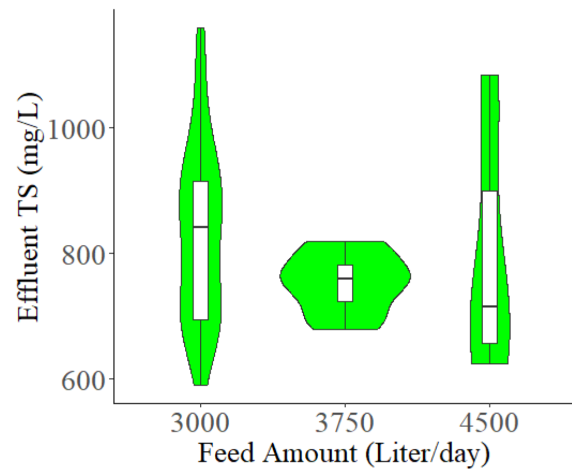
The effluent from the BBR operated with the feed amounts of 3000, 3750, and 4500 LPD was analyzed in terms of the parameters used for wastewater characterization (Figure 8 (a-m)). The results are presented using box plots with density curves (violin plots) created by R software. The plots show the data distribution around the mean value. The average values of the parameters with their standard deviations are listed in Table S4. Water quality parameters were statistically analyzed to determine any changes in the performance of the BBR as the feed amount was increased from 3000 to 4500 LPD. Normality and equal variance tests were performed on each parameter before running ANOVA and Kruskal-Wallis tests. The statistical analysis shows that there are no significant ($P>0.05$) differences between three feed amounts on turbidity, total solids (TS), total suspended solids (TSS), COD, TOC, BOD, TN, total coliform, and *E. coli* concentrations in the effluent from BBR. Average turbidity, TS, TSS, COD, TOC, BOD_5, TN, total coliform, and *E. coli* of the effluent are 26.5 NTU, 792.9 mg/L, 40.9 mg/L, 151.5 mg/L, 55.7 mg/L, 138.6 mg/L, 9.36 mg/L, 6.1 log/100 ml, and 5.1 log/100 ml, respectively, with corresponding removals of 98.0%, 57.0%, 95.9%, 94.2%, 90.9%, 92.9%, 89.7%, 1.73 log, and 1.89 log (Table 7). Similar COD and BOD_5 concentrations indicate that no recalcitrant organics were dissolved in the effluent, which is attributed to high performance of the internal settling tank of the BBR unit.

However, concentrations of nitrogen compounds (ammonia, nitrate, and nitrite) and phosphorus were significantly ($P<0.05$) influenced by feed amount. Ammonia concentrations in the effluent for 3000, 3750, and 4500 LPD were 6.74 ± 2.84, 4.96 ± 1.81, and 1.89 ± 0.90 mg/L, which were significantly ($P<0.05$) different from each other. Increasing the feed amount in the testing range certainly enhanced both nitrification (ammonia removal) and denitrification (nitrate removal). The ammonia removal was improved from 85.0% at 3000 LPD to 94.7% at 4500 LPD (Table 7). Nitrate removal was also increased from 35.3% at 3000 LPD to 46.5% at 4500 LPD (Table 7). One of the major factors increasing nitrification rate could be the amount of activated sludge in the treatment. With higher organic loading (higher feed amount), more activated sludge is produced and remains in the aeration chamber. With the unique reactor configuration of the BBR (Figure 7), retention of the activated sludge in the reactor is enhanced via recirculating the sludge back to the aeration chamber via the internal settler. The MLSS of the aeration zone was increased from 7.02 g/L at the feed amount of 3000 LPD to 11.18 g/L at the feed amount of 4500 LPD. More activated sludge means more organic carbon contents and electron donors, which can facilitate nitrate reduction [63,64]. The corresponding microbiology of nitrification and denitrification is discussed in section 3.1.2.

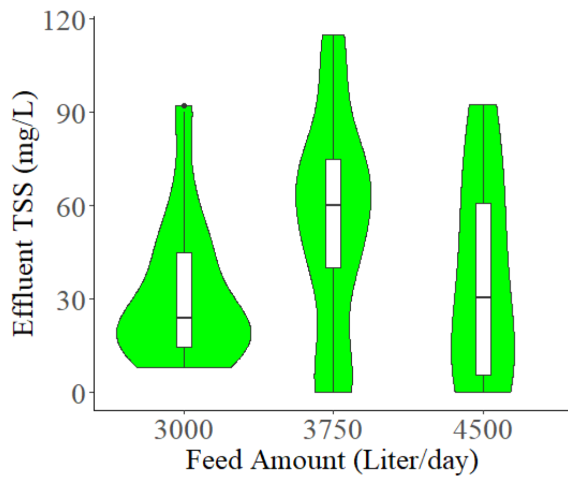
The total phosphorous (TP) results for 3000, 3750, and 4500 LPD were 2.01 ± 1.49, 1.71 ± 1.09, and 0.99 ± 0.52 mg/L, respectively, which were significantly ($P<0.05$) different from each other. Similar to ammonia removal, increasing the organic loading enhanced phosphorus removal. Phosphorous removal was increased from 92.5% at 3000 LPD to 96.1% at 4500 LPD (Table 7). This could also be attributed to the unique reactor configuration of sludge retention encouraging the biological uptake of P under higher organic loadings [65].



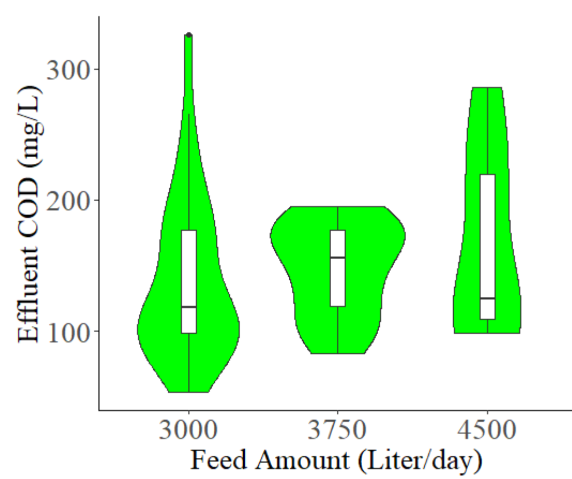
A



B



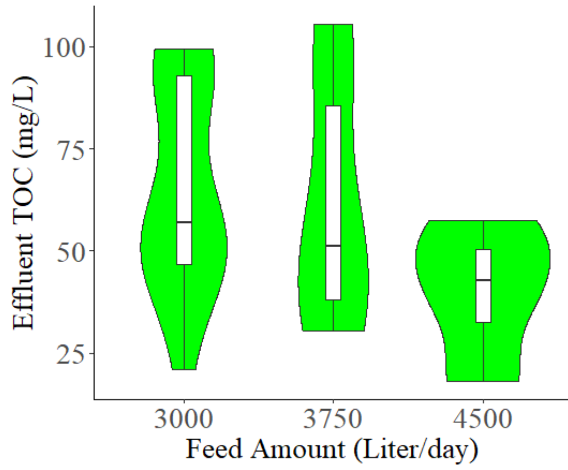
C



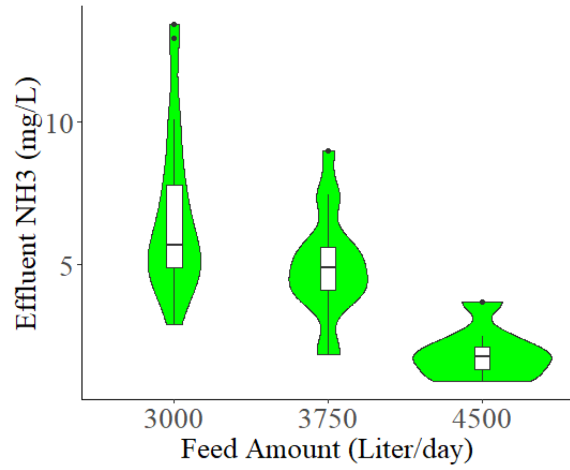
D

*Figure 8. Quality of treated water at different feed amounts *. A. Turbidity; B. TS; C. TSS; D. COD; E. TOC; F. NH3; G. Nitrite; H. Nitrate; I. TN; J. TP; K. Total coliform; L. E. coli; M. BOD5.*

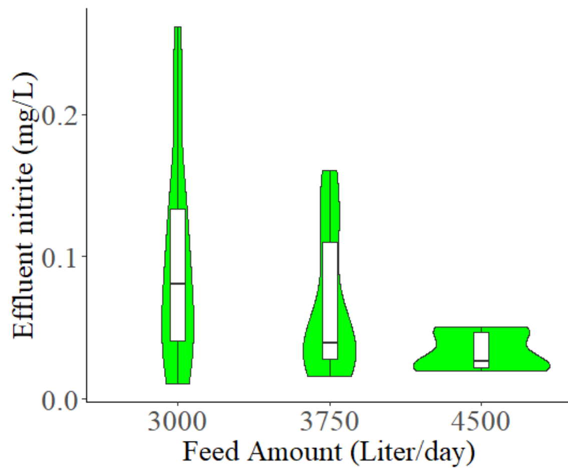
Figure 8 (cont'd)



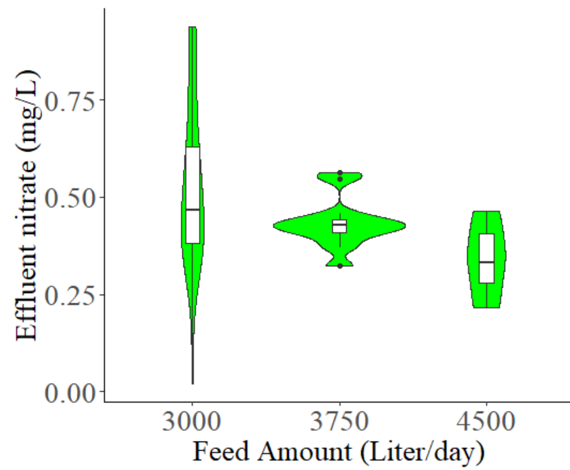
E



F

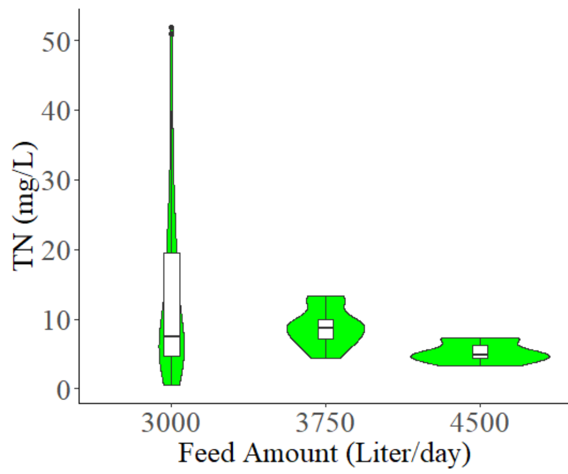


G

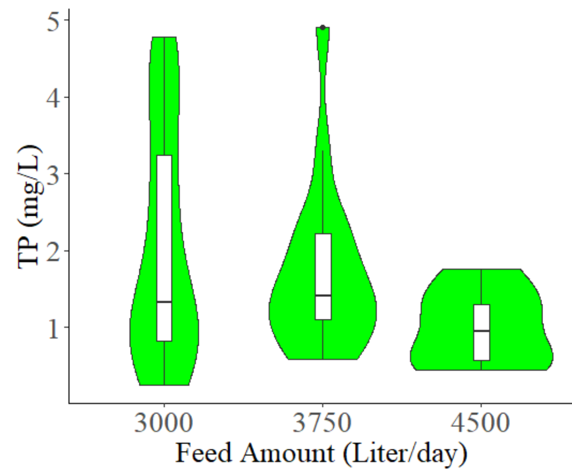


H

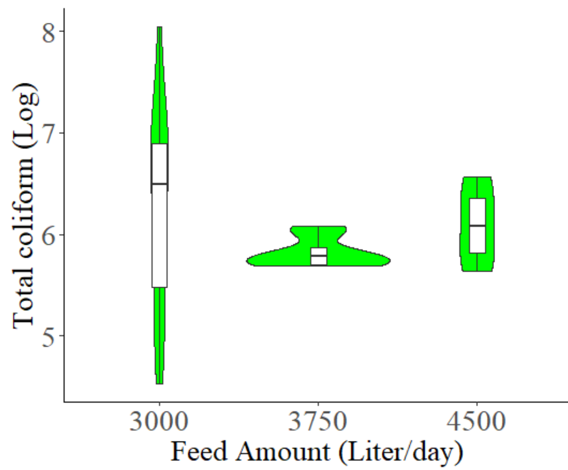
Figure 8 (cont'd)



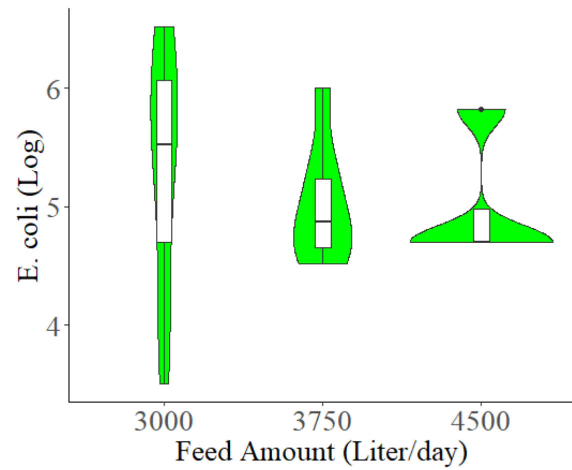
I



J

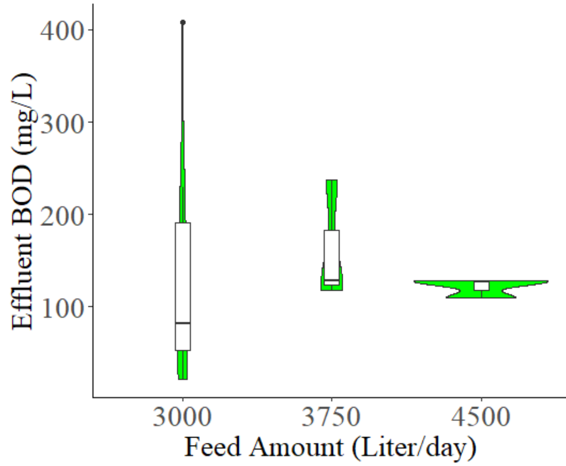


K



L

Figure 8 (cont'd)



M

Table 7. Pollutant removal percentages of the treated wastewater.

Parameter	**Treated wastewater**		
	3000 LPD	**3750 LPD**	**4500 LPD**
Turbidity (%) [a]	97.68 ± 3.95	97.54 ± 1.75	98.70 ± 0.65
TS (%) [b]	52.74 ± 10.89	61.92 ± 5.35	56.21 ± 17.79
TSS (%) [c]	96.02 ± 4.10	94.39 ± 4.57	97.23 ± 2.26
COD (%) [d]	94.09 ± 2.84	94.70 ± 1.74	93.90 ± 1.49
BOD_5 (%) [e]	91.36 ± 6.13	92.18 ± 4.27	95.13 ± 1.46
NH_3 (%) [f]	84.97 ± 7.40	86.11 ± 5.89	94.72 ± 2.09
NO_2^- (%) [g]	59.15 ± 28.46	70.25 ± 23.38	78.98 ± 6.26
NO_3^- (%) [h]	35.32 ± 20.33	36.74 ± 15.87	46.48 ± 7.62
Table 3 (cont'd). Pollutant removal percentages of the treated wastewater.			
TOC (%) [i]	89.99 ± 5.59	89.15 ± 3.74	93.51 ± 3.57
TN (%) [j]	84.10 ± 16.36	90.58 ± 3.31	94.32 ± 1.68

Table 7 (cont'd)

TP (%) [k]	92.48 ± 5.97	93.91 ± 4.45	96.05 ± 1.71
Total coliform (Log) [l]	1.08 ± 1.01	2.05 ± 0.39	2.04 ± 0.39
E. coli (Log) [m]	1.42 ± 1.04	2.03 ± 0.41	2.22 ± 0.28

a. *Turbidity data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 24, 13, and 8 samples, respectively, with standard deviations.*
b. *TS data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 36, 12, and 7 samples, respectively, with standard deviations.*
c. *TSS data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 31, 14, and 8 samples, respectively, with standard deviations.*
d. *COD data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 31, 17, and 7 samples, respectively, with standard deviations.*
e. *BOD data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 11, 3, and 3 samples, respectively, with standard deviations.*
f. *NH_3 data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 21, 17, and 8 samples, respectively, with standard deviations.*
g. *NO_2 data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 25, 17, and 5 samples, respectively, with standard deviations.*
h. *NO_3 data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 29, 13, and 7 samples, respectively, with standard deviations.*
i. *TOC data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 14, 6, and 4 samples, respectively, with standard deviations.*
j. *TN data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 37, 16, and 7 samples, respectively, with standard deviations.*
k. *TP data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 35, 17, and 7 samples, respectively, with standard deviations.*
l. *Total coliform data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 18, 8, and 4 samples, respectively, with standard deviations.*
m. *E. coli data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 13, 8, and 4 samples, respectively, with standard deviations.*

The effluent quality at different feed amounts shows that the removal of solids (TSS) was sufficient to meet requirements of the federal secondary treatment regulation, while biodegradable organics concentrations were above the required concentration for BOD_5 (7-day average of 45 mg O_2/L) [66]. Since the concentrations of total COD and BOD_5 in effluent were in similar levels for all three feed amounts, removal of carbonaceous BOD_5 during the treatment needs to be further improved to meet the regulations. As for N, its removal was increased with

the increase in feed amount and maintained at a high level for all three feed amounts tested (Table 7). In addition, TP content in the effluent is a key parameter in controlling eutrophication in water resources. The data indicated that TP removal was more than 90% regardless of different feed amounts. Moreover, the total coliform and *E. coli* were monitored, and there was a significant improvement in the *E. coli* removal when the feed amount was increased from 3000 to 3750 LPD.

3.1.2. Microbial community during treatment

The results of the treatment performance show that nitrification and denitrification were significantly influenced by a change in feed amount. To better understand the effects of different feed amounts on the black water treatment, the relationship between microbial community and treatment performance was studied.

The 16S rRNA gene sequencing result shows that the reads of gene sequences in a sample ranged from 1675 to 3996 (Figure 9 and Table 8). The sequences were rarified at 3990 reads. The numbers of sequenced microbial species stabilized after sampling 1,500 sequences for all samples, which demonstrates good sample coverage. The rank abundance analysis concludes a richness of approximately 300 species (Figure 9). Statistical analysis on diversity and evenness of microbial communities concludes that feed amount had a significant ($p<0.05$) influence on diversity (Shannon's index, H) and evenness (Pielou's index, J) among all samples (Table 9, Figure 10). The microbial diversity results demonstrate that the feed amount influenced treatment performance through changes of both the evenness and diversity of the microbial community. Both H and J of microbial communities were significantly ($P<0.05$) increased with the increase of feed amount, which means that significantly ($P<0.05$) more microbial species were evenly distributed in the communities with higher feed amounts.

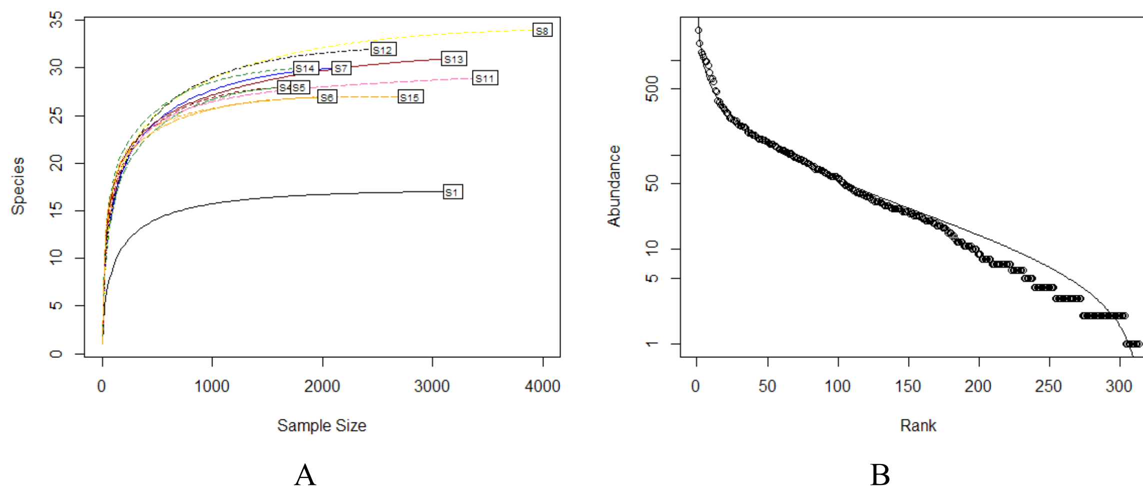
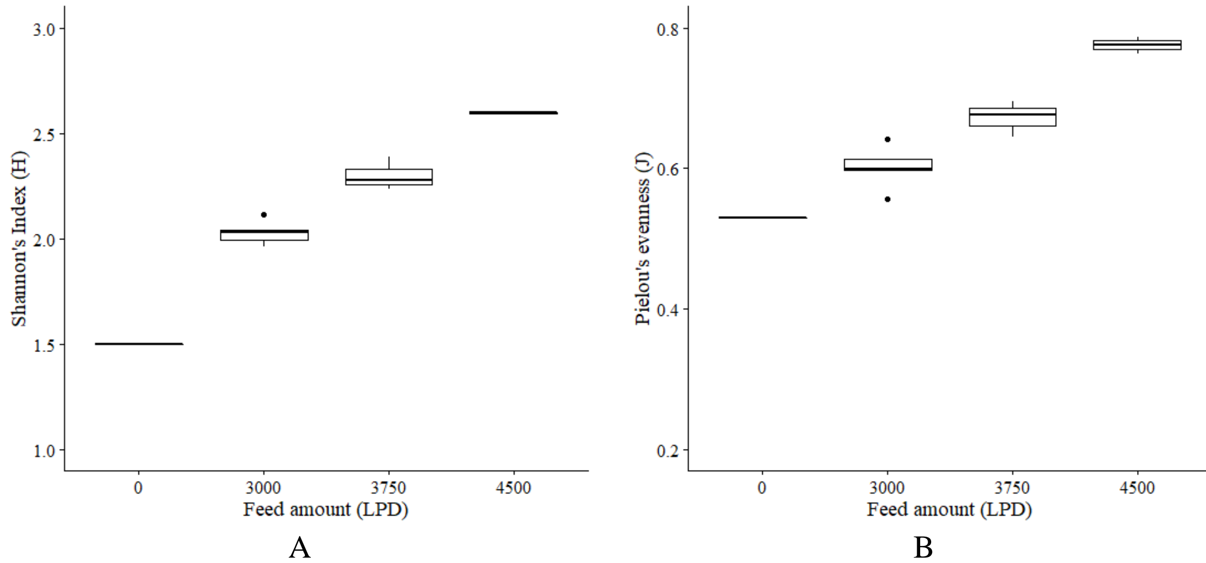


Figure 9. Rarefaction and rank abundance. A. Rarefaction curves for gene sequences of all samples; B. Rank abundance.

Table 8. Diversity and evenness of microbial communities.

Sample ID	N[a]	Frequency [b]	H[c]	J[d]
Blackwater	3182	17	1.500365	0.529563
3000 LPD at day 20	1675	28	1.996186	0.599059
3000 LPD at day 27	1795	28	2.043491	0.613255
3000 LPD at day 30	2045	27	2.115112	0.641753
3000 LPD at day 31	2171	30	2.033089	0.597757
3000 LPD at day 34	3996	34	1.964235	0.557015
3750 LPD at day 50	3479	29	2.277383	0.676324
3750 LPD at day 51	2551	32	2.237085	0.645486
3750 LPD at day 58	3192	31	2.387105	0.695141
4500 LPD at day 73	1849	30	2.597247	0.763627
4500 LPD at day 74	2798	27	2.594241	0.787127

[a] *N: total 16S rRNA gene sequences in the samples.*
[b] *Frequency: numbers of observed frequency.*
[c] *H: Shannon's index which indicates the diversity of the microbial community.*
[d] *J: Pielou's index which indicates the evenness of the microbial community.*



A B

Figure 10. Diversity and evenness of microbial community under different feed amounts. A. Shannon's index; B. Pielou's index.

Table 9. One-way ANOVA of feed amount on diversity and evenness of microbial communities.

Parameter		HRT
H	Degree of freedom	4
	Sum square	0.8632
	F value	7.076
	P	0.00569 *
J	Degree of freedom	1
	Sum square	0.04795
	F value	3.907
	P	0.0366 *

"" means significant difference.*

A dendrogram was generated to determine the similarity of microbial communities across all samples (Figure 11). The first and second separation of clades shows a clear sign of community shift regarding the change of feed amounts. Communities in all three feed amounts are different from the microbial community in the blackwater, and the communities in the higher feed amount (4500 LPD) show differences from those in lower feed amounts (3000 and 3750 LPD). The dendrogram demonstrates that feed amount changed microbial communities and led to different treatment performances (Table 7).

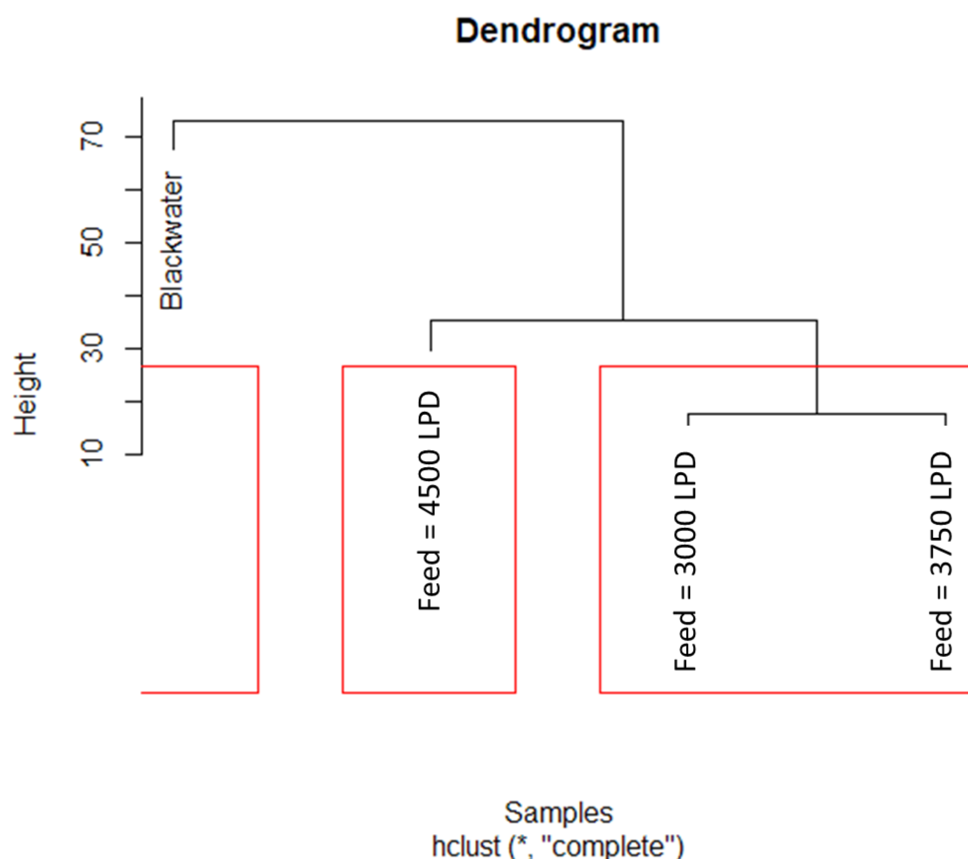


Figure 11. Dendrogram of microbial communities between different feed amounts.

A total of 49 bacterial genera were identified in the samples from the treatment (Table S5). Predominant phyla in the raw blackwater were mainly *Bacteroidetes* and *Proteobacteria* with relative abundances of 43.75% and 54.21%, respectively (Table S6 and Figure 13). Data also showed that feed amount significantly ($P<0.05$) changed microbial communities (Figure 13 and Table S6). The abundances of *Bacteroidetes* (17-21%) and *Proteobacteria* (30-47%) were reduced during the treatment compared to the blackwater. Unclassified *Bacteria* (18.5-42.6%) and *Verrucomicrobia* (6.1-11.2%) were enriched during the treatment (Table S6 and Figure 13).

The phylum *Proteobacteria*, one of the most abundant microbial groups in the blackwater and treatment, includes species from the families of unclassified *Proteobacteria* (in phylum Proteobacteria), unclassified *Rhizobiales* (in the order *Rhizobiales*), unclassified

Sphingomonadales (in the order *Sphingomonadales*), unclassified *Betaproteobacteria* (in the class *Betaproteobacteria*), unclassified *Burkholderiales* (in the order *Burkholderiales*), unclassified *Gammaproteobacteria* (in the class *Gammaproteobacteria*), and *Xanthomonadaceae* (Figure 13C). Among them, unclassified *proteobacteria* were the dominant proteobacteria family in the blackwater feed (47.8% of relative abundance). However, the abundance of the unclassified *proteobacteria* (1.95 – 2.95%) was significantly ($P<0.05$) reduced in the treatment (Figure 13C). Unclassified *Rhizobiales*, unclassified *Sphingomonadales*, and unclassified *Burkholderiales* became dominant proteobacteria families during the treatment with an increase of feed amount. At the feed amount of 4500 LPD, the corresponding abundances of these three families are 10.1, 11.4, and 15.8%. Species in the orders of *Rhizobiales* and *Sphingomonadales* are known to use different and complex carbon sources, such as polymers, chloro- and nitro-phenolic compounds, polyacrylamides, quaternary ammonium alcohols, in aerobic conditions during the oxidation of ammonia nitrogen [67,68,69]. It has also been reported that many species in order *Burkholderiales* have strong denitrifying activity [70,71]. The qPCR data shows that there are no significant differences ($P>0.05$) in relative concentrations (C_t) of *amoA* and *nirK* genes between different feed amounts (Figure 12). This means that relative abundances of nitrifiers and denitrifiers in Phylum *Proteobacteria* were not different between different feed amounts. However, MLSS data showed that the amount of activated sludge was increased with an increase in the feed amount. More bacterial biomass in the higher feed amounts means higher amounts of nitrifiers and denitrifiers in the treatment. Therefore, changes of these proteobacterial microbes and genes match the performance data that ammonia, nitrate, and nitrite were significantly ($P<0.05$) removed under higher feed amounts (3750 and 4500 LPD) (Table 10).

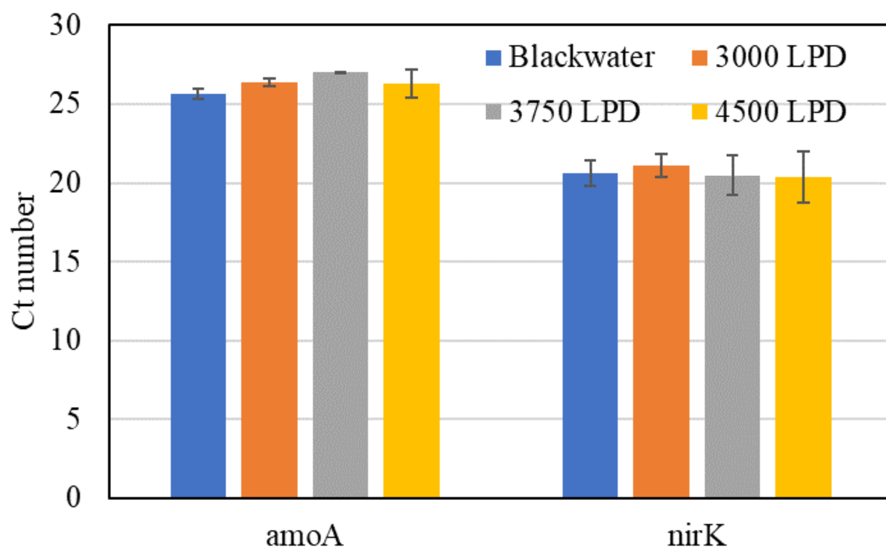


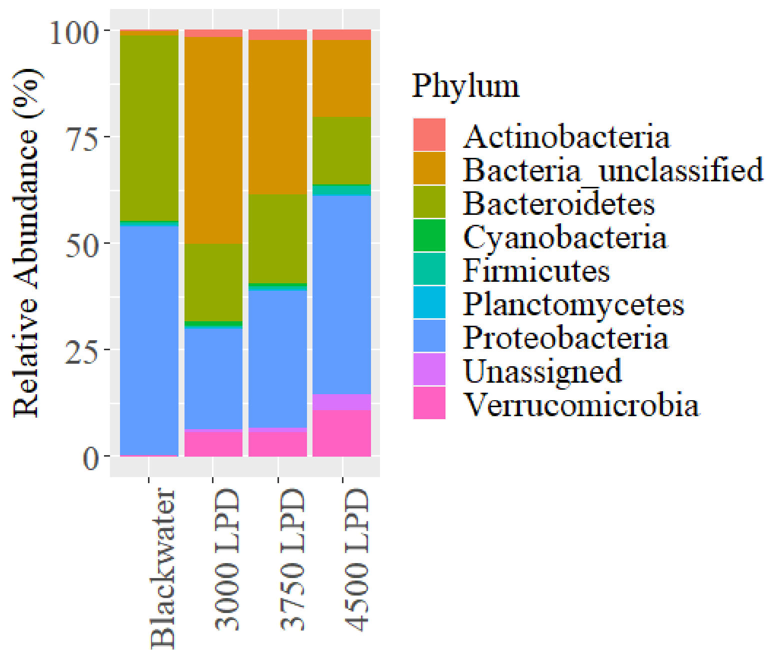
*Figure 12. Threshold cycles (Ct) of amoA and nirK genes in the activated sludges of three feed amounts. **

**: Data are averages of 2-6 replicates with standard deviation.*

The phylum *Bacteroidetes* is another abundant microbial group in the blackwater. Three dominant families of unclassified *Bacteroidetes* (in the phylum *Bacteroidetes*), *Flavobacteriaceae*, and *Chitinophagaceae* were determined in the samples (Figure 13A). Microbes in the phylum *Bacteroidetes* are primarily responsible for degrading carbohydrates in wastewater. Similar to the phylum *Proteobacteria*, there were more *Bacteroidetes* in the blackwater (43.8%) than in the treatment (15.8 – 20.8%) (Figure 13B). During the treatment, abundances of both unclassified *Bacteroidetes* and *Chitinophagaceae* were decreased with the increase in feed amount, while the abundance of *Flavobacteriaceae* was increased with higher feed amounts. As a filamentous bacterium, a high abundance of *Flavobacteriaceae* could cause the issue of sludge bulking [72]. The accumulation of *Flavobacteriaceae* increased with organic loading (with correspondingly increased carbohydrates) under high feed amounts. Even though the sludge bulking was not observed during the treatment under the feed amount of 4500 LPD,

the growth of filamentous bacteria such as *Flavobacteriaceae* needs to be closely monitored to prevent sludge bulking.

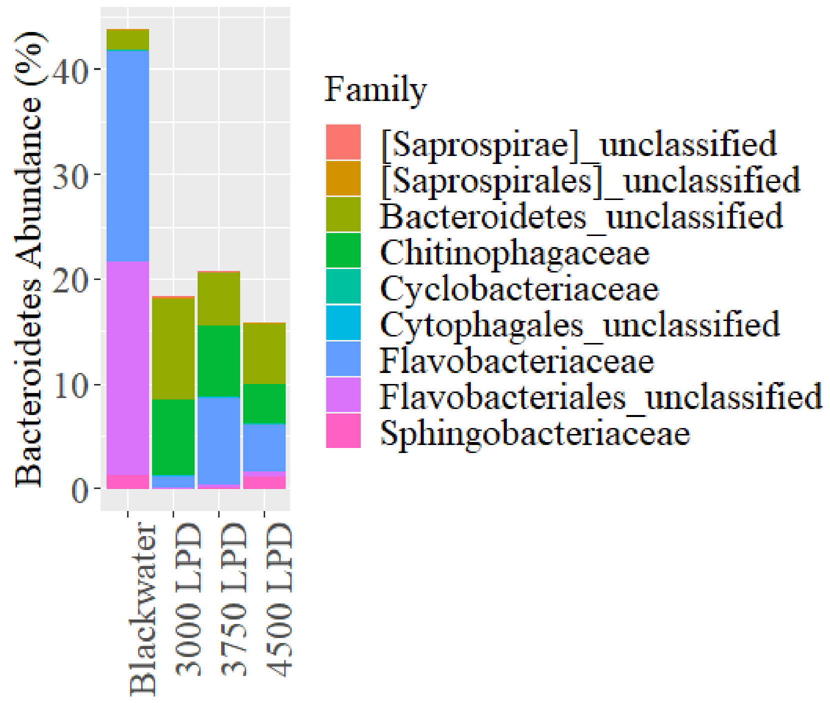
The phylum *Verrucomicrobia* was the third most abundant phylum in the treatment (6.0 – 11.2%). There are two families of unclassified *Verrucomicrobia* and *Verrucomicrobiaceae* in the treatment (Figure 13D). In contrast to *Bacteroidetes* and *Proteobacteria*, the family of *Verrucomicrobia* was not detected in the blackwater. During the treatment, the abundance of *Verrucomicrobia* was significantly (P<0.05) increased with an increase in the feed amount, particularly at 4500 LPD (11.2%) (Figure 13D). *Verrucomicrobia* is widely distributed in a wide range of ecosystems [73]. However, their functions and metabolisms are still not very clear. It has been reported that *Verrucomicrobia* can degrade carbohydrates as well as possess nitrogen fixation enzymes that may contribute to the nitrogen cycle of blackwater treatment [74].



A

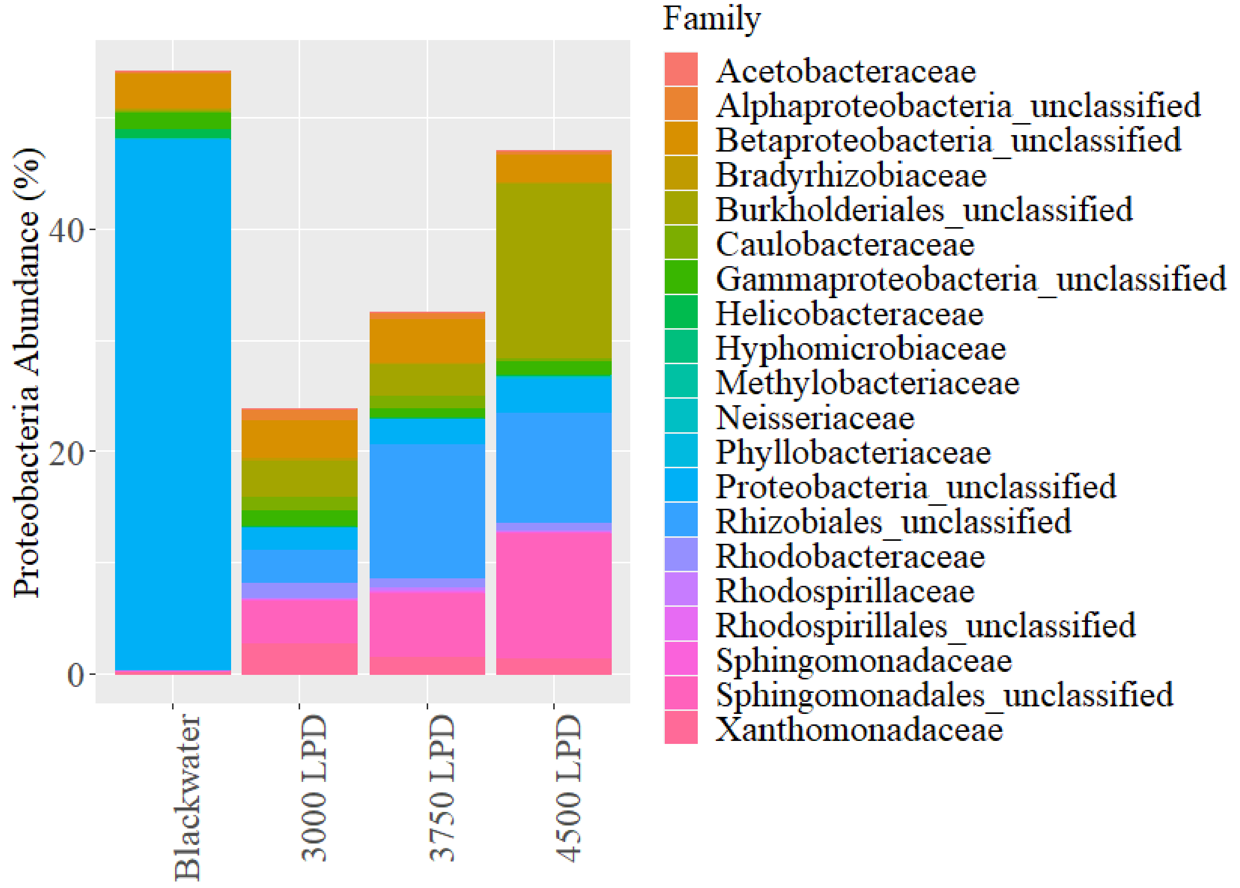
*Figure 13. Microbial communities during the treatment *. A. Phylum; B. Families in Phylum Bacteroidetes; C. Families in Phylum Proteobacteria; D. Families in Phylum Verrucomicrobia.*

Figure 13 (cont'd)



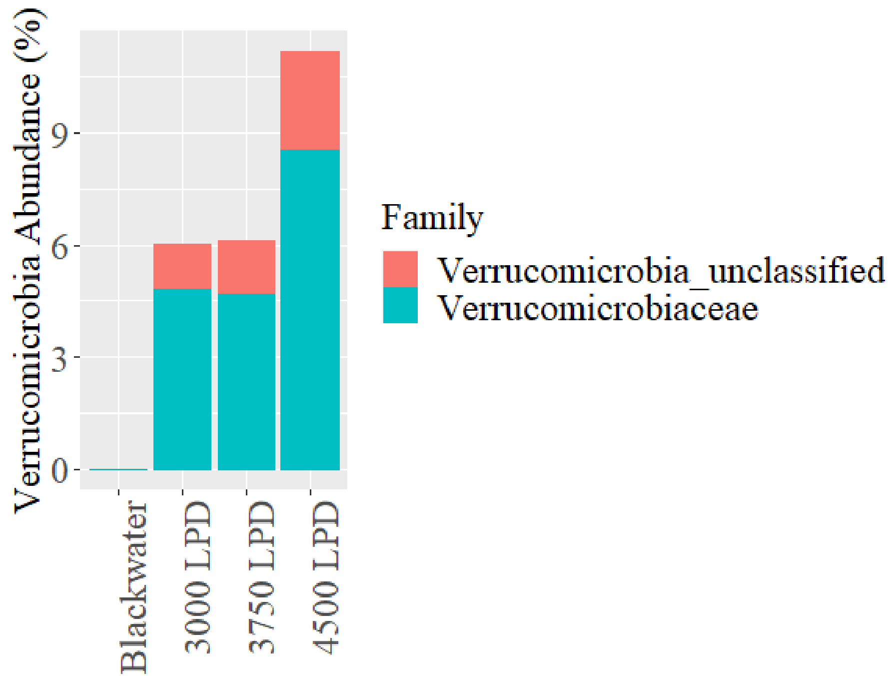
B

Figure 13 (cont'd)



C

Figure 13 (cont'd)



D

3.1.3. Relationship between microbial community and chemical parameters during the treatment

Non-metric multidimensional scaling (NMDS) analysis was applied to elucidate the dynamic relationships between microbial community, feed amount, and treatment performance (Figure 14). The results show that after 20 random runs, two convergent ordination solutions were concluded. The final stress of the best fit (best solution) between sample community distances and ordination distances was 0.098, which indicates that the ordination distances explain 90.2% of the variability in the community distance matrix. Major patterns of microbial communities in the samples were encapsulated. The permutation test of fitting the experimental conditions on the ordination indicates that the feed amount was correlated (Permutation $P<0.05$) to the community structure of the treatment samples. In addition, the permutation test of fitting

performance parameters and several key microbial communities on the ordination concludes that TS, NO_3^-, TKN, *Chitinophagaceae*, *Verrucomicrobiaceae, unclassified Sphingomonadales, unclassified Burkholderiales, and unclassified Proteobacteria* were also correlated (Permutation $P<0.05$) to the community structure (Figure 14). The NMDS analysis reveals that an increase in the feed amount enhanced the relative abundance of *Verrucomicrobiaceae, unclassified Sphingomonadales, and unclassified Burkholderiales* in the community, which also facilitated the removal of TS, TKN, and NO_3^-. As discussed in the previous section, *Verrucomicrobiaceae, unclassified Sphingomonadales,* and *unclassified Burkholderiales* are all related to nitrification/denitrification and carbohydrate degradation. The NMDS results demonstrate that the design of the reactor configuration increased the retention time of the activated sludge and further enabled and enhanced the treatment performance under higher feed amounts (higher organic loading).

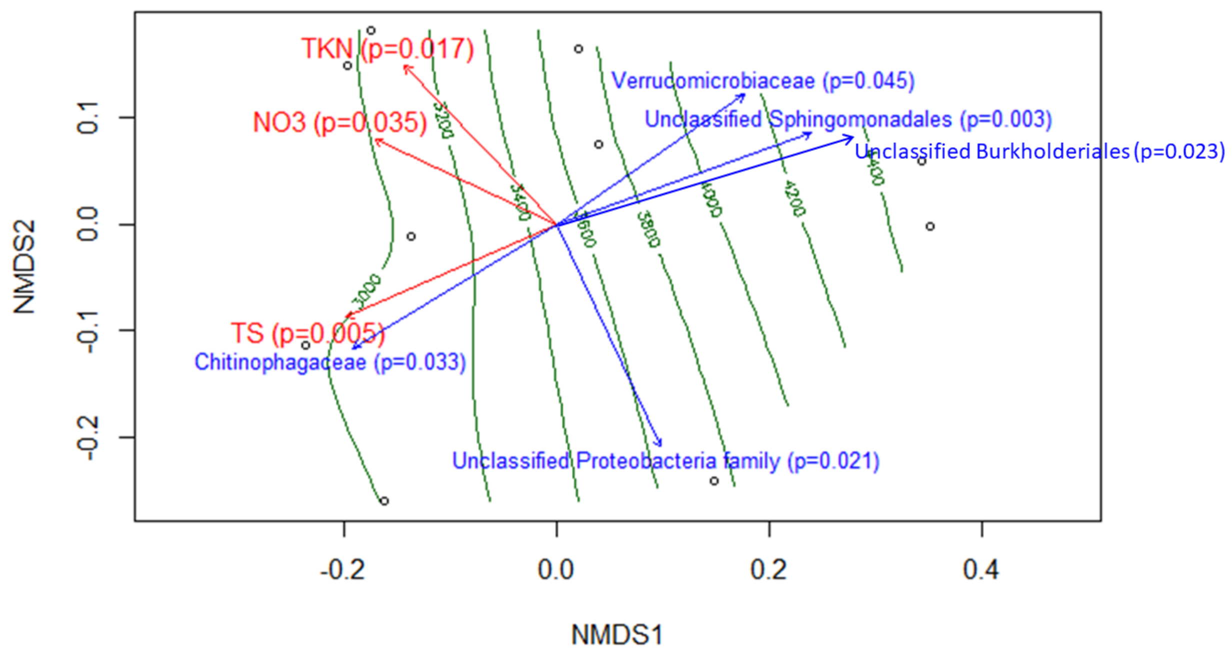


Figure 14. NMDS of microbial communities, feed amount, and treatment performance.

3.2.Mass, energy, and exergy analyses

The mass, energy, and exergy analyses were conducted to evaluate the treatment performance of the BBR. Besides the treated wastewater, activated sludge is another effluent stream, and collected and wasted from the final clarifier (Figure 7). The formation rate of sludge with a typical 95-99% H_2O content was measured as 182, 257, and 284 LPD during 3000, 3750, and 4500 LPD respectively (Table 10). Regardless of the feed amounts, 94% (v/v) of the inflowing blackwater was reclaimed as the effluent from the BBR. The energy required for the BBR operation included pumping the influent blackwater and its treatment. The combined energy requirement for the BBR was monitored and there was a declining trend in energy consumption with the increase of feed amount. Energy balance calculations resulted in the energy consumption of 6.31, 5.06, and 4.34 Wh/L for the treatment of blackwater for the feed amounts

of 3000, 3750, and 4500 LPD, respectively (Table 10). The typical energy consumption of conventional activated sludge wastewater treatment can be as high as 3.74 Wh/L for medium and large-scale treatment plants [75,76]. The studied decentralized process had higher energy consumption compared to the large-scale treatment. However, it is known that nutrient removal requires extended aeration, which increases energy consumption. Considering the high nutrient contents (3-4 times higher than regular sewage) of the blackwater, energy consumptions of the studied process based on unit nutrient removal (i.e., kWh/kg BOD removed, and kWh/kg COD removed) were much lower than large-scale sewage treatment. The mass and energy balance results show that the studied process is a comparable and efficient decentralized system to treat high-strength wastewater.

Table 10. Mass and energy balance and exergy analysis of the treatment at different feed amounts.

	3000 LPD	**3750 LPD**	**4500 LPD**
Mass balance			
Treated water (LPD) [a]	2847	3528	4259
Sludge removal (LPD) [b]	182	257	284
Energy balance			
Electricity consumption of the BBR unit (Wh/L) [c]	6.31	5.06	4.34
Electricity consumption of the feed pump (Wh/L) [d]	0.015	0.019	0.022
Exergy analysis			

Table 10 (cont'd)

Exergy rate of the blackwater (W) [e]	1417	1771	2125
Exergy rate of the electricity for the treatment and feeding pump (W) [f]	799	802	826
Exergy rate of the treated water (W) [g]	73	90	118
Exergy rate of the sludge (W) [h]	1059	1478	1654
Universal exergy efficiency (%) [i]	51	61	60
Exergy destruction during the process (W) [j]	1084	1005	1179

a. *The amount of treated water was the daily average of the treated effluent for each feed amount.*
b. *The amount of sludge removal was the daily average of the removed sludge from the BBR for each feed amount. The sludge is intended to be used on-site as an organic fertilizer and transportation of the sludge to other locations was not considered.*
c. *The electricity consumption of the BBR unit was recorded by the voltmeter.*
d. *The electricity consumption of the feeding pump was recorded by the voltmeter.*
e. *The exergy rate of the blackwater was calculated using Equation 3. Average COD, TN, and TP concentrations of the blackwater in Table 6 were used to multiply with each feed amount and corresponding special chemical exergy (Table 8).*
f. *The exergy rate of the electricity for the treatment or feeding pump was calculated using the recorded electricity consumption (Table 8).*
g. *The exergy rate of the treated water was calculated using Equation 3 again. Average COD, TN, and TP concentrations of the treated wastewater in Table S1 and Figure 8 were used to multiply with each treated water amount and corresponding special chemical exergy to obtain the exergy rate of the treated water (Table 8).*
h. *The exergy rate of the sludge was calculated using Equation 3. Average COD, TN, and TP concentrations of the sludge were used to multiply with each feed amount and corresponding special chemical exergy (Table 8).*
i. *Universal exergy efficiency (η) was calculated using Equation 4. The total exergy output (B_{total}^{out}, W) and the total exergy input (B_{total}^{in}, W) were calculated using Equations 5 and 6.*
j. *Exergy destruction during the process (I) was calculated using Equation 7.*

Energy balance analysis has shortcomings in the evaluation of efficient use of the physical resources because some of the energy is either converted or conserved during the process. The portion of the energy converted to work is called exergy. Exergy analysis has

become a benchmark study to compare the efficiencies of the wastewater treatment plants as it provides a rational basis for process optimization according to both minimum exergy destruction (better energy efficiency) and minimum exergy remained in the treated water (cleaner water) [62,77,78]. Exergy destruction (irreversibility) is calculated using exergy rates for inflows and outflows (Table 10). Three feed amounts of 3000, 3750, and 4500 LPD had exergy destruction of 1,084, 1,005, and 1,179 W, respectively. Based on the exergy destruction and other exergy values listed in Table 10, universal exergy efficiencies were calculated using the Equations from section 2.7. Universal exergy efficiency, which accounts for total mass inflows and outflows (the difference between them is the exergy destruction), increased from 51 to 61% with feed amount increasing from 3000 to 3750 LPD and did not show any considerable difference between 3750 and 4500 LPD. However, exergy rates of the treated water were increased with the increase in feed amount. The exergy rate of the treated water for 4500 LPD was 118 W, which was higher than the 73 and 90 W of the treated water for 3000 and 3500 LPD, respectively. According to the wastewater treatment performance, the preferred treatment process should simultaneously achieve both higher universal exergy efficiency (minimum exergy destruction) and lower exergy rate in the treated water. Therefore, considering mass and energy balance and exergy efficiency, it is concluded that 3750 LPD is the preferred feed amount among the tested feed amounts to treat the blackwater.

3.3. Economic analysis

Economic feasibility is another important factor that determines commercial applicability of the compact high-strength wastewater treatment. The treatment cost consisting of CapEx and OpEx, are the parameters for assessment of the economic performance. Since the tricon-based treatment unit is designed for remote areas with limited or no connection to electrical grids, on-

site electricity generation is needed to power the wastewater treatment system. Four energy case scenarios of electricity from the grid, propane gas engine for remote rural communities, diesel engine (I) using standard US market diesel fuel costs for remote rural communities and scientific research bases and military bases (not contingency operation), and diesel engine (II) using the fully burdened military cost of diesel fuel for military bases of contingency operation were selected to compare with the control being grid power supply. As presented in Table 11, the CapEx to establish the pilot unit is $172,000 with no difference between the three feed amounts due to the fact that all feed amounts are realized by the same compact wastewater treatment unit. Due to the differences in energy type and treatment application, the energy costs for individual case scenarios greatly varied from $0.10/kWh of the grid electricity to $0.82/kWh of the diesel engine II for a contingency operation. The corresponding treatment costs are changed accordingly. Under the feed amount of 3000 LPD, the treatment costs with four energy scenarios of the grid, propane gas engine, diesel engine I, and diesel engine II are $8.9, $9.8, $9.1, and $13.4 per 1000-Liter backwater (Table 11). The data clearly shows that reducing power consumption and providing a continuous power supply are critical to sustaining such an operation at a small scale. Meanwhile, compared to the treatment scenario powered by propane gas engine, the diesel engine scenario with high thermal efficiency demonstrates much less energy consumption (8% reduction of the treatment cost) than the propane gas engine, which means that a diesel engine for electricity generation is preferred to power the treatment system if available. In addition, the economic analysis concludes that increasing the feed amount from 3000 LPD to 3750 LPD and 4500 LPD could greatly reduce the treatment cost by approximately 20.0 and 33.3%, respectively

Table 11. Economic analysis of the treatment unit at different feed amounts based on different energy scenarios.

	3000 LPD				3750 LPD				4500 LPD			
Capital expenditure (CapEx)												
The baffle reactor ($) [a]	170,000				170,000				170,000			
The feeding unit ($) [b]	2,000				2,000				2,000			
Operational expenditure (OpEx) [c]												
Energy scenarios (Electricity source)	The grid	The propane gas engine	The diesel engine (I)	The diesel engine (II)	The grid	The propane gas engine	The diesel engine (I)	The diesel engine (II)	The grid	The propane gas engine	The diesel engine (I)	The diesel engine (II)
Energy consumption ($/year)	700 [d]	1,782[e]	921[f]	5,757[g]	700 [d]	1,782[e]	921[f]	5,757 [g]	700 [d]	1,782 [e]	921[f]	5,757 [g]
Maintenance ($/year) [h]	1,000				1,000				1,000			

Table 11 (cont'd)

Treatment cost ($/1000 L blackwater)	8.86	9.84	9.06	13.43	7.09	7.87	7.25	10.75	5.92	6.56	6.04	8.96

a. *The cost of the baffled reactor is based on the manufacturing cost of the unit. The costs for the diffuser and air pumps are included in the CapEx. The electricity generation unit is not included in the CapEx.*
b. *The feeding unit includes a feeding pump and a timer. The cost is based on the sale prices of the pump and timer.*
c. *The OpEx includes both energy consumption and maintenance costs.*
d. *The grid power is used for residential or small community scenarios. The electricity cost is $0.1/kWh for the grid.*
e. *The propane engine is used for remote and rural scenarios. The electricity cost is based on 30% of thermal efficiency, 87.7 MJ/gallon liquid propane of lower heating value, and $1.86/gallon liquid propane in the U.S. market.*
f. *The diesel engine (I) is also used for remote and rural scenarios. The electricity cost is based on 47% of thermal efficiency, 139.7 MJ/gallon diesel of lower heating value, and 2.40$/gallon diesel in the U.S. market.*
g. *The diesel engine (II) is used for military contingency bases and other extreme environmental scenarios. The electricity cost is based on 47% of thermal efficiency (based on the U.S. Army Advanced Medium Mobile Power Source (AMMPS), 139.7 MJ/gallon diesel of lower heating value, and $15.00/gallon diesel.*
h. *The maintenance cost is mainly for labor to clean up the BBR a few times per year, which is based on the testing operation.*

The sensitivity analysis further elucidates the economic impacts of capital expenditure and operational expenditure on the treatment cost between the four case scenarios (Table 12). For the scenarios with relatively low energy costs (the grid, propane gas engine, and diesel engineer (I)), Changing the capital expenditure (the cost of the treatment unit) would have more significant influences on the treatment cost than the operational expenditure. The data shows that a 25% change to the capital expenditure led to treatment cost changes of 16.3, 14.7, and 16.0% for the cases of the grid, propane gas engine, and diesel engine (I), respectively, which are much higher than corresponding changes (4.4, 6.4, and 4.8%) from a 25% change of operational expenditure for the same case scenarios. Reducing the cost of the treatment system could significantly improve the economic performance of these case scenarios. However, for the case scenario of diesel engine (II) with a high energy cost, the impact of operational expenditure (the energy cost) had a much larger impact (11.4% change on the treatment cost based on a 25% change of the operational expenditure) on the treatment cost than other case scenarios. In addition, the impact of the operational expenditure also exceeds that of the capital expenditure (10.8%). This result indicates that reducing energy cost is critical to sustain the treatment operation for the case using diesel engine (II) for military contingency operations. Improving energy efficiency and using on-site renewable energy (solar, wind, and bio-energy) would be potential ways to advance the treatment technology and significantly reduce the cost burden of waste transportation and logistics.

*Table 12. Sensitivity analysis of different energy scenarios on the treatment cost for the feed amount of 3750 LPD. **

Feed amount (LPD)	Energy scenario		Base value ($)	Sensitivity range (%)	Base treatment cost ($)	Change on treatment cost (%)
3750	The grid	CapEx	172,000	25	26.83	±16.3
		OpEx	1,700	25	26.83	±4.4
	The propane gas engine	CapEx	172,000	25	29.79	±14.7
		OpEx	2,782	25	29.79	±6.4
	The diesel engine (I)	CapEx	172,000	25	27.43	±16.0
		OpEx	1,921	25	27.43	±4.8
	The diesel engine (II)	CapEx	172,000	25	40.68	±10.8
		OpEx	18,271	25	40.68	±11.4

**: The other two feed amounts have the same changes on treatment cost regarding 25% changes on CapEx and OpEx.*

4. Conclusions

A decentralized blackwater treatment system based on a baffled bioreactor was comprehensively studied. The study concluded the baffled bioreactor enhanced microbial communities that facilitated removal of total solids, and inorganic and organic nitrogen. Increasing the feed amount in the range of 3000-4500 LPD improved the treatment performance. The mass, energy, and exergy analyses concluded that the feed amount of 3750 LPD is the preferred feed amount to treat the black water in a technically feasible and environmentally sound way. Treatment with a feed amount of 3750 LPD consumes 5.1 Wh/L wastewater with a universal exergy efficiency of 61%. An economic analysis further elucidated that at 3750 LPD,

the corresponding treatment costs were $7.1, $7.9, $7.3, and $10.8 per 1000 liters blackwater for four studied energy case scenarios of electricity from the grid, propane gas engine for remote rural communities, diesel engine (I) for remote rural communities and scientific research bases, and diesel engine (II) for military contingency bases and other extreme environmental scenarios.

This chapter represents published work: Thomas, Benjamin D., Uludag-Demirer S, Frost H, Liu Y, Dusenbury JS, Liao W. Decentralized High-Strength Wastewater Treatment Using a Compact Aerobic Baffled Bioreactor. J Environ Manage. 2022 Mar 1;305:114281. doi: 10.1016/j.jenvman.2021.114281. Epub 2021 Dec 26. PMID: 34965502.

CHAPTER 3: EVALUATION OF ULTRAFILTRATION MEMBRANE FOULING FOR GREYWATER RECYCLING USING A MULTIPLE-OBJECTIVE OPTIMIZATION APPROACH

1. Introduction

Freshwater resources are becoming increasingly stressed due to factors associated with climate change and drought [79]. A recent study indicates a projected 55% increase in global water demand [80], exacerbating pressure on already strained freshwater reservoirs. In response, exploring alternative sources of water becomes imperative to growing freshwater scarcity. Greywater, defined as wastewater from showers and laundries, that does not contain contributions from latrine wastewater [81], emerges as a viable resource for an alternative water source. Greywater typically contains household cleaning agents such as soaps, detergents, and other household personal care products but does not contain fecal matter. It also has lower contaminant concentrations than other wastewater types. Moreover, it constitutes a substantial portion (approximately 75%) of household wastewater. Given its relatively simpler treatment process, greywater stands out as a prime candidate for recycling [12]. By adopting greywater recycling practices, it is possible to mitigate the dependence on freshwater reserves while curbing pollution resulting from untreated greywater discharge into the environment [11].

Ultrafiltration membranes are a promising option for greywater recycling due to their operational consistency and ability to maintain water quality. Nevertheless, a significant challenge in utilizing ultrafiltration for greywater recycling is membrane fouling, which can escalate energy demand and maintenance costs [82]. Greywater contains various potential fouling agents, including organic and inorganic particulates, dissolved organic matter, salts, surfactants, and pathogens [83]. Fouling of submerged ultrafiltration membranes has been investigated in previous studies showing the complex mechanisms that greywater can pose on

membrane filtration processes [82, 83, 84, 85]. Calcium has been identified as an important multivalent cation contributing to membrane fouling during the treatment of greywater [83]. Moreover, organic matter in greywater can cause significant fouling and their concentrations in the source water correlate strongly with membrane fouling [83]. Despite numerous studies investigating fouling on submerged ultrafiltration membranes, research on the effects of greywater with spiral wound ultrafilters for direct filtration and fouling remains limited.

Therefore, a comprehensive assessment of spiral wound ultrafiltration membranes on greywater recycling can provide a better understanding of the relationship between various membrane types and greywater characteristics. Such an evaluation can yield valuable insights into optimizing operational strategies. This study aims to apply a multi-objective optimization (MOO) approach to evaluate three membranes (PPG, PVDF, and PES) in treating three different greywater sources (shower, laundry, and combined shower/laundry).

2. Materials and methods

2.1. Membranes

Three ultrafiltration membranes were selected for operation on greywater based on previous field testing and manufacturer recommendations. The membranes and their characteristics are shown in Table 13. The PPG ultrafilter was selected based on its superior performance in relevant field testing for a greywater recycling operation at a military base. Commercial PVDF and PES membranes were selected based on the manufacturer's recommendations for greywater treatment and the desire to test commercially available and conventional membranes. Cut sheet membranes were procured to fit in the Sterlitech SEPA cell.

Table 13. Membrane Characteristics.

Membrane	**Material**	**Pore Size**
PPG - UMA4040-DD1PFEM11FF	Proprietary Mixed Matrix	0.05um nominal

Table 13 (cont'd)

PVDF - Synder BY YMBY1905	PVDF (Polyvinylidene Fluoride), C2H2F2	100,000 Daltons
PES - Snyder LY YMLY1905	PES (Polyethersulfone), C12H8O2S	100,000 Daltons

2.2. Greywater sources

National Sanitation Foundation (NSF) 350 recipe waters were utilized during this study in place of real greywater. The recipe water allows for the reevaluation of each water source during testing for comparison on each membrane. NSF has created three different recipe waters for greywater which can be seen in detail in the NSF 350 document. The ingredients are shown in Tables 14, 15, and 16 for this study. The recipe waters were batched in the laboratory and mixed in a 60-gallon tank. The water was then pumped through a 5-micron cartridge filter for pre-filtration prior to sending the water to the feed tank that was used for the test. Pre-filtration was implemented to mimic the solids removal step that would occur prior to the ultrafiltration process in actual greywater recycling operations. Each of the three selected membranes was operated on each greywater source. Table 17 shows the characteristics of the raw recipe water.

Table 14. NSF 350 Shower Water Recipe.

Component	Quantity/100L	Unit
Secondary Effluent	2	L
Lactic Acid	3	g
Bodywash	30	g
Toothpaste	3	g
Deodorant	2	g
Shampoo	19	g
Conditioner	21	g
Bathroom Cleaner	10	g
Hand Soap	23	g

Table 15. NSF 350 Laundry Water Recipe.

Component	Quantity/100L	Unit
Laundry Detergent	40	mL
Fabric Softener	21	mL
Na_2SO_4	4	g
Na_2PO_4	4	g
Secondary Effluent	2	L
$NaHCO_3$	2	g

Table 16. NSF 350 Combined Shower/Laundry Recipe.

Component	Source	Quantity/100L	Unit
Laundry Detergent	L	18.8	mL
Fabric Softener	L	9.87	mL
Na_2SO_4	L	1.88	g
Na_2PO_4	L	1.88	g
Secondary Effluent	L/S	2	L
$NaHCO_3$	L	0.94	g
Lactic Acid	S	1.59	g
Bodywash	S	15.9	g
Toothpaste	S	1.59	g
Deodorant	S	1.06	g
Shampoo	S	10.07	g
Conditioner	S	11.13	g
Bathroom Cleaner	S	5.3	g
Hand Soap	S	12.19	g

Table 17. Characteristics of shower, laundry, and combined water.

	Laundry water	Shower water	Combined water
pH	7.01±0.30	7.09±0.67	7.05±0.84
Turbidity (NTU)	11.44±2.65	13.88±0.56	12.84±1.73
Conductivity (μS /cm)	503.22±68.62	387.80±7.86	425.64±56.98
COD (mg/L)	291.50±52.74	303.50±24.13	252.04±49.70
TP (mg/L)	19.16±4.02	0.97±0.15	12.82±1.01
TN (mg/L)	3.71±0.89	3.73±0.96	3.82±1.16
UV_{254}	0.21±0.05	0.12±0.01	0.18±0.03

2.3. Flat sheet test setup

A membrane flat sheet test setup was established for this study (Figure 15 and Figure 16). A 60-gallon feed tank (with a mixer) was used to batch the recipe water that feeds the flat cells. The raw water pump (Hydra-Cell M03SASGSNSCA, Wanner Engineering, Inc – Minneapolis, MN) transferred the water in series to all three of the test cells (Sterlitech SEPA CF, Sterlitech Corporation - Auburn, WA). IFM PX322X pressure sensors were utilized to measure the pressure in and out of each cell to monitor any fouling during the run. The pressure was maintained at 70 PSI for each test and the cross flow was kept at 0.16 GPM utilizing an IFM SM6601 flow meter. The flow was recorded every 10 seconds to calculate flux decline during the test. The effluent water from each cell was transferred to a 1-gallon tank that sat on top of a scale (Mettler Toledo PBA655-A6) to measure the effluent flow rate. The effluent water tanks were automatically drained back into the feed tank based on the measured weight of the full tank so that the system could be operated continuously for 48 hours. A clear acrylic cell was utilized as the second cell in the series so that imaging could be conducted. A Nikon DS-Fi3 camera was set up above the acrylic cell and time-lapse pictures were taken every hour to monitor the fouling on the surface of the membrane (Figure 16b).

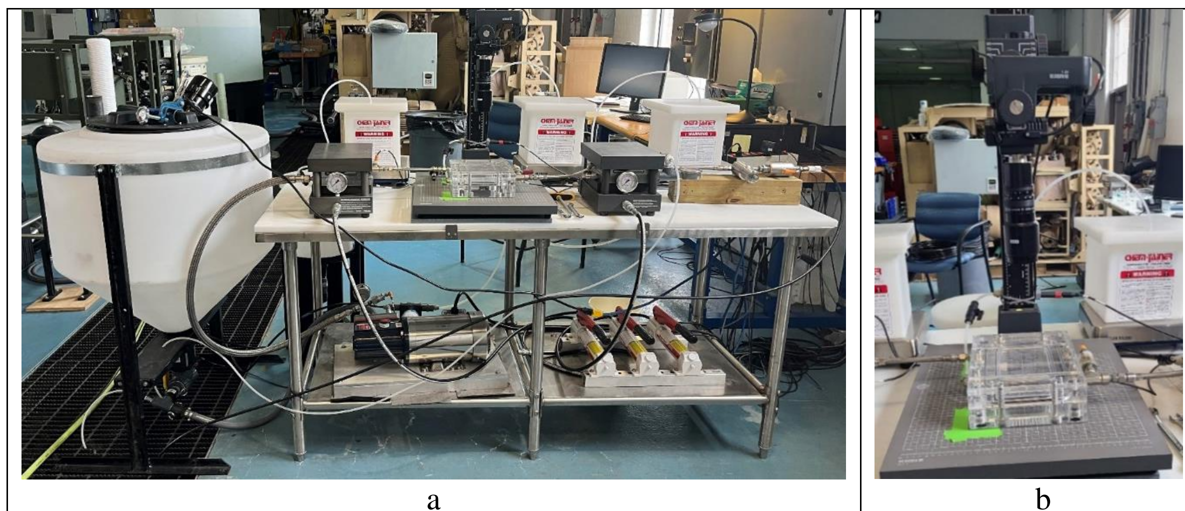


Figure 16. Flat cell test setup. (a). The flat cells. (b) the time lapse camera.

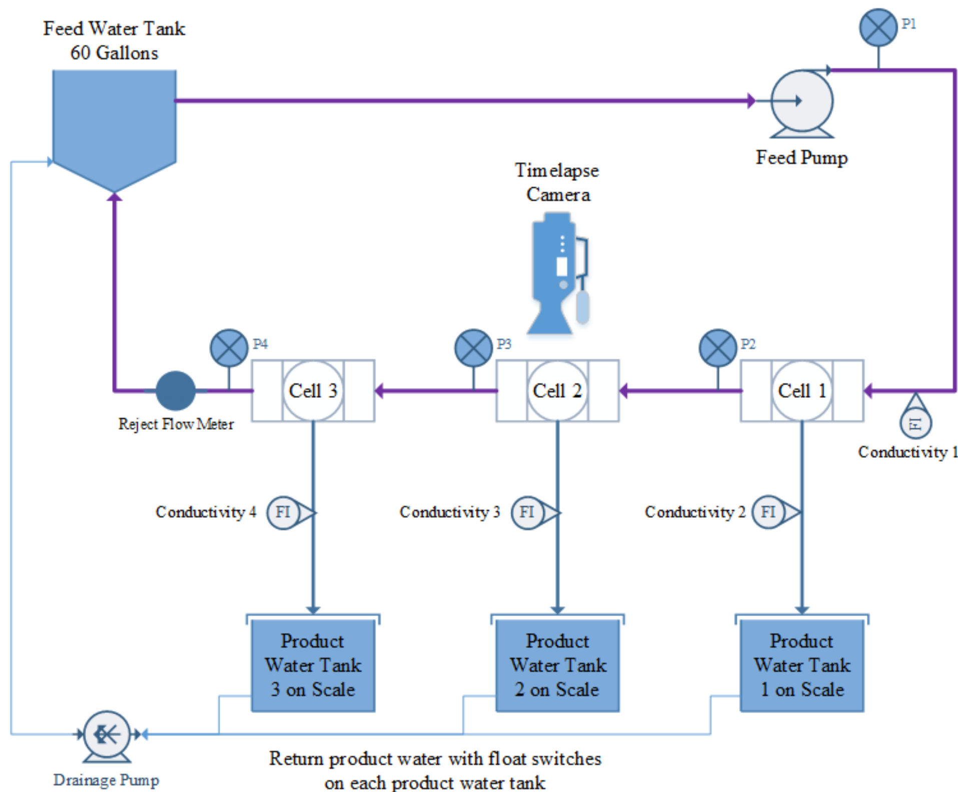


Figure 15. The flow diagram of the flat cell setup.

2.4.Water quality analysis

Wastewater samples were collected twice daily using 1 L Nalgene bottles from the influent, effluent, and reject streams. Turbidity was measured using the nephelometric method (Method 2130) [19] with a portable turbidimeter (HACH, 2100Q). The concentration of chemical oxygen demand (COD) and total organic carbon (TOC) was analyzed using a wet oxidation-colorimetric method based on standard Method 5520-D and 5310 respectively [19] and kits (HACH) were used for the measurement. All nutrients (TN and TP) were measured using colorimetric methods using HACH kits prepared based on Standard Methods for the Examination of Water and Wastewater analyses [19]. All wet oxidation reactions were carried out in a digester (HACH DRB200) and colorimetric measurements were fulfilled by a spectrophotometer (HACH DR3900). UV 254 Absorbance measurements were taken using Real Tech – REAL UV254 meter (Standard Method 5910). The pH measurements were taken

utilizing Hach PHC201 probe on the Hach HQ40d (Standard Method 4500H-B). Conductivity measurements were also taken on the Hach HQ40d utilizing the Hach CDC401 probe (EPA 120.1).

2.5. Membrane analysis

After each run was completed, the membranes were removed from the test cell and freeze-dried prior to analysis. The samples were analyzed using a JEOL 6610LV (tungsten hairpin emitter) scanning electron microscope (SEM), an Oxford Instruments Aztec system energy dispersive X-ray spectrometer, and an ATR-FTIR spectrometer (Jasco, FT/IR-660 ATR PRO ONE, Oklahoma City, OK). The X-ray spectroscopy resulted in the elemental composition of the fouling layer for each membrane. This will help determine the performance of each membrane and determine which one resulted in the least amount of fouling when operating on each greywater source. The SEM produced images of the fouling surface which can help visualize the layer. ATR-FTIR was used to determine the molecular constituents of the fouling layer. The chemical bond information produced from the FTIR analysis along with the elemental percentage from the X-ray spectroscopy will elucidate the chemical and molecular composition of the fouling materials for each membrane.

The SEM and X-ray spectroscopy samples were cut from the freeze-dried membranes and mounted on aluminum stubs using adhesive tabs (M.E. Taylor Engineering, Brookville, MD). They were then coated with osmium in a Tennant20 osmium CVD (chemical vapor deposition) coater (Meiwafosis Co., Ltd., Osaka, Japan). The SEM image was taken first followed by the X-ray spectroscopy. SEM imaging was performed at 15kV, WD11mm, SS55, and x1200 zoom. X-ray spectroscopy was performed at 15kV, WD11mm, SS55, and x40 zoom.

The FTIR analysis was performed by scraping the fouling layer from the freeze-dried membrane to collect a powder. This powder was then used to conduct the FTIR analysis. The FTIR spectra were analyzed using peak wavelengths, intensities, and broadness. These categories were then compared to a reference IR spectrum table provided by the Chemistry Department at MSU to determine the group and compound class [86].

2.6. Multiple-objective optimization

During the filtration test, the flux is often in conflict with the fouling (i.e., powder mass of fouling) and water quality of the treated water (i.e., COD, turbidity, UV254, etc.). To simultaneously optimize these conflicting criteria and select the preferred membranes that are capable of maintaining a high flux with a minimum fouling and a good treatment performance, Pareto frontier was applied to carry out multiple-objective optimization [87]. Pareto frontier is an approximation set that consists of distinct objective vectors that are nondominated by each other [88]. In this study, objective vectors include flux, powder mass, turbidity reduction, COD reduction, and UV254 reduction. Flux was paired with powder mass, turbidity reduction, COD reduction, and UV254 reduction to form four pairs of objectives (flux vs powder mass, flux vs turbidity reduction, flux vs COD reduction, and flux vs UV254 reduction) for the optimization. Visualization is one of the most effective measures for Pareto frontier optimization, and it was used to assess the quality of the approximation set. R function "psel" was used to run the optimization and to output and visualize the results.

2.7. Statistical analysis

The statistical analyses conducted for this study were performed using R software. The data with normal distribution and equal variance were analyzed using a one-way analysis of variance (ANOVA). Tukey and Conover's pair-wise rank comparison post-hoc tests were used

following ANOVA. A significance value of $\alpha = 0.05$ was used for all tests. The GGPLOT library in R was used to generate the plots in this study.

3. Results and discussion

3.1. Effluent water quality and flux

The effluent from each membrane on the three different water sources was sampled and analyzed for the parameters discussed in 2.4. The results of the sample analyses are presented in Figure 17 (a-u) using box plots with density curves (violin plots) that were created using R software. Data distribution around the mean value is shown in these violin plots. These results were also analyzed to determine if there were any significant differences between the three different membranes on the three source waters for the quality of the effluent water.

For shower water, the statistical analysis showed that there were no significant ($p>0.05$) differences between the three membranes on pH, however, the other water quality parameters showed a significant ($p<0.05$) difference on one or more membranes. UV254 and COD measurements had a significant difference on the PVDF membrane compared to PES and PPG. Turbidity measurements showed a significant difference between the PVDF and PPG membranes. Total phosphorous (TP) had a significant difference on the effluent measurements for all three membranes. Total nitrogen (TN) results saw a significant difference between the PPG and PES membranes, and the PVDF and PPG membranes. Conductivity measurements saw a significant difference on PES compared to PVDF and PPG membranes. Average effluent measurements for the PES membrane on shower water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0238, 80.1, 0.458, 0.238, 1.123, 268.3, and 6.87 respectively. The PPG membrane on shower water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0216, 84.1, 0.137, 0.964, 2.88, 325.4, and 6.91 respectively. The PVDF membrane on

shower water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0348, 198.6, 0.470, 0.444, 1.82, 316.0, and 6.99 respectively.

Statistical analysis on laundry water showed that there were significant ($p<0.05$) differences on all three membranes for UV254, COD, and Conductivity. Turbidity and total phosphorous (TP) measurements showed a significant difference on the PPG membrane compared to PES and PVDF. Total nitrogen measurements had a significant difference between PPG and PES membranes, and pH showed a significant difference between the PVDF and PES membranes. Average effluent measurements on laundry water for the PES membrane for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.055, 121.5, 0.629, 9.5, 1.616, 259, and 6.99 respectively. The PPG membrane on laundry water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0295, 259.17, 0.196, 20.3, 2.984, 365.6, and 7.11 respectively. The PVDF membrane operating on laundry water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.081, 175.86, 0.881, 11.1, 1.799, 451.5, and 7.31 respectively.

For combined shower and laundry water, the statistical analysis showed that there were significant differences on UV254, COD, and Turbidity for the PES membrane compared to both the PVDF and PES membranes. Conductivity and total phosphorous showed a significant difference on the PPG membrane compared to PES and PVDF. PH and total nitrogen showed a significant difference on the PVDF membrane compared to PES and PPG. Average effluent measurements on the combined shower and laundry water source with the PVDF membrane for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0394, 130.87, 0.38, 6.94, 2.69, 272.9, and 7.21 respectively. The PPG membrane on the combined shower and laundry water for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0785, 187, 1.35, 10.98, 3.44, 418.5, and 7.05 respectively. The PVDF membrane on the combined shower and laundry water

for UV254, COD, Turbidity, TP, TN, Conductivity, and pH were 0.0628, 166, 1.635, 6.66, 1.72, 277.9, and 6.76 respectively.

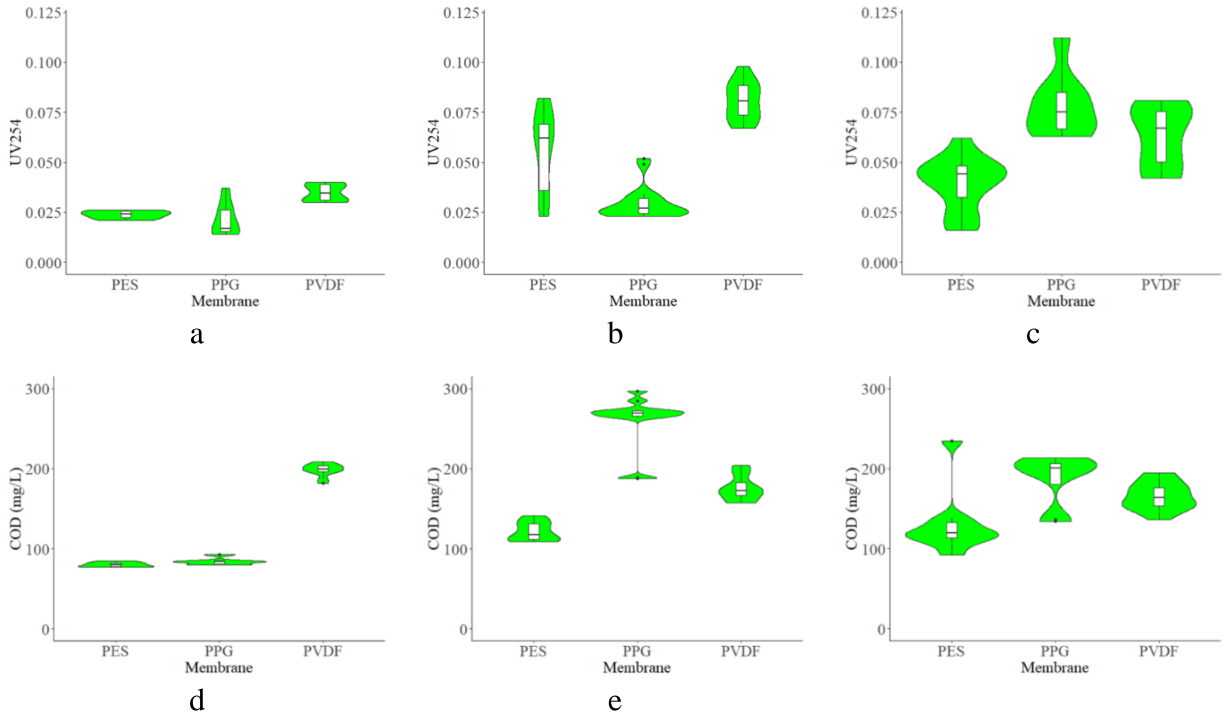
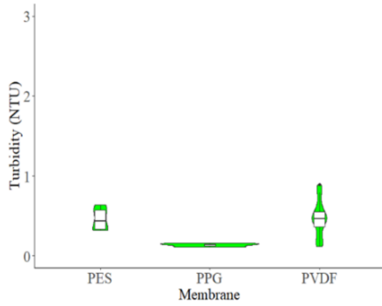


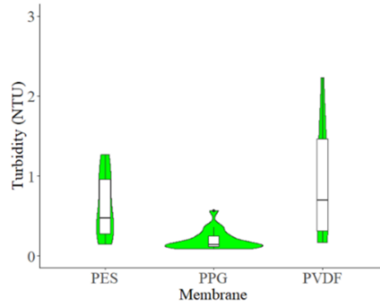
Figure 17. Treatment performance of three membranes on different wastewaters.

(a). UV254 of treated shower water; (b). UV254 of treated laundry water; (c). UV254 of treated shower/laundry water; (d). COD of treated shower water; (e). COD of treated laundry water; (f). COD of treated shower/laundry water; (g) Turbidity of treated shower water; (h) Turbidity of treated laundry water; (i) Turbidity of treated shower/laundry water; (j) TP of treated shower water; (k) TP of treated laundry water; (l) TP of treated shower/laundry water; (m) TN of treated shower water; (n) TN of treated laundry water; (o) TN of treated shower/laundry water; (p) Conductivity of treated shower water; (q) Conductivity of treated laundry water; (r) Conductivity of treated shower/laundry water; (s) pH of treated shower water; (t) pH of treated laundry water; (u) pH of treated shower/laundry water.

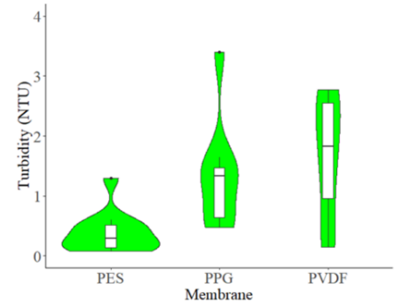
Figure 17 (cont'd)



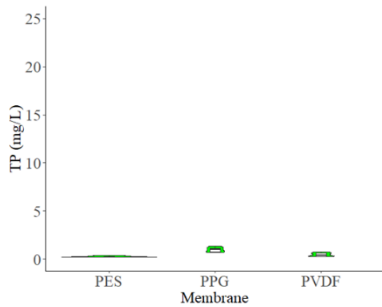
g



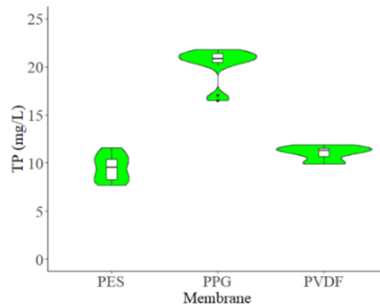
h



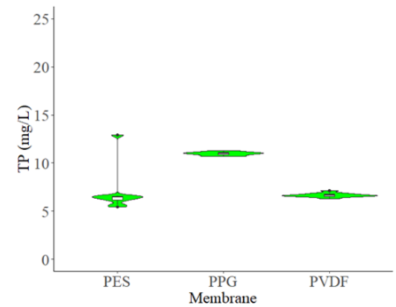
I



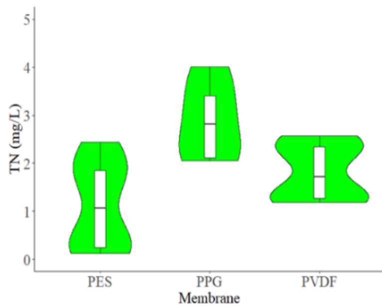
j



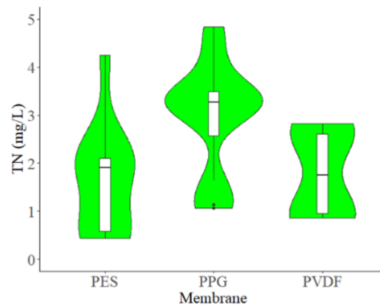
k



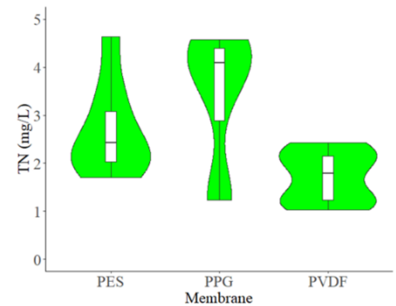
l



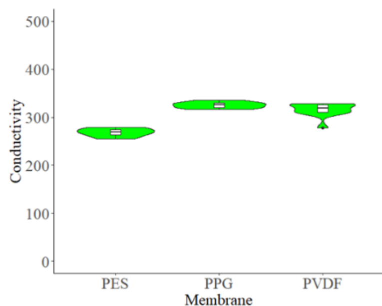
m



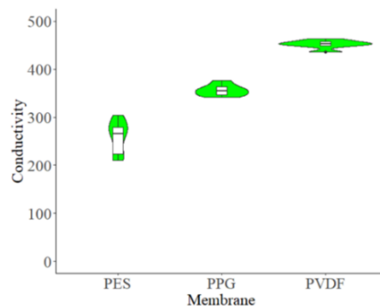
n



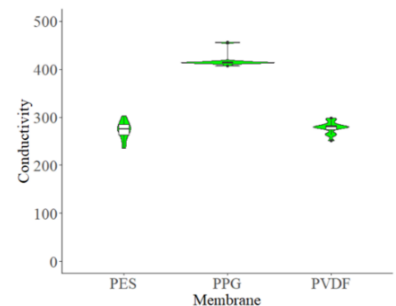
o



p

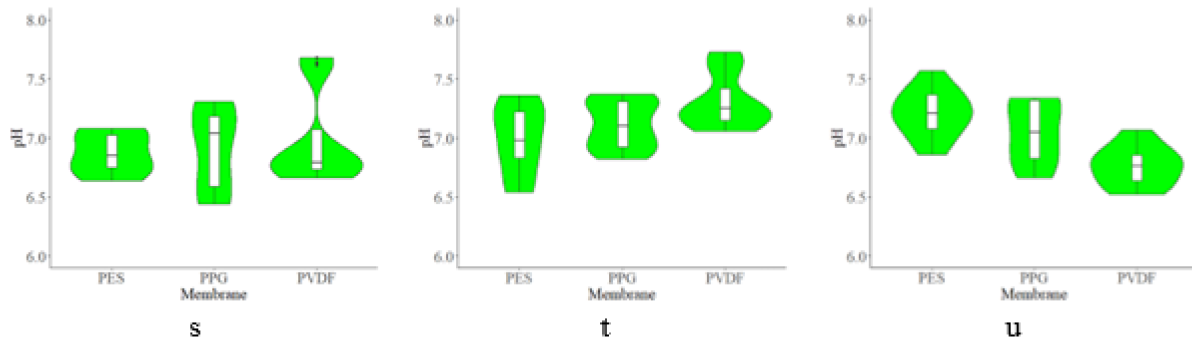


q



r

Figure 17 (cont'd)



The three different membranes operated during this study also resulted in significant differences on the flux (m3 wastewater/m2 membrane/min), shown in Figure 18. For the shower water test, there was only a significant difference ($p<0.05$) between PPG and PES membranes. The laundry water test showed a significant difference in flux between all three membranes. The combined shower and laundry water operation did not result in any significant difference for flux between the three membranes.

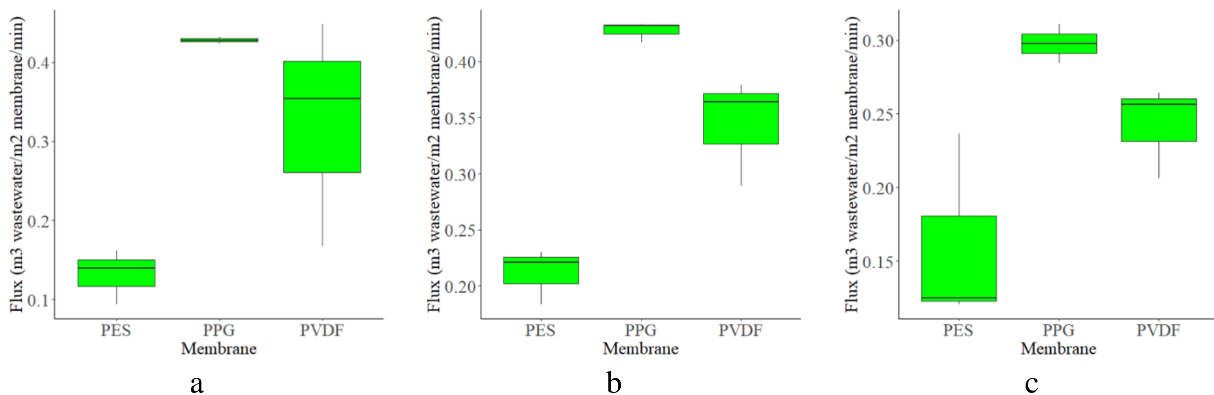


Figure 18. Effects of wastewaters on the flux of individual membranes.

(a) Flux of shower water filtration; (b) Flux of laundry water filtration; (c) Flux of shower/laundry water filtration

3.2. Fouling characteristics of three membranes on shower, laundry, and shower/laundry wastewaters

The total fouling mass accumulated on the surface of the three membranes was analyzed during the treatment of three wastewaters (Figure 19). Membrane mass before and after

treatment are shown in Table 18. For the treatment of shower wastewater, there were significant ($P<0.05$) differences on fouling mass between the three membranes. The fouling mass on PES, PPG, and PVDF were 2.99, 0.31, and 1.37 g/m2 membrane/m3 wastewater, respectively. PPG had less fouling mass than PES and PVDF. As for the laundry and laundry/shower wastewaters, there were no significant ($P>0.05$) differences between the three membranes. The fouling masses were 1.00, 0.65, and 1.03 g/m2 membrane/m3 wastewater for PES, PPG, and PVDF, respectively for the laundry wastewater. The treatment of shower/laundry combined wastewater led PES, PPG, and PVDF to accumulate the fouling mass of 0.72, 1.21, and 1.69 g/m2 membrane/m3 wastewater, respectively. Similar trends were observed for the accumulation of elements (C, O, N, and P) on three membranes treating different wastewaters (Figure 20 and Table 19).

The FT-IR data further illustrated the functional groups from wastewater that have accumulated on the membrane surfaces (Table 29). PPG membrane shows higher percent transmittance (%T) on all functional groups of alcohol OH, alkene CH, alkane CH, allene C=C, nitrogen compound, alkane methyl group, carboxylic acid OH, anhydride, and halo compound when compared to PES and PVDF membranes for all three wastewaters. The results indicate that the PPG membrane accumulated the compounds with these functional groups on all three wastewaters to a lesser degree than the PES and PVDF membranes. Meanwhile, considering both fouling mass accumulated on the membrane (Figure 19) and FT-IR data (Table 20), all of the functional groups accumulated on three membranes from the combined shower/laundry treatment were much less than the shower and laundry treatments separately, which shows that different wastewater sources significantly influenced the accumulation of functional groups on the membranes. During the shower/laundry treatment, PES had the lowest fouling mass on the membrane and elemental contents among the three membranes. The PES membrane also had

relatively low contents of functional groups (higher %T than PVDF and lower %T than PPG). The data demonstrates that PES could be a good option to treat combined shower/laundry without considering flux (PES has the slowest flux among three membranes) (Figure 18). The result of fouling characteristics elucidates that membrane selection needs to consider both wastewater characteristics and membrane properties.

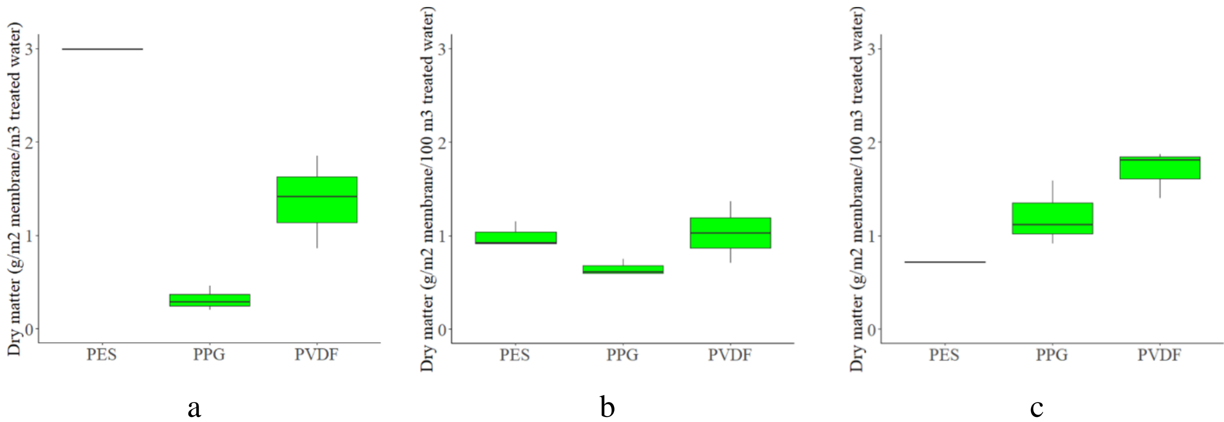


Figure 19. Mass accumulated on the membrane after the treatment. (a) Shower; (b) Laundry; (c) Combined shower and laundry.

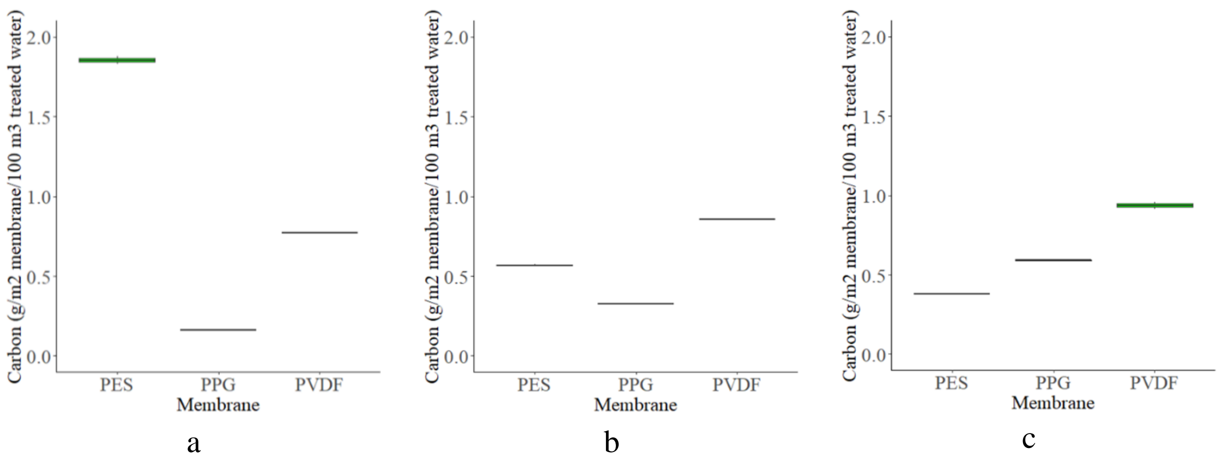
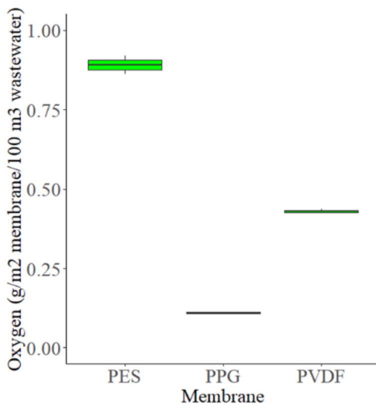


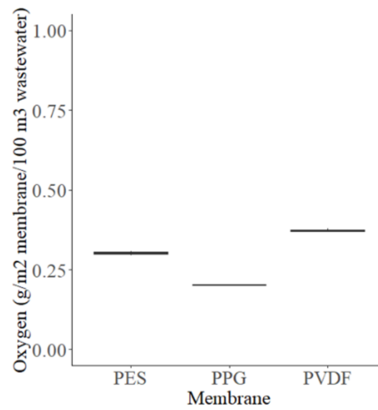
Figure 20. Element mass accumulated on the membrane after treatment.

(a) Carbon accumulation from shower water; (b) Carbon accumulation from laundry water; (c) Carbon accumulation from combined shower and laundry water; (d) Oxygen accumulation from shower water; (e) Oxygen accumulation from laundry water; (f) Oxygen accumulation from combined shower and laundry water; (g) Phosphorous accumulation from shower water; (h) Phosphorous accumulation from laundry water; (i) Phosphorous accumulation from combined shower and laundry water; (j) Nitrogen accumulation from shower water; (k) Nitrogen accumulation from laundry water; (l) Nitrogen accumulation from combined shower and laundry water.

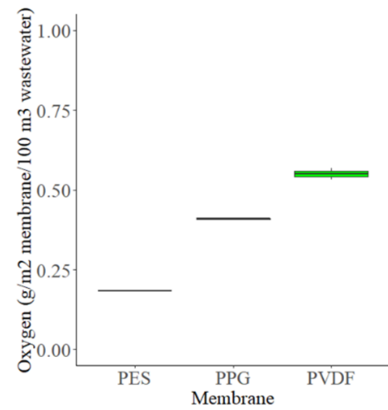
Figure 20 (cont'd)



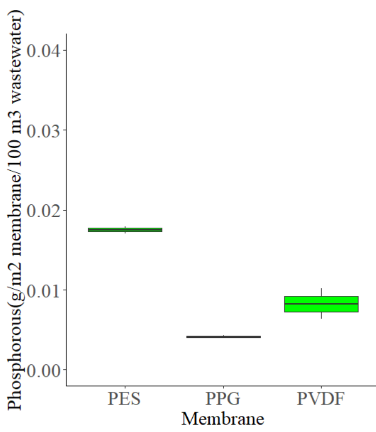
d



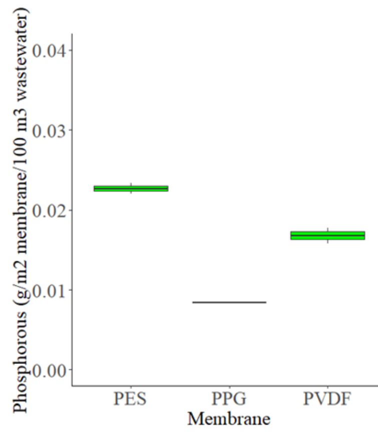
e



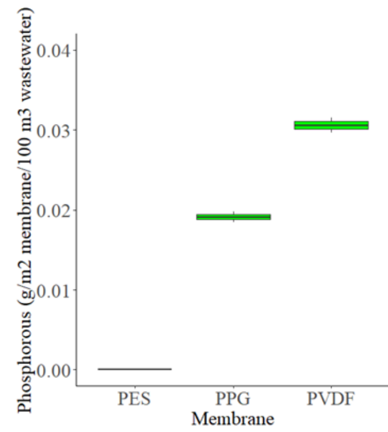
f



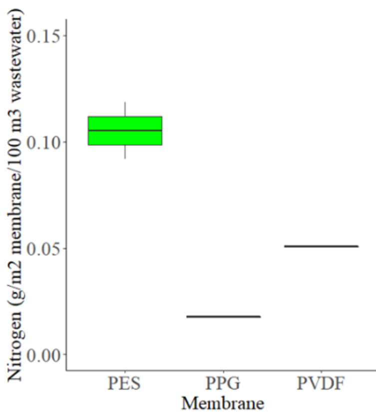
g



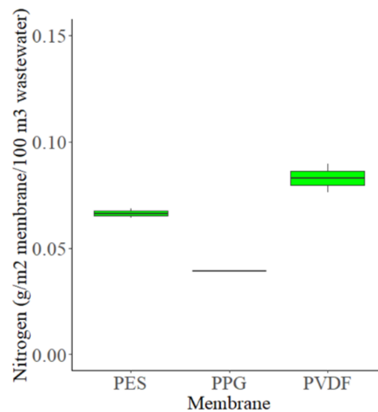
h



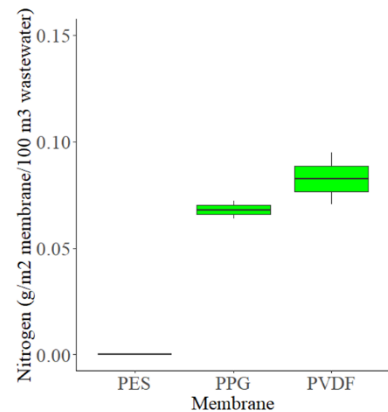
i



j



k



l

Table 18. Membrane mass before and after the treatment.

Wastewater	Membrane	Dry mass/membrane (g/m^2)
Control	PES	106.66 ± 1.00
	PPG	51.87 ± 1.52
	PVDF	124.19 ± 0.94
S	PES	107.81 ± 0.99
	PPG	54.02 ± 0.72
	PVDF	135.81 ± 3.42
L	PES	112.51 ± 0.66
	PPG	59.62 ± 0.86
	PVDF	133.54 ± 2.43
SL	PES	109.42 ± 3.40
	PPG	60.05 ± 1.90
	PVDF	135.28 ± 1.36

Table 19. Element data of fouling substances on the three membranes for shower, laundry, and shower/laundry wastewaters.

Wastewater	S			L			SL		
Membrane	PES	PPG	PVDF	PES	PPG	PVDF	PES	PPG	PVDF
C (g/m^2 membrane/m^3 treated water)	1.85±0.03	0.16±0.01	0.77±0.00	0.57±0.01	0.33±0.00	0.86±0.00	0.38±0.00	0.59±0.01	0.94±0.03
O (g/m^2 membrane/m^3 treated water)	0.89±0.04	0.11±0.01	0.43±0.01	0.30±0.01	0.20±0.00	0.37±0.01	0.18±0.00	0.41±0.00	0.55±0.03
P (g/m^2 membrane/m^3 treated water)	0.02±0.001	0.004±0 00002	0.008±0.003	0.02±0.0009	0.008±0.00	0.017±0.001	0.0±0.0	0.019±0.0009	0.031±0.001
N (g/m^2 membrane/m^3 treated water)	0.11±0.02	0.02±0.00	0.05±0.00	0.07±0.00	0.04±0.00	0.08±0.01	0±0	0.07±0.01	0.08±0.02
S (g/m^2 membrane/m^3 treated water)	0.024±0.002	0.001±0.0004	0.024±0.0001	0.010±0.0004	0.003±0.00	0.016±0.0004	0.024±0.00	0.008±0.002	0.016±0.00

Table 20. FT-IR data (%T) of fouling substances of three membranes on shower, laundry, and shower/laundry wastewaters.

Wastewater	S			L			SL		
Membrane	PES	PPG	PVDF	PES	PPG	PVDF	PES	PPG	PVDF
Alcohol OH	89.76±2.87	97.26±1.55	91.62±3.77	88.85±4.16	97.07±0.94	86.32±0.61	95.44±1.94	99.36±0.16	91.16±3.6
Alkene CH	82.91±5.35	97.99±1.79	83.16±4.83	85.69±5.57	97.94±0.82	81.67±1.59	95.36±1.96	99.83±0.23	87.96±5.76
Alkane CH	86.73±4.34	98.5±1.77	86.81±3.67	89.61±4.27	98.66±0.84	86.7±1.29	97.16±1.64	100±0.4	90.89±3.7
Allene C=C	81.28±4.94	93.4±2.75	85.82±6.13	76.87±7.22	91.59±2.58	73.94±1.58	91.83±4.12	96.54±0.48	86.13±9.12
Nitrogen compound	83.76±4.36	95.74±1.71	89.02±5.97	81.79±6.53	93.84±2.27	79.37±4.65	95.58±3.1	98.48±0.39	86.82±4.42
Alkane methyl group	87.11±3.88	99.66±0.62	89.81±3.55	88.19±5.54	98±1.63	86.45±0.86	96.83±2.89	100.5±0.4	91.21±2.51

Table 20 (cont'd)

Carboxylic acid OH	87.23±3.94	99.66±0.62	89.75±4.09	88.66±5.24	98±1.63	86.83±0.84	97±3.08	100.5±0.4	91.33±2.35
Anhydride COOCO	73.77±6.61	87.82±5.71	74.96±12.7	71.06±8.23	91.16±2.99	71.07±0.9	82.52±8.32	91.58±1.85	73.96±7.34
Halo compound	72.15±8.1	93±1.63	68.53±13.26	71.93±8.76	90.16±5.03	72.46±0.75	86.21±13.28	90.83±2.01	69.99±4.31

SEM imaging was also utilized to visualize the fouling layer on the surface of the membrane. Figure 21 shows the result of the SEM imaging on the different membranes operating on the three greywater sources. It is apparent that the PPG membrane had minimal surface fouling mass accumulation on all three wastewaters compared to the other two membranes, except PES from the shower/laundry wastewater, which also showed minimal surface accumulation (Figure 21d). The SEM data are consistent with the results discussed in the previous section.

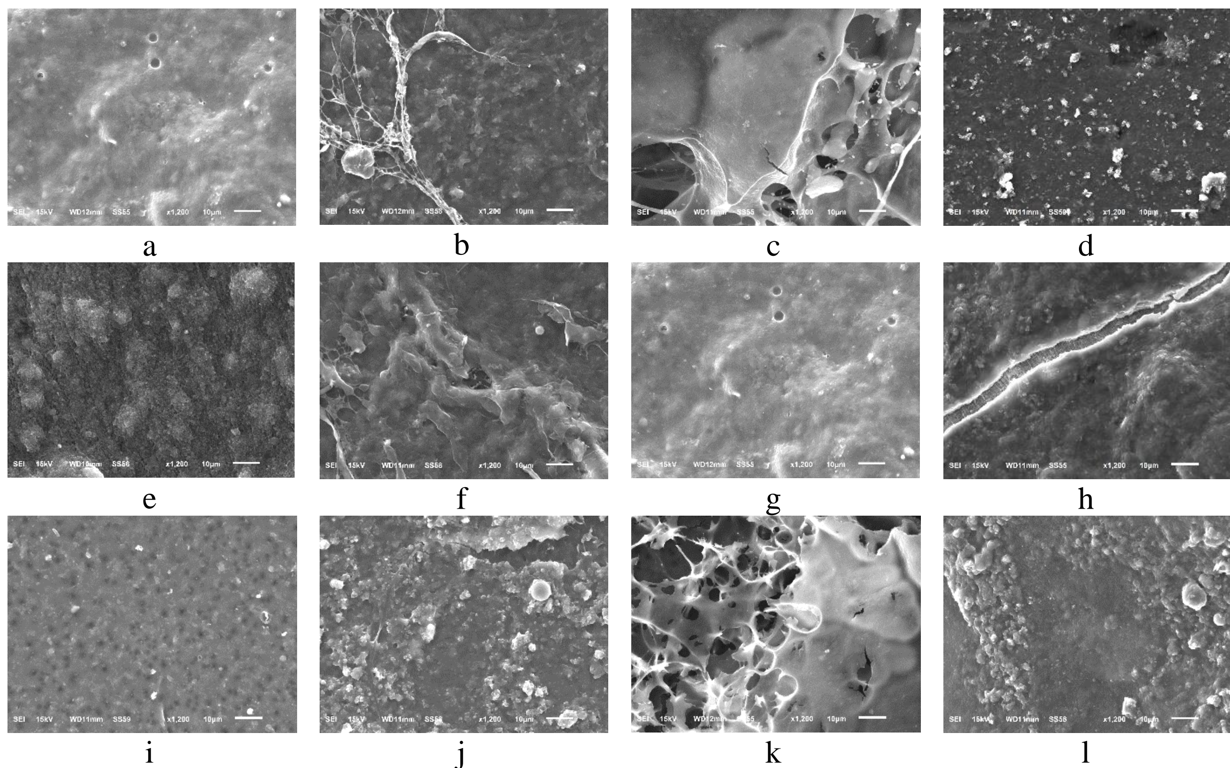


*Figure 21. SEM images of membranes on shower, laundry, and shower/laundry wastewaters *.*

(a) PES control; (b) PES shower; (c) PES laundry; (d) PES shower/laundry; (e) PPG control; (f) PPG shower; (g) PPG laundry; (h) PPG shower/laundry; (i) PVDF control; (j) PVDF shower; (k) PVDF laundry; (l) PVDF shower/laundry.
**: The crack shown on the layer in Figure 6h resulted from the drying process when the layer shrunk on the surface.*

3.3. Effects of membrane types and wastewater sources on fouling characteristics and treatment performance

Considering the importance of the flux, fouling, and quality of the treated water, a two-objective optimization approach, Pareto frontier, was applied to identify membranes that have the potential to possess good flux, minimal fouling, and high quality of the treated water from three different wastewaters. Figure 22 summarizes the Pareto frontier results.

The Pareto frontier analysis showed that the PPG membrane was the best for flux, powder mass accumulation, turbidity, and UV254 on both shower and laundry wastewater treatment (Figure 22a, b, and c). PPG had a flux of 0.43 m3 wastewater/m2 membrane/min for all three wastewaters of shower, laundry, and combined shower and laundry. Under this flux, the shower and laundry wastewater treatments accumulated 0.32 and 0.65 g/m2 membrane/100 m3 wastewater of fouling mass on the PPG membrane surface, respectively. Turbidity reduction and UV254 reduction of the PPG treatment of the shower wastewater were 99% and 82%, respectively. The PPG treatment of the laundry wastewater had a 98% and 86% reduction of turbidity and UV254, respectively. PPG also had a good COD reduction (72%) for the shower wastewater treatment under the flux of 0.43 m3 wastewater/m2 membrane/min (Figure 22d). However, the reduction of TN and TP of the PPG treatment on the shower wastewater were low (Figure 22e and f). Meanwhile, PPG was not very efficient at the removal of COD, TN, and TP from the laundry wastewater. In addition, the Pareto Frontier analysis also concludes that PPG is not a preferred membrane for combined shower and laundry wastewater treatment.

The two-objective optimization analysis shows that PES was on the Pareto Frontier lines to remove COD, TN, and TP from different wastewaters with better reduction efficiency compared with the other two membranes (Figure 22d, e, and f). However, the flux was much

slower than the other two membranes. PES removed 74% of COD, 70% of TN, and 75% of TP in shower wastewater with a low flux of 0.13 m3 wastewater/m2 membrane/minute. PES also resulted in a good performance to remove COD and TN from the laundry wastewater as well as the combined shower and laundry wastewater. The COD reduction with PES on the laundry wastewater and the combined shower and laundry wastewater were 58% and 48%, respectively at the flux of 0.21 wastewater/m2 membrane/minute. The TN reduction with PES on the laundry wastewater was 56% with a flux of 0.21 wastewater/m2 membrane/minute.

The PVDF membrane was also on the Pareto frontier lines of TN and TP reduction for the treatment of shower wastewater and combined shower/laundry wastewater (Figure 22e and f). A TN reduction of 55% was achieved from the treatment of the combined shower/laundry wastewater at the flux of 0.32 wastewater/m2 membrane/minute. PVDF removed 54% of TP at a flux of 0.34 wastewater/m2 membrane/minute, and 48% at a flux of 0.34 from the shower wastewater and combined shower/laundry wastewater, respectively.

The two-objective optimization analysis elucidates that PPG is very efficient in preventing fouling and remove turbidity and UV254 with a high flux. PES and PVDF are efficient in removing COD, TN, and TP with a tradeoff of lower fluxes. The analysis also demonstrates that the combined shower and laundry wastewater is more difficult to treat compared to separate shower wastewater and laundry wastewater. Nevertheless, Pareto frontier in this study is clearly presented as a useful multi-objective optimization tool that can be used to select the right membranes and treat targeted wastewater with better and more efficient treatment methods. Besides membrane selection, it can also be used as a preliminary screening tool to conclude membrane combinations that have good potential to efficiently treat different types of wastewater.

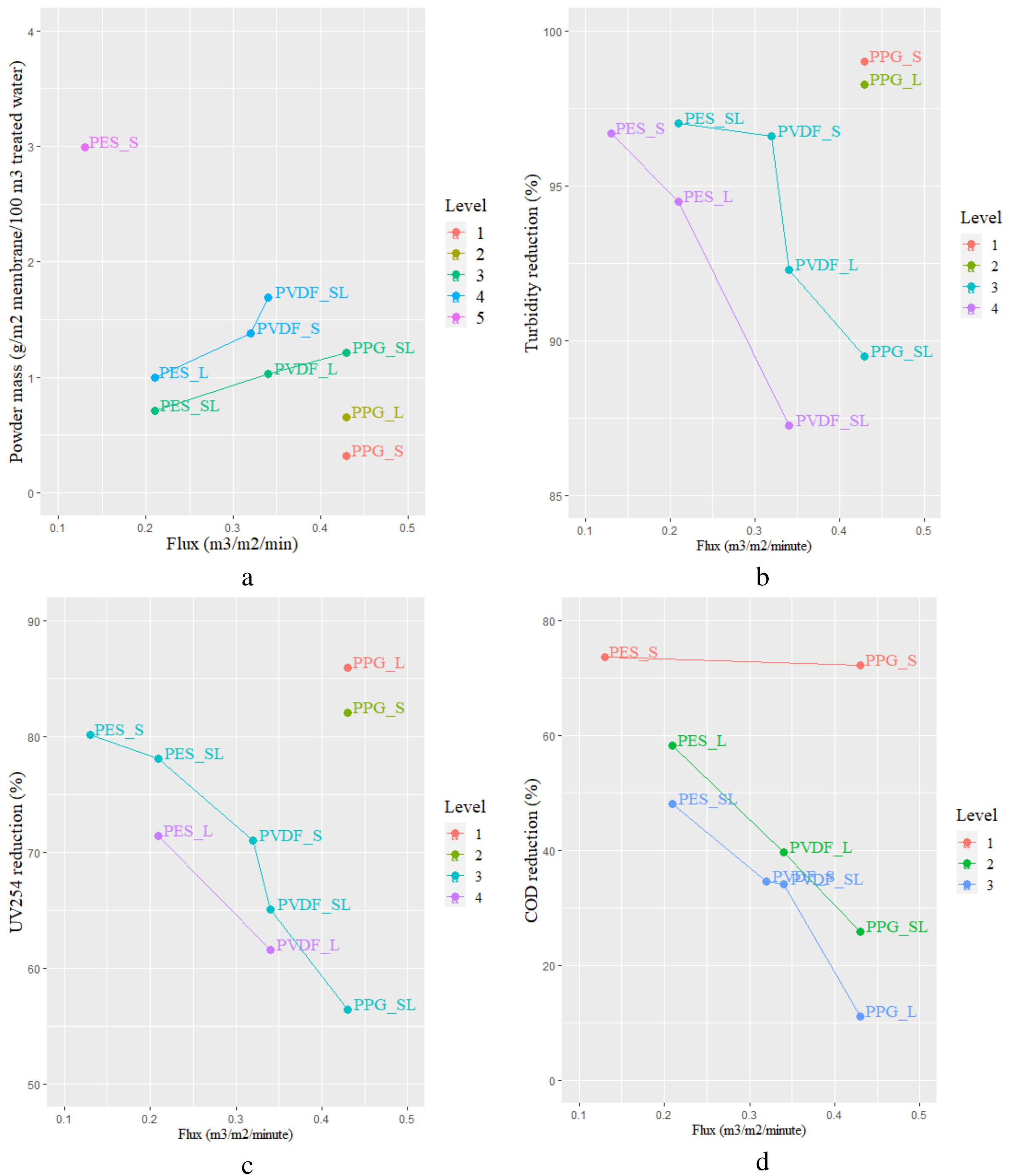
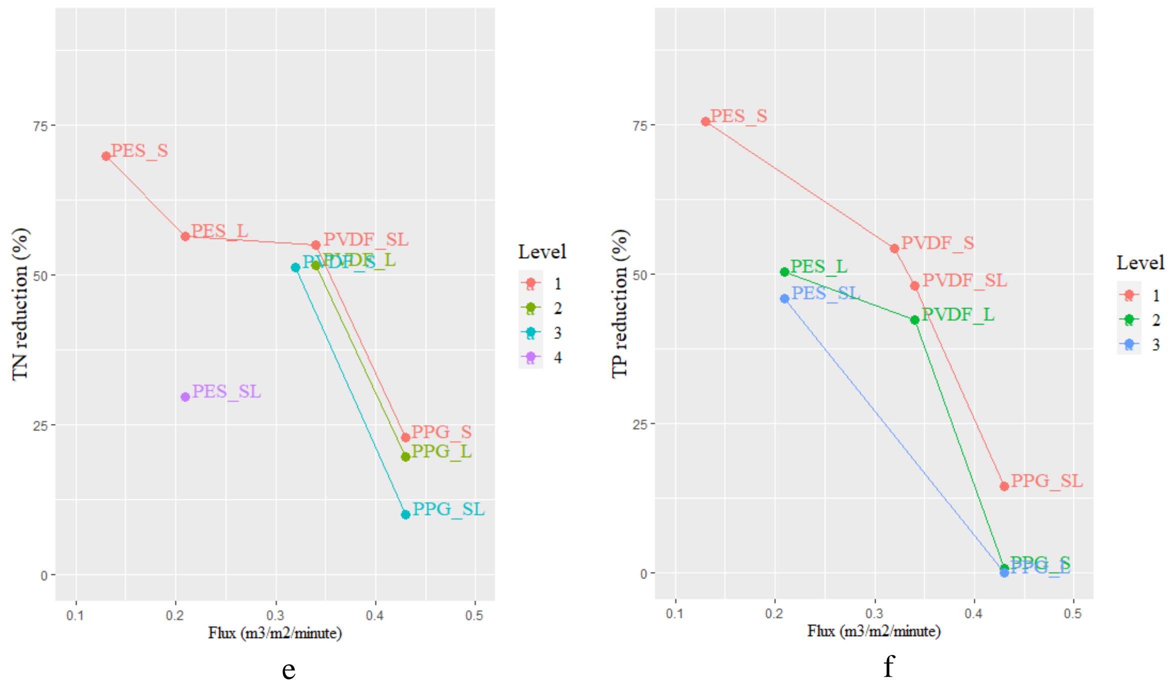


Figure 22. Pareto frontier and level of two-objective optimization.

a. Flowrate and powder mass accumulated on the membrane; b. Flowrate and turbidity reduction; c. Flowrate and UV254 reduction; d. Flowrate and COD reduction; e. Flowrate and TN reduction, f. Flowrate and TP reduction.

Figure 22 (cont'd)



4. Conclusions

Three ultrafiltration membranes of PPG, PVDF, and PES have been evaluated to treat shower, laundry, and combined shower/laundry wastewaters. The results elucidate that among the three membranes, PPG had the fastest flux for all three wastewaters. PPG also accumulated the least surface mass (fouling) compared to PVDF and PES on individual wastewaters, not including the combined shower/laundry. PES accumulated less surface mass than PPG and PVDF on the combined shower/laundry wastewater. Additionally, PES and PVDF showed better removal performance of COD, TN, and TP compared to the PPG membrane. Wastewater type also had significant influences on membrane performance. In general, the three membranes had relatively poor performance on the combined shower/laundry wastewater compared to the individual wastewaters. This study shows that membrane selection is extremely important for optimal treatment performance depending on the water source, even different types of greywater

had a substantial influence on treatment performance. Based on wastewater types and the treatment performance of individual membranes, using MOO to optimize membrane selection could be a solution to effectively treat different types of greywater.

CHAPTER 4: MULTI-OBJECTIVE OPTIMIZATION OF A MODULAR BASED TREATMENT SYSTEM TOWARDS SUSTAINABLE WASTE AND WASTEWATER MANAGEMENT

1. Introduction

Wastewater management is a critical but expensive operation for remote environments, due to the high capital and maintenance costs, restricted local budgets, lack of local expertise, and a lack of funding [89]. According to a recent study, the focus on rural development in India has been on establishing schools and healthcare facilities, with the absence of wastewater management systems due to financial considerations [90]. Communities in remote environments create a unique opportunity for wastewater and waste management. The cost of infrastructure for rural communities to integrate into a centralized wastewater treatment system is often cost-prohibitive, and the wastewater that is generated in such a community is often more concentrated than typical municipal wastewater due to a lack of dilution [52,53]. Given the limitations of centralized wastewater treatment in remote areas, decentralized wastewater management provides a potential alternative. On-site systems can be tailored to target the specific needs and resource constraints of an individual community's need, providing a cost-effective and environmentally friendly solution [91]. Treating wastewater closer to the source can significantly reduce the environmental risk from contamination during transport through miles of sewer pipelines and an increase in the energy efficiency of the system [92]. Considering that wastewater and organic wastes contain energy that can be utilized, the development of robust and energy-positive treatment systems is needed to turn the wastes from an environmental, health, and political liability into a valuable resource for water supply and renewable energy production to sustain wastewater and waste management operations in remote austere locations. Blackwater

and food waste were selected as the representative waste streams to study a modular-based decentralized treatment system.

Many studies have been conducted on blackwater treatment. Biological, physical, chemical, and electrochemical methods such as aerobic activated sludge, filtration, flocculation, and coagulation have been developed and used to treat blackwater [93, 94]. However, long start-up time (activated sludge), membrane fouling (membrane filtration), and additional chemical demand (flocculation and coagulation) make these methods difficult to implement at remote locations. Compared to these conventional treatment processes, electrocoagulation (EC) is an emerging technology to remove solid particles and other contaminants from wastewaters (e.g., pulp and paper wastewater, animal wastes) [95]. It has been applied to remove organic matter [96,97,98], nutrients [99], and microorganisms [100] from a variety of wastewaters. EC has several advantages, such as in-situ coagulant production induced by dissolving metal using electric current, the combination of three processes (coagulation, flocculation, settling) in a single step, short reaction/retention time, removal of small particles and color-causing compounds, and no additional sludge production [96, 101, 102]. The iron-rich EC sludge as a supplemental feed to an anaerobic digestion (AD) unit can stimulate the indirect interspecies electron transfer (IIET) between bacteria and archaea, so that the performance of AD (less TS in the AD effluent, more CH_4 and less CO_2 and H_2S in the biogas) would be significantly enhanced, and high carbon conversion efficiency could be achieved.

Despite the advantages of the EC technology, soluble compounds such as NaCl and ammonia are not able to be efficiently removed by the EC. To remove those soluble compounds and achieve higher water quality, additional treatment is needed. Electrodialysis (ED) has been widely reported to efficiently remove ions and impurities from water streams. It applies an

electric field across ion-selective membranes, causing ions to migrate towards electrodes of opposite charge. This migration facilitates the separation and extraction of dissolved salts and other soluble contaminants. Electrodialysis has several advantages in wastewater treatment. It operates at ambient temperatures and pressures, reducing energy consumption compared to traditional methods such as stripping and evaporation. Due to its unique separation mechanism, ED was selected to be a module in this study to polish the EC water to improve water quality.

Food wastes have high chemical oxygen demand (COD) and BOD contents. They are very good feedstocks for AD to produce biogas. Biogas can be used to provide energy on-site for decentralized wastewater treatment systems. There are a wide variety of digestion configurations for the treatment of different wastewater streams, such as plug-flow reactor for high-solid concentration streams (animal manure), completed stirred tank reactor (CSTR) for municipal sludge, anaerobic membrane bioreactor (AnMBR) for low-solid wastewater, upflow anaerobic sludge blanket (UASB) reactor and upflow fixed film reactor (UFFR) for food wastes, etc. [103]. Among these reactor configurations, CSTR has the advantages of less sensitivity to temperature change, efficient COD/BOD reduction, and good capability of handling both low and high-strength wastewater (providing the flexibility to treat EC sludge and food wastes). The digestion effluent with reduced volume and low TS and VS can be mixed with the blackwater, which can be treated by EC and ED to reclaim the water.

Considering the variation of blackwater amount and concentration for these decentralized treatment systems, biogas production can vary from time to time. The electricity from the Stirling engine of biogas conversion may not be sufficient to satisfy the need of the integrated system. Therefore, a secondary energy source is needed as an additional power supply to ensure stable operation of the system. Solar energy, as one of the most abundant renewable energy

sources on this planet, is used for this study. Several solar power technologies have been developed and implemented to generate electricity such as PV, parabolic trough systems, power tower systems, dish solar systems, Fresnel reflectors, etc. Among them, PV is a technology that satisfies the requirements of remote communities (scalable, simple, and easy to use).

During waste and wastewater treatment, key factors such as water quality, energy consumption, and treatment costs often conflict with each other. For example, high water quality typically demands more energy and requires more sophisticated and expensive equipment to achieve it. To optimize such a multiple objective system, trade-off(s) between these conflicting factors need to be considered. Therefore, a multi-objective optimization (MOO) approach was adopted in this study to carry out the optimization and selection of suitable treatment combinations. The MOO approach has been applied to optimize various aspects of wastewater treatment systems such as improving pollutant removal efficiency with the minimal use of resources, enhancing energy efficiency while maintaining treatment effectiveness, developing robust treatment systems to minimize the risk of non-compliance, and reducing the treatment cost while meeting treatment requirements.

This study focuses on analyzing and optimizing the integration of four modular operations (AD, EC, ED, and PV) to develop sustainable decentralized wastewater and waste management strategies for remote environments. Pilot-scale units for individual modules have been fabricated and tested by this study to generate the data. Based on the data obtained from the pilot unit, techno-economic analysis, life-cycle assessment, and multi-objective optimization were applied to conclude the preferred management strategies.

2. Materials and methods

2.1. Food waste and blackwater

Food waste was collected from Michigan State University (MSU) food services. Both pre-consumable and post-consumable food wastes were mixed as the food waste feed for this study. The synthetic blackwater was made using primary sludge from the East Lansing Wastewater Resource Recovery Facility. The primary sludge was diluted with fresh water by a factor of 20. Based on data of real blackwater, the synthetic blackwater used for the bench EC system and the selected pilot EC system was dosed with 0.89 g/L of ammonium chloride (NH_4Cl) to increase the ammonia nitrogen (NH_3-N) concentration. Characteristics of blackwater and food waste are listed in Table 21.

Table 21. Characteristics of blackwater and food waste.

Characteristics	Food waste	Blackwater
Total solids (TS, %)	20 ± 1	0.197 ± 0.05
Volatile solids (VS, %)	18 ± 2	-
TSS (mg/L)	-	970 ± 576
COD (mg/L)	317,543 ± 81,675	2,050 ± 616
TN (mg/L)	15,458 ± 240	125 ± 45
TP (mg/L)	2,000 ± 120	59 ± 7
E.coli (CPU/mL)	-	160,000
T. coliform (CPU/mL)	-	100,000
Somatic phage (pfu/mL)	-	157 ± 31
F-amp phage (pfu/mL)	-	85 ± 4

2.2. Modules and the treatment combinations for blackwater and food waste

The studied system includes four modules: electrocoagulation (EC) treatment of blackwater, anaerobic digestion (AD) for treatment of food waste and EC sludge, electrodialysis (ED) membrane treatment for final water treatment, energy generation from biogas using a Stirling engine, and photovoltaic (PV) solar energy for additional electricity generation. All modules were installed in a 20-foot iso-container at the East Lansing Water Resource Recovery Facility

(Figure 23). The system has been running for 11 months. The details of the individual modules are described as follows.

2.2.1. Individual modules and operation procedures

2.2.1.1. Electrocoagulation (EC) treatment of blackwater

One 16 L continuous-flow EC unit along with a 100 L settler was fabricated for the system (Figure 23f). The electrodes are connected in mono-polar mode. The power supply is a 40A and 24V DC power supply. The current density is 10-15 A/m^2 electrode, and the ratio of electrode surface area to solution volume is 1 m^2/0.1 m^3. The EC reactor is made of PVC. An aluminum EC sludge separator fabricated using aluminum was used to separate the EC sludge from the EC water. The EC sludge separator is placed on the top of the EC reactor.

During the EC operation, the blackwater is fed to the EC reactor in a continuous mode. The retention time of the EC treatment is 8.5 minutes. The EC effluent overflows to the EC sludge separator. The supernatant from the settler is collected as the EC water. The iron-rich EC sludge is also collected and used as a feed for the AD unit.

2.2.1.2. Anaerobic digestion (AD) of food waste and EC sludge

The two-stage AD (acidification and methanogenic stages) is adopted to carry out the digestion of food wastes and EC sludge (Figure 23c). Food waste and EC sludge are heated in the feeding vessels (200 L each) and then pumped to the AD module (Figure 23b). The reactor volume of the acidification tank is 930 L with the dimension of L×W×H = 0.7 × 0.7 ×1.9 m. The effective volume of the acidification tank is 750 L. The reactor volume of the methanogenic tank is 1,860 L with the dimension of L×W×H = 1.4 × 0.7 × 1.9 m. The effective volume is 1,500 L. The reactor vessels are made of high-density polyethylene. Both reactors are insulated without internal heating elements.

Food waste is directly fed to the acidification tank during the operation. The organic loading rate (OLR) of the acidification stage is 27 g VS/L/day. The effluent from the acidification stage is mixed with the EC sludge and fed to the methanogenic tank. The organic loading rate (OLR) of the methanogenic stage is 19 g VS/L/day. Iron from the EC sludge stabilizes and enhances digestion performance of the methanogenic stage, which makes the AD unit robust and flexible. The biogas from the AD system is stored in a 10 m^3 biogas bag (Figure 23e). A Stirling engine Combined Heat and Power (CHP) (Qnergy Co.) with a power capacity of 1.2 kW is used to directly utilize the raw biogas to generate electricity and heat (Figure 23d). The electricity and heat generated from the Stirling engine are used to satisfy the energy demands of unit operations.

2.2.1.3. Electrodialysis (ED) treatment of water reclamation

Since the EC cannot efficiently remove ammonia and other soluble compounds, the ED module is used to further treat the EC water. ED1000H from PCcell, Germany is used as the ED module (Figure 23g). The ED1000H has 50 cell pairs of ion exchange membranes with a membrane size of 30 x 50 cm. The total active membrane area is 1,500 cm^2 per membrane. The ED cell uses Pt/Ir-coated titanium as the anode and V4A steel as the cathode. The cell housing material is polypropylene. A 300 W DC power supply is used to power the ED unit. The recirculation flow rate of the EC effluent is 250 L/hour. The average voltage used for the ED process is 20V, and the max current is 3A. The treatment capacity is 75 L/hour.

2.2.1.4. PVs and batteries

PV panels (EVPV360PK, Panasonic) with a maximum voltage of 33.9V and maximum current of 10.6A for each panel are used to generate electricity to address the issue of insufficient energy and ensure the stable and continuous operation of individual modules (Figure 23a). 14 m^2 of PV panels were installed. The electricity conversion efficiency of the PVs can reach up to

22.1%. A power center with 8 Simpliphi® batteries and an Outback gateway controller is installed to store electricity and manage energy generation and consumption (Figure 23h). The Outback gateway is used to manage energy generation and consumption.

a

b

c

d

e

Figure 23. All modules in the iso-container.

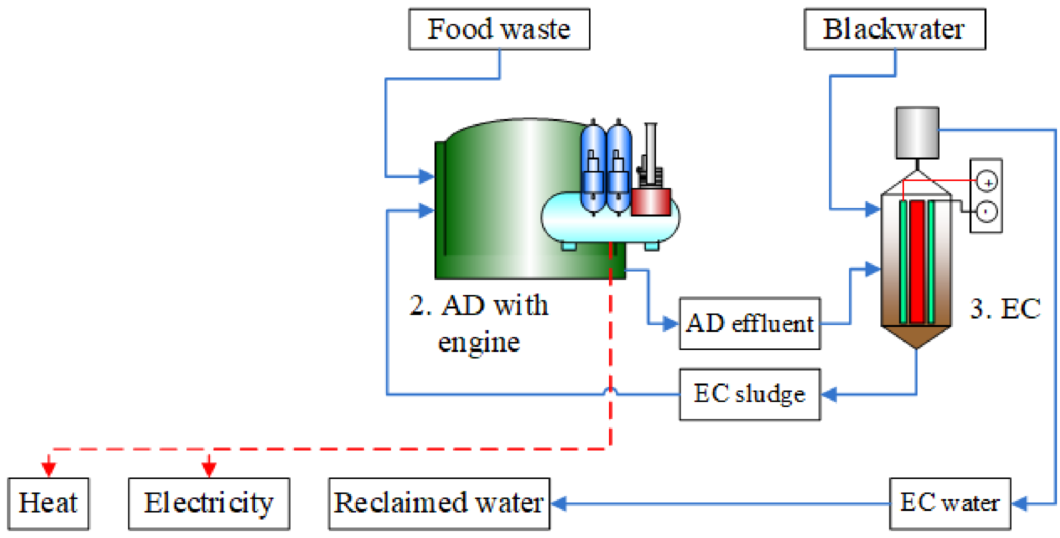
(a). the iso-container with PV panels; (b). Food waste feeding/heating; (c). Two-stage AD; (d). Stirling engine CHP for AD; (e). Biogas storage; (f). EC unit; (g). ED unit; (h). Battery storage

Figure 23 (cont'd)

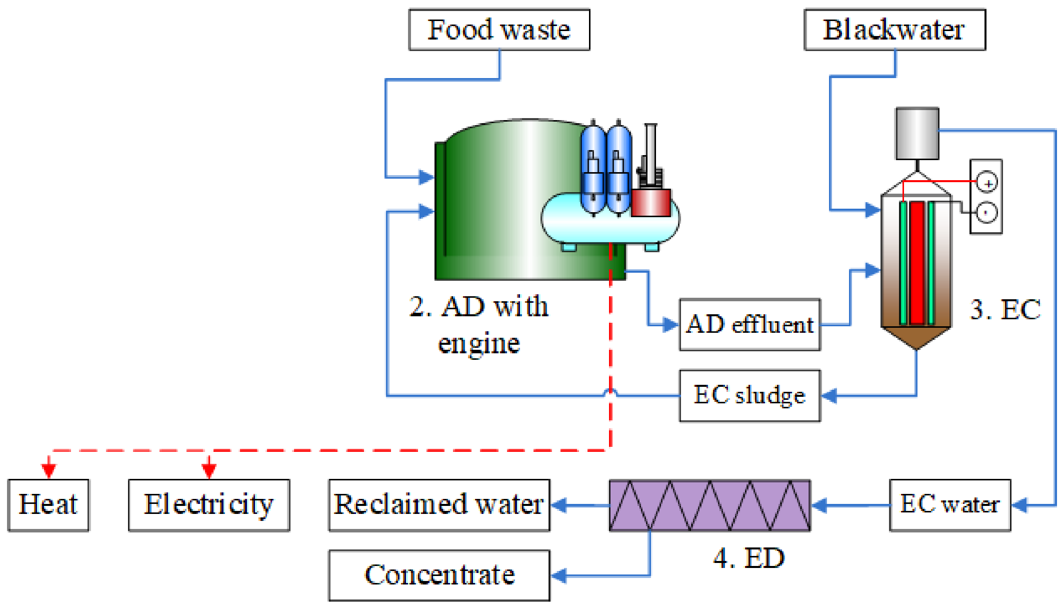
f g h

2.2.2. Treatment combinations

PV, AD, EC, and ED modules are arranged into four treatment combinations of AD+EC, AD+EC+ED, PV+AD+EC, and PV+AD+EC+ED (Figure 24). The combination of AD+EC uses AD and EC to treat food waste and blackwater, respectively. The EC sludge and AD effluent are circulated back to the AD and EC respectively to be treated with food waste and blackwater. Electricity and heat are generated by the Stirling engine CHP of AD biogas to power the treatment (Figure 24a). The combination of AD+EC+ED uses AD and EC to treat food waste and blackwater and generate energy first, which is the same as the combination of AD+EC. ED is then applied to treat and reclaim water from the EC effluent (Figure 24b). The combinations of PV+AD+EC and PV+AD+EC+ED follow the same patterns of AD+EC and AD+EC+ED and add PV as the second path of energy generation (Figure 24c & d).



a



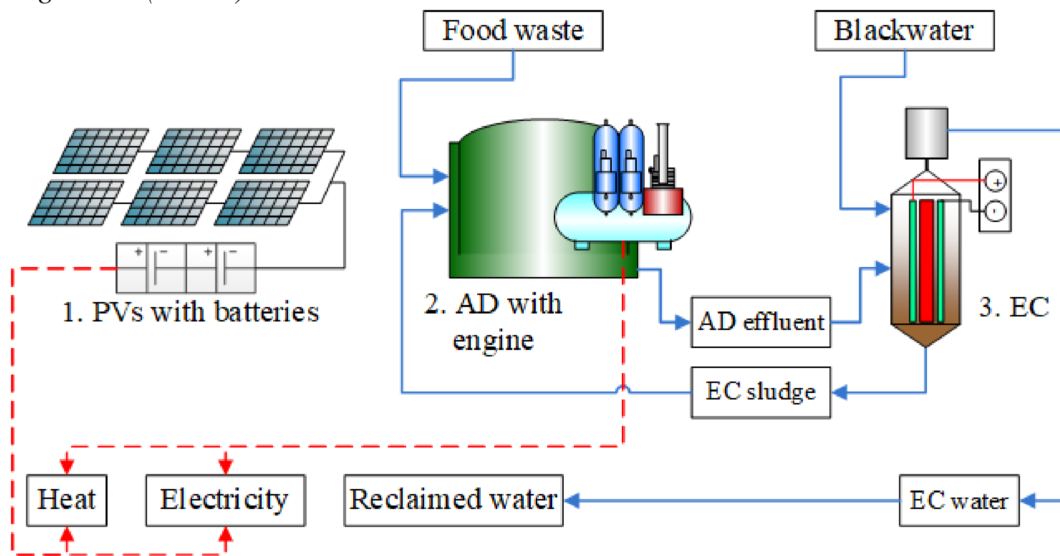
b

Figure 24. Four treatment combinations of individual modules to treat blackwater and food waste.*

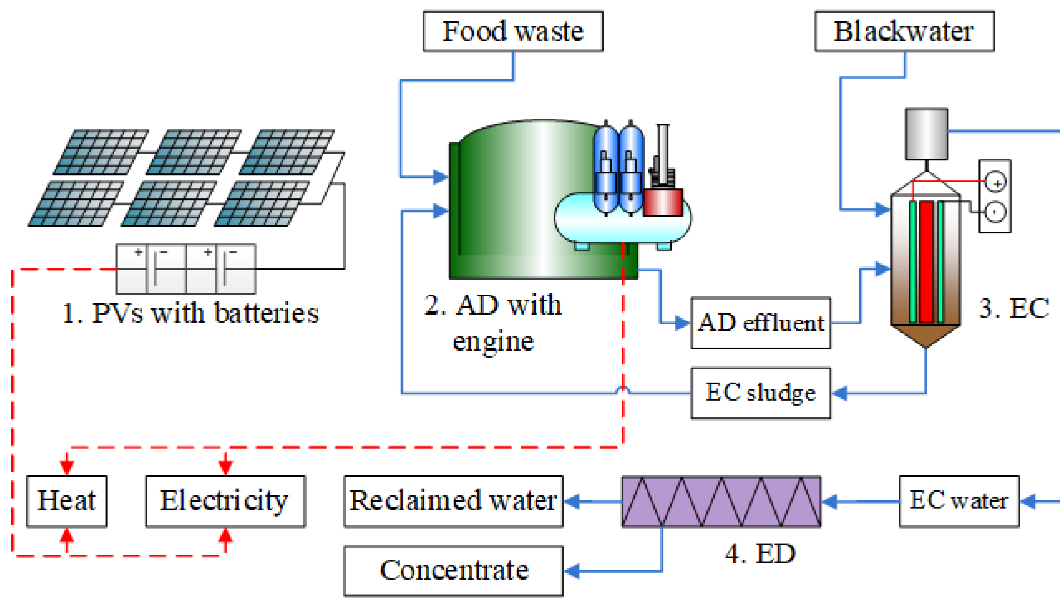
(a) AD+EC; (b) AD+EC+ED; (c) PV+AD+EC; (d) PV+AD+EC+ED

**: Blue solid lines represent the mass flows; Red dash lines represent the energy flows*

Figure 24 (cont'd)



c



d

2.3. Mass and energy balance

Mass and energy balance analyses were carried out for AD+EC and AD+EC+ED since the combinations of PV+AD+EC and PV+AD+EC+ED have the same mass flows as AD+EC and AD+EC+ED, respectively. The mass balance includes the following flows in (kg/day): blackwater, food wastes, EC sludge, AD effluent, EC treated water, ED treated water, and ED

concentrate. All flow data and characteristic data were obtained from the pilot operation. Following the mass balance analysis, an energy balance was conducted for each combination. Energy input data included energy consumption for EC, ED, and AD. Energy output data included energy generation from PV and AD. All energy data for the analysis was obtained from the pilot operation. The mass and energy balance analysis determined the energy demand per day or the energy demand/m^3 treated blackwater for individual treatment combinations.

2.4. Life cycle impact assessment (LCIA)

With the detailed mass and energy balance analysis, an LCIA was carried out to evaluate the environmental impacts of individual combinations compared to the conventional treatment practices of activated sludge treatment of blackwater and the landfill of food waste. The boundary of the LCIA is from the wastewater to the end products of the individual treatment combinations including treated water, renewable energy, and concentrate (Figure 24). Two impact categories related to carbon emission and water quality were chosen for the life cycle impact assessment: Global Warming Potential (GWP) and Water Eutrophication Potential (WEP). The data generated from the mass and energy balance was used to establish a life cycle inventory. All emission factors for individual compounds are listed in Table 22. The EPA Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (TRACI) version 2.1 was used for the LCIA. To calculate the impact for each category being considered, the substance mass from each emission source was multiplied by the listed characterization factors. Summing the total emissions within each impact category resulted in the total impact score for each category. Contribution analysis was performed to elucidate the influences of different treatment combinations on each impact category.

Table 22. Parameters for life cycle impact analysis.

	Item	Value	Unit	Data source
GWP	CH_4 emission factor of the activated sludge treatment of blackwater	0	Kg CO_2-e/kg TS	[105]
	CH_4 emission factor of the land application of food waste	2.3	Kg CO_2-e/kg TS	
	N_2O emission factor	0.005	g N emitted as N_2O/g TN in the food waste or wastewater	[105]
	Molecular weight conversion of N_2O per N_2	1.5714		
	GWP factor of N_2O emission	298	Kg CO_2-e/kg N_2O	[51]
	GWP factor of CH_4 emission	25		[51]
	GWP factor of natural gas electricity	0.491	Kg CO_2-e/kWh	[51, 106]
	GWP factor of diesel electricity	0.731	Kg CO_2-e/kWh	[51, 106]
WEP	WEP factor of TN	0.9864	Kg N-eq/kg TN	[51]
	WEP factor of TP	7.29	Kg N-eq/kg TP	[51]
	WEP factor of COD	0.05	Kg N-eq/kg COD	[51]

2.5. Economic analysis

The economic assessment is important to determine the viability of real-world application for the systems being analyzed. Capital Expenditure (CapEx), Operational Expenditure (OpEx), and cost-savings are the parameters used to assess the economic performance of different treatment combinations. The CapEx and OpEx data were collected from the system fabrication and the demonstration operation (Table 23). The current electricity cost of $0.18/kWh-e was used to calculate energy cost. The Modified Accelerated Cost Recovery System (MACRS) was used to calculate the annual depreciation of CapEx. The MACRS annual depreciation rates were 0.100, 0.188, 0.144, 0.115, 0.092, 0.074, 0.066, 0.066, 0.065, 0.065, 0.033, and 0.033 (after 10 years). Twenty years was set as the lifetime for individual treatment combinations. Annual inflation of

3.2% was set for OpEx, with a tax rate of 35%. The net cash flow based on depreciated CapEx and inflated OpEx was conducted to determine the treatment cost.

Table 23. Capital cost of individual units.

	The cost
PV unit	**$28,920**
23 m^2 PV panel and batteries	$4,900
8 batteries	$19,200
Control panel and software	$2,500
Unit installation (20% of the capital cost)	$4,820
AD unit	**$97,200**
20 ft containers	$3,000
Feeding unit with grinder	$3,000
Digesters with vessels, valves, pumps, and insulation	$15,000
Biogas storage (one 10 m^3 gas bags)	$5,000
Stirling engine CHP of direct biogas utilization	$45,000
Control panel and software	$10,000
Unit installation (20% of the capital cost)	$16,200
EC unit	**$16,200**
EC reactor with electrodes and valves	$3,500
EC sludge separator with electrodes and valves	$3,500
Pumps (feeding pump)	$4,500
Control panel and software	$2,000
Unit installation (20% of the capital cost)	$2,700
ED unit	**$18,500**
ED unit with valves and flow meters	$16,000
Power unit	$500
Control panel and software	$2,000
Unit installation (20% of the capital cost)	$2,300

2.6.Chemical analysis

Wastewater samples were collected daily using 1 L Nalgene bottles from the influent and effluent streams. Samples for total coliform and Escherichia coli analyses were collected using sterilized sample containers (250 mL, Nalgene). All parameters used for the characterization of

wastewater were completed immediately after their transfer to the laboratory. Total solids (TS) and total suspended solids (TSS) concentrations were measured using the standard gravimetric method (Method 2540 B &D) from Standard Methods for the Examination of Water and Wastewater [19]. Turbidity was measured using the nephelometric method (Method 2130) (APHA, 2012) with a portable turbidimeter (HACH, 2100Q). The concentration of chemical oxygen demand (COD) and total organic carbon (TOC) was analyzed using a wet oxidation-colorimetric method based on standard Method 5520-D and 5310 respectively [19] and kits (HACH) were used for the measurement. All nutrients (TN, TKN, TP, NH3-N, NO3-N, NO2-N) were measured using colorimetric methods using HACH kits prepared based on Standard Methods for the Examination of Water and Wastewater analyses [19]. Five-day BOD5 tests were carried out based on the respirometry technique using BODTrakII Respirometric BOD apparatus and a fresh seed was collected from the activated sludge process in Delhi WWTP (Holt, MI) for every measurement. Total coliforms and E-coli were detected using the membrane filter technique (Method 9222) [19] in a biosafety cabinet with laminar flow. All wet oxidation reactions were carried out in a digester (HACH DRB200) and colorimetric measurements were fulfilled by a spectrophotometer (HACH DR3900). Samples for microbial analysis were stored at -20 ºC until they were analyzed.

2.7. Multiple-objective optimization

The Pareto frontier, a MOO approach, was adopted for this study to carry out the optimization of the treatment combinations. The Pareto frontier represents the set of non-dominated solutions, where no other solution in the feasible solution space simultaneously improves one objective vector without worsening at least one other objective vector. Five objective vectors of water recovery, water quality, GWP, WEP, and treatment cost are used for

the MOO. They formed 25 pairs of two-vector combinations. The Pareto frontier allows the visualization of the trade-offs between these vectors. R function "psel" was used to run the optimization and to output and visualize the results. Each point on the Pareto frontier represents a solution that offers a different balance between the vectors.

3. Results and discussion

3.1. Performance of different treatment combinations

3.1.1. Mass and energy balance

According to the data obtained from the demonstration operation based on a small military contingency base, a mass and energy balance was conducted to evaluate the performance of individual treatment combinations (Figure 25). Since the PVs are for energy generation and do not contribute to the mass balance, the mass balance analysis was on treatment combinations of AD+EC and AD+EC+ED. For the combination of AD+EC (Figure 25a), the amount of blackwater fed to the EC unit for both combinations was 800 kg/day. The EC unit generated 780 kg of EC treated water/day with TSS of 29±12 mg/L, turbidity of 3.5±3.1 NTU, COD of 202±57 mg/L, BOD of 106±0 mg/L, NH_3-N of 87±18 mg/L, TP of 0.52 ±0.04 mg/L, E.coli of 23 CFU/100 mL, total coliform of 74 CFU/100 mL, somatic phage of 20 PFU/mL, and F-amp phage of 9 PFU/mL (Table 24). The EC unit significantly improved the water quality; however, the quality of the EC water does not satisfy the EPA wastewater discharging standards. Meanwhile, the EC unit generated 20 kg/day of EC sludge with TS of 1.73±0.52%, TSS of 20,425±265 mg/L, COD of 16,146 mg/L (Table 24). The EC sludge was fed into the methanogenic stage of the AD unit. The amount of food waste fed to the acidification stage of the AD unit was 20 kg/day. The AD generated 2,000 L raw biogas/day with 70.0±2.2% (v/v) of

CH_4, and 30±2.4% of CO_2, and 2.4±4.7 of H_2S. The raw biogas was directly used by the Stirling engine CHP to generate electricity and heat.

Due to the incompetence of the EC unit to achieve water quality to reclaim the water, an ED unit was included in the combinations of AD+EC+ED and PV+AD+EC+ED, the AD and EC treatments are the same as the combinations of AD+EC and PV+AD+EC. The ED unit treated the EC water and generated 772 kg/day of the ED water and 8 kg/day of the nutrient-rich EC concentrate. The ED treated water had TSS of 7.4±5.2 mg/L, turbidity of 2.7±1.0 NTU, COD of 64±7 mg/L, BOD of 14.5±0 mg/L, NH_3-N of 3.7±1.5 mg/L, TP of 0.59±0.39 mg/L, somatic phage of 6 PFU/mL, and F-amp phage of 1 PFU/mL (Table 24). The nutrient-rich EC concentrate contained COD of 281±121 mg/L, NH_3-N of 406±131 mg/L, and TP of 0.68±0.23 mg/L (Table 24). The ED treated water is clean enough to satisfy the EPA discharging standards.

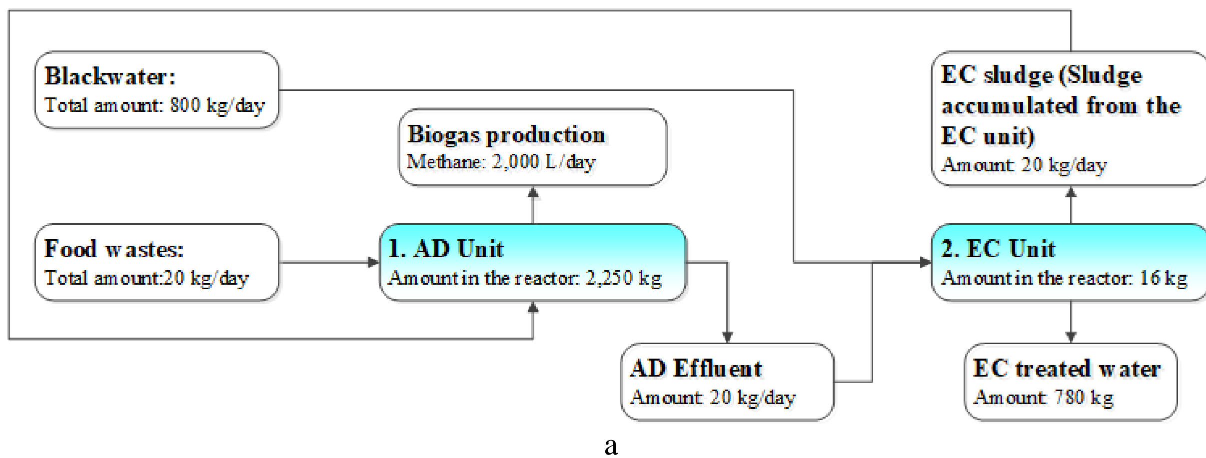
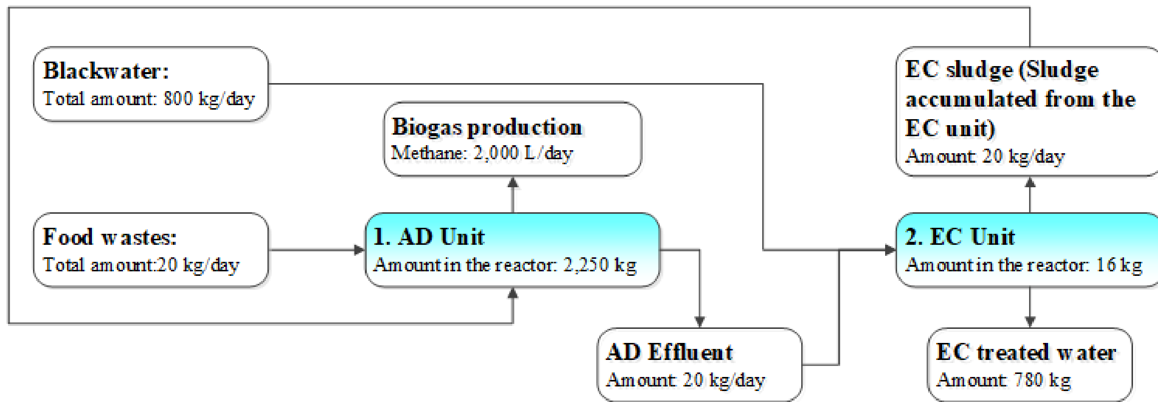


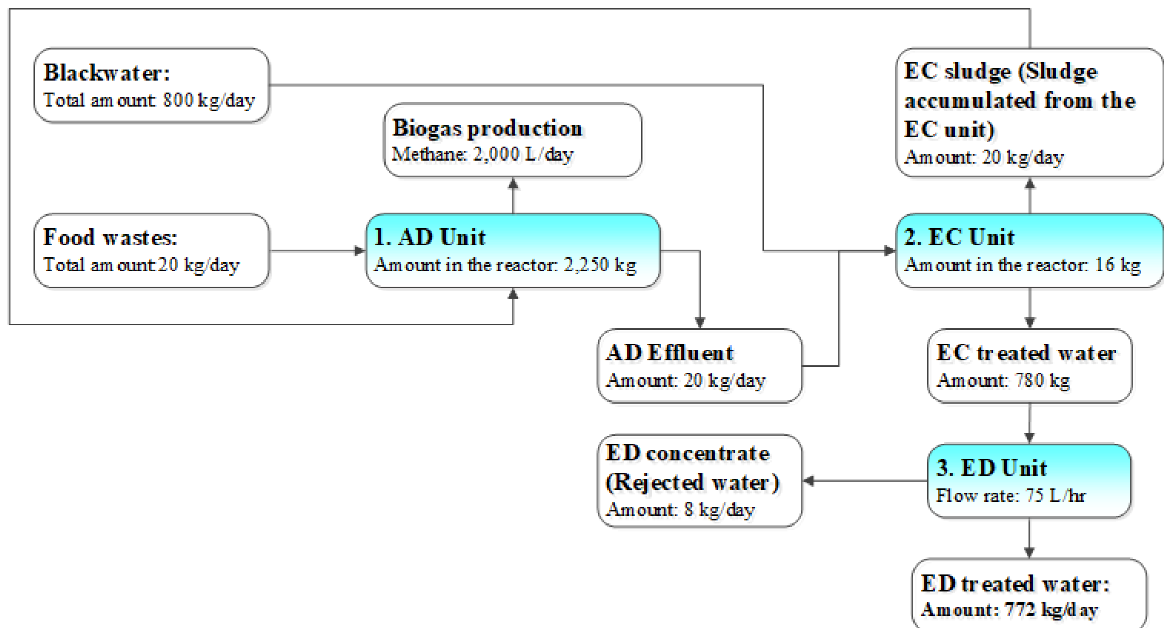
Figure 25. Mass balance of different combinations.

(a). For the combinations of AD+EC and PV+AD+EC); (b). For the combinations of AD+EC+ED and PV+AD+EC+ED.

Figure 25 (cont'd)



a



b

Table 24. Characteristics of the treated water and collected sludge.

Characteristics	AD effluent	EC treated water	EC sludge	ED treated water	ED concentrate
Total solids (TS, %)	0.78 ± 0.28	-	1.73 ± 0.52	-	-
Volatile solids (VS, %)	0.39 ± 0.23	-	-	-	-
Total suspended solids (TSS, mg/L)		29 ± 12	20,425 ± 265	7.4 ± 5.2	21 ± 12
COD (mg/L)	11,130 ± 7,299	202 ± 57	16,146	64 ± 7	281 ± 121
BOD (mg/L)	-	106 ± 0	-	14.5 ± 0	-

Table 24 (cont'd)

TN (mg/L)	2,024 ± 720	-	-	-	-
NH_3-N (mg/L)	-	87 ± 18	-	3.7 ± 1.5	406 ± 131
TP (mg/L)	583 ± 111	0.52 ± 0.04	443 ± 0	0.59 ± 0.39	0.68 ± 0.23
Turbidity (NTU)	-	3.5 ± 3.1	-	2.7 ± 1.0	5.2 ± 2.0
Somatic phage (PFU/ml)	-	20 ± 4	-	6 ± 1	-
F-amp phage (PFU/ml)	-	9 ± 2	-	1 ± 0	-

The energy balance analysis was then concluded to evaluate the energy performance of the system (Table 25). The operational data shows that the Stirling engine CHP has electricity and heat conversion efficiencies of 13.6 ± 1.6% and 25.7 ± 5.4% to utilize the raw biogas. The CHP generated 5.1 and 2.6 kWh-e/day of heat and electricity from the combustion of 2,000 L/day of the raw biogas from the two-stage AD. Meanwhile, 14 PVs generated an average of 22 kWh/day in East Lansing, MI (only considering the year-round sunny days). For energy consumption, The AD unit consumed 6 and 0.8 kWh-e/day of heat and electricity to maintain the digestion temperature and operate the mixer and pumps. The EC unit used 6.5 kWh-e/day to treat 800 kg blackwater, and the ED unit demanded 0.6 kWh-e/day to reclaim 772 kg/day of the ED treated water.

The energy balance results show that the net energy outputs of -7, -7.8, 20.5, and 19.8 kWh-e/m^3 treated blackwater are for the combinations of AD+EC, AD+EC+ED, PV+AD+EC, and PV+AD+EC+ED, respectively (Table 25). The combinations without PVs (AD+EC and AD+EC+ED) cannot self-sustain their operations. Additional energy sources are needed to support both systems. However, considering the relatively low organic loading rates (27 g VS/L/day and 19 g VS/L/day for acidification and methanogenic stages, respectively) for the AD unit in this study, increasing food waste loading could significantly increase the AD energy outputs, which could make both combinations energy neutral.

Meanwhile, the combinations with PVs (PV+AD+EC and PV+AD+EC+ED) clearly show the benefits of net energy output. The positive energy outputs of both combinations indicate that PVs can significantly enhance the energy performance of the treatment combinations. With the PVs, 16.4 and 15.8 kWh-e/day of extra energy are generated from PV+AD+EC and PV+AD+EC+ED, respectively, during the treatment. Consequently, energy-positive waste and wastewater treatment solutions have been achieved.

Table 25. Energy balance of different combinations [a, b].

System	Energy input (kWh-e/day)				Energy output (kWh-e/day)			Net energy output (kWh-e/day)	Net energy output (kWh-e/m^3 treated blackwater)
	AD-Heat	AD-electricity	EC	ED	PV [c]	AD-Heat [d]	AD-electricity [d]		
AD + EC	-6	-0.8	-6.5	-	-	5.1	2.6	-5.6	-7
AD + EC + ED	-6	-0.8	-6.5	-0.6	-	5.1	2.6	-6.2	-7.8
PV + AD + EC	-6	-0.8	-6.5	-	22	5.1	2.6	16.4	20.5
PV+ AD + EC + ED	-6	-0.8	-6.5	-0.6	22	5.1	2.6	15.8	19.8

a. *The positive numbers are energy outputs, and the negative numbers are energy inputs. The energy consumption is based on the average during a year-round operation and only considers the sunny days.*
b. *Data was collected from the demonstration operation.*
c. *The solar panels can collect 22 kWh/day of electricity on a sunny day.*
d. *The lower heating value of methane is 35 MJ/m^3 (9.72 kWh-e/m^3). The electricity and heat efficiencies of the Stirling engine are 13.6 ± 1.6% and 25.7 ± 5.4%.*

3.1.2. Life cycle impact assessment (LCIA) of different combinations

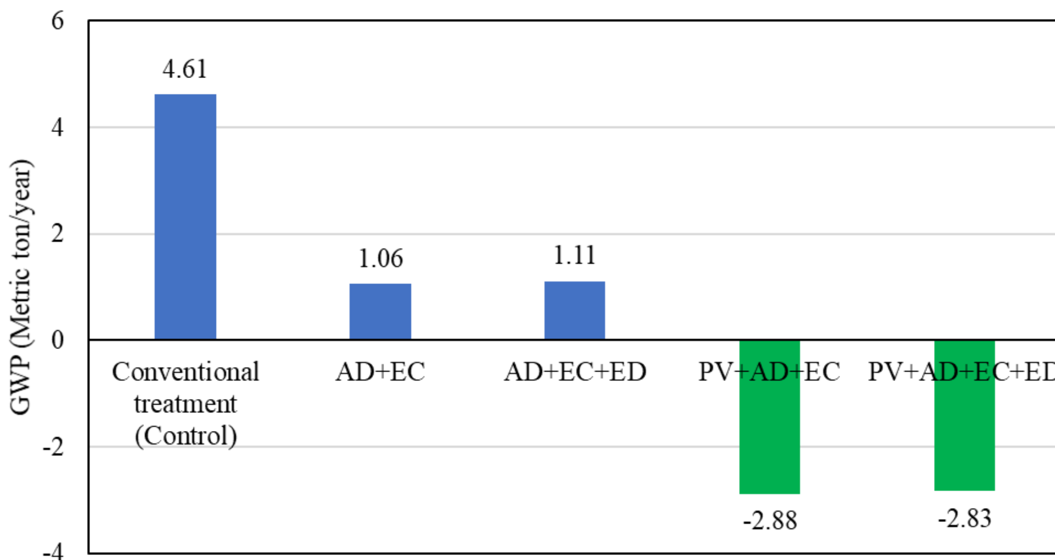
The LCIA was conducted to evaluate and compare the environmental impacts of these treatment combinations. GWP and WEP are the two impact factors evaluated in this study (Figure 26).

GWPs of individual treatment combinations were calculated based on individual modules of AD, EC, ED, PV, energy usage, and carbon and nitrogen contents of the treated water and EC sludge (Figure 26a). For the control, CO_2 emissions from the activated sludge treatment of blackwater and the landfill of food waste are biogenic and therefore have no impact on the treatment emissions. CH_4 emissions from the land application of food waste, N_2O emissions from the treated blackwater discharging and food waste were counted for GWP. In addition, natural gas electricity for the activated sludge treatment was also counted for GWP. The GWP for the control is 4.61 metric tons CO_2-e/year. For the four treatment combinations, since all carbon flows are contained in the treatment, the greenhouse gas emissions were N_2O emissions from residual nutrients in the EC water and ED water, and the natural gas electricity used for AD+EC and AD+EC+ED. Due to the natural gas electricity usage, the GWPs for AD+EC and AD+EC+ED were positive at 1.06 and 1.11 metric tons CO_2-e/year, respectively. As for PV+AD+EC and PV+AD+EC+ED, the inclusion of solar energy made both treatment combinations energy-positive. No external fossil-based energy was needed to operate the treatment. The GWPs for PV+AD+EC and PV+AD+EC+ED were -2.88 and -2.83 metric tons CO_2-e/year, respectively. The results demonstrate that the PV addition of the treatment combinations (PV+AD+EC and PV+AD+EC+ED) enabled both treatments to be carbon-negative.

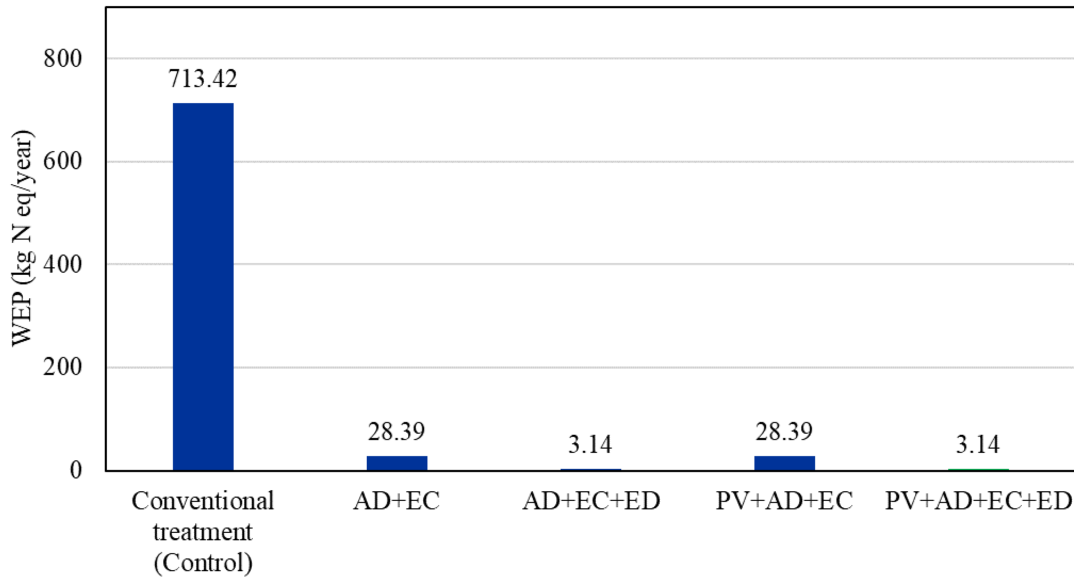
WEP was calculated based on the total amount of N and P discharged to the environment from different treatment combinations (Figure 26b). The WEP of the control was 713 kg N-eq/year due to high nitrogen and phosphorus contents in blackwater and food waste (Table 21). Among four treatment combinations, the EC water from the PV+AD+EC and AD+EC had higher nitrogen and phosphorus contents than the PV+AD+EC+ED and PV+AD+EC+ED. The

WEP of the PV+AD+EC and AD+EC was 28 kg N-eq/year, which is 9 times higher than the WEP (3.14 kg N-eq/year) from the PV+AD+EC+ED and AD+EC+ED. ED is the key module to remove nutrients and reduce WEP in the treatment combinations.

The life cycle impact assessment elucidates that different combinations of modules had significant impacts on the environment. The results demonstrate that the modules of PV and ED are the key components to significantly reduce environmental impacts and achieve sustainable treatment operations.



a



b

Figure 26. Contribution analysis of GWP and WEP for different combinations.

(a). Global warming potential with electricity from natural gas; (b). Water eutrophication potential

3.1.3. Economic analysis of different treatment combinations

Economic performance is another important factor to determine the viability of the potential real-world application of the different treatment combinations. As presented in Table 26, the CapExs of AD+EC, AD+EC+ED, PV+AD+EC, and PV+AD+EC+ED are $113,400, $131,900, $142,320, and $160,820, respectively. With the addition of PV and ED, the combinations of PV+AD+EC and PV+AD+EC+ED are more expensive than the other two combinations. The corresponding OpExs are $14,488, $15,452, $12,623, and $13,603, respectively. The lower OpExs of PV+AD+EC and PV+AD+EC+ED are due to the energy savings from PV electricity.

The cash flow analysis demonstrates that considering a 20-year payback period, The treatment costs of AD+EC, AD+EC+ED, PV+AD+EC, and PV+AD+EC+ED are $86, $96, $89,

and $98/m^3 treated water. The higher CapEx of the ED module led to higher treatment costs of the treatment combinations with ED.

Since the studied treatment combinations are all for small-scale operations, the amounts of the reclaimed water and carbon credits were small. The savings on both items were not considered and included in the OpEx for this analysis.

Table 26. Economic performance of different combinations.

	AD+EC	AD+EC+ED	PV+AD+EC	PV+AD+EC+ED
Capital expenditure (CapEx) ($)	113,400	131,900	142,320	160,820
Operational expenditure (OpEx) ($/year)	14,488	15,452	12,623	13,603
Maintenance ($/year) [a]	5,670	6,595	5,670	6,595
Labor cost ($/year) [b]	8,450	8,450	8,450	8,450
Energy demand or saving ($/year) [c]	368	565	-1,497	-1,442
Treatment cost ($/m^3 treated water) [d]	86	96	89	98

a. *The maintenance cost is based on the demonstration operation.*
b. *It requires 1 hour/working day to feed the system and check the operation based on the pilot operation. The hourly payment for the operator is $25/hour with a 30% fringe benefit.*
c. *The cost of energy demand is assigned as positive numbers, and the cost of energy generation is assigned as negative numbers. The energy cost is $0.18/kWh-e based on the market price of electricity in Michigan in 2024.*
d. *The treatment cost is calculated based on 20 years of lifetime for individual combinations.*

3.2. Multiple-objective optimization of system performance

Considering the importance of five objective vectors: water quality, water recovery, GWP, WEP, and treatment cost, a two-objective optimization approach, Pareto frontier, was applied to delineate the relationship between them and select preferred treatment combinations and conditions. Figure 27 summarizes the Pareto frontier results.

The Pareto frontier analysis showed that the combination of AD+EC was the best for treatment cost (Figure 27g). The AD+EC has the lowest treatment cost of $86/m^3 treated water and the best water recovery of 780 kg/day among the four treatment combinations. However, the

other vectors of water quality (not satisfying the EPA discharging standards), net energy output (5.6 kWh-e/day), GWP (1.06 metric ton CO_2-e/year), and WEP (28.4 kg N-eq/year) were not as good as other combinations.

The combination of AD+EC+ED showed the best performance on two vectors of water quality (satisfying the EPA discharging standards) and WEP (3.14 kg N-eq/year among four combinations (Figure 27c). While AD+EC+ED performed less efficiently on the water recovery (772 kg treated water/day), treatment cost (\$96/m^3 treated water), and energy output (6.2 kWh-e/day) than other combinations.

Meanwhile, two combinations with PVs show different performance from the combinations without PVs. Both energy output and GWP were greatly improved. The combination of PV+AD+EC demonstrates the best performance on three vectors of water recovery (780 kg treated water/day), energy output (-16.4 kWh-e/day), and GWP (-2.88 Metric ton CO_2-e/year) (Figure 27e & h). Since the combination does not include ED, it had poor performance on the water quality (not satisfying the EPA discharging standards) and WEP (28 N-eq/year).

The combination of PV+AD+EC+ED indicates the best performance on two vectors of water quality (satisfying the EPA discharging standards) and WEP (3.14 kg N-eq/year). It also performed well on GWP (-2.83 Metric ton CO_2-e/year) and energy output (-15.8 kWh-e/day), even though they are slightly lower than the combination of PV+AD+EC. Due to the fact that all four modules are included in this combination, it had the highest treatment cost (\$98/m^3 treated water) among the four combinations.

Considering the priority of water quality, the Pareto frontier analysis elucidates that both AD+EC+ED and PV+AD+EC+ED are the preferred combinations. PV+AD+EC+ED also has better energy output than AD+EC+ED with the drawback of higher treatment cost. Nevertheless,

Pareto frontier in this study demonstrates a useful multi-objective optimization tool that can be used to select treatment combinations that recover water with targeted quality and good performance efficiency.

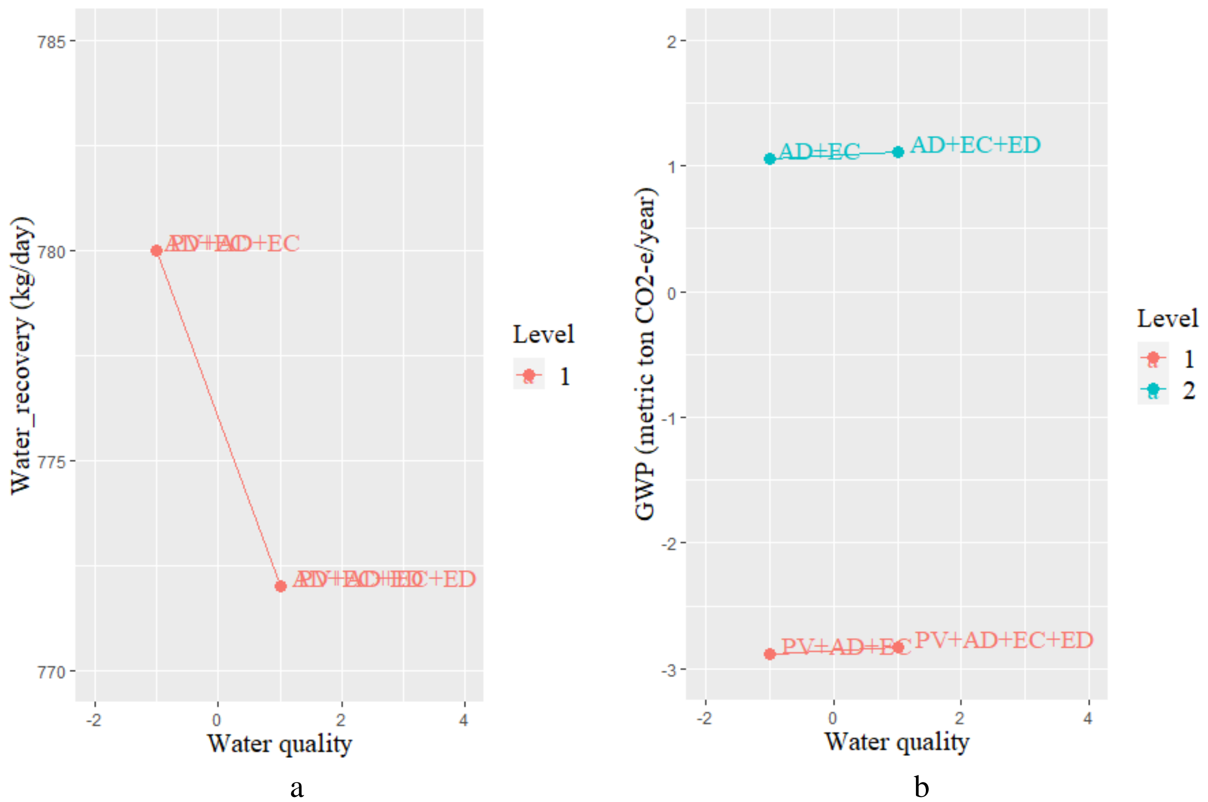
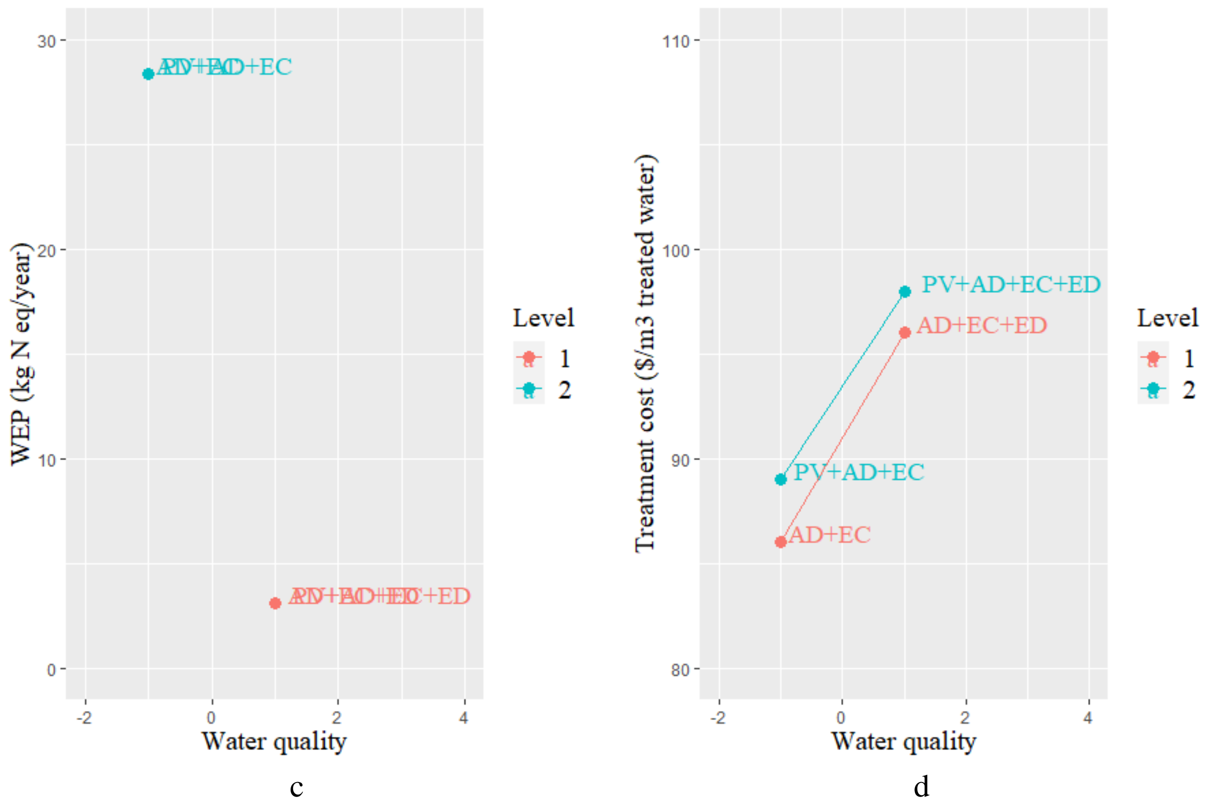
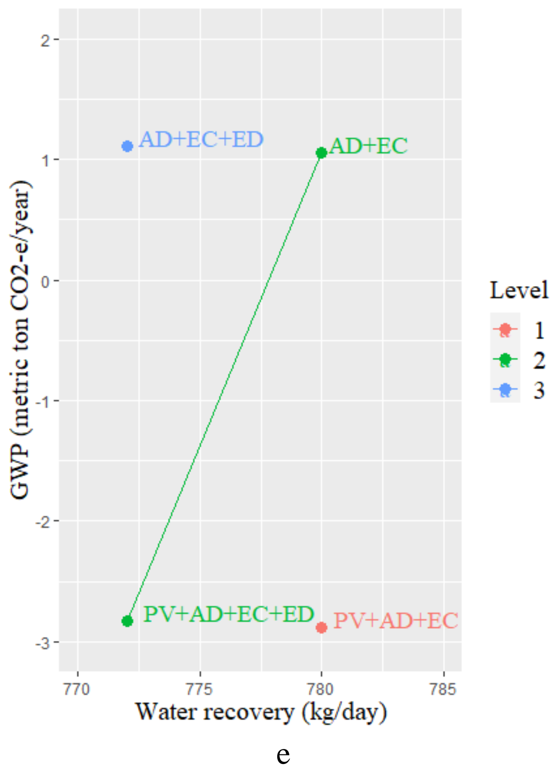


Figure 27. Pareto frontier lines for the intersection preference of different objectives.

(a)maximum water quality and maximum water recovery; (b)maximum water quality and minimum GWP; (c) maximum water quality and minimum WEP; (d) maximum water quality and minimum treatment cost. (e) maximum water recovery and minimum GWP; (f). maximum water recovery and minimum WEP; (g). maximum water recovery and minimum treatment cost; (h). minimum WEP and minimum GWP; (i). minimum WEP and minimum treatment cost; (j). minimum GWP and minimum treatment cost

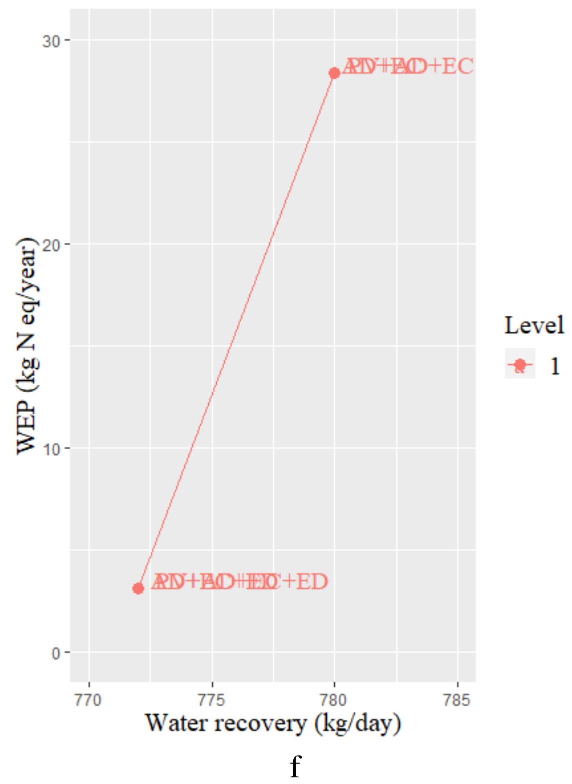
Figure 27 (cont'd)



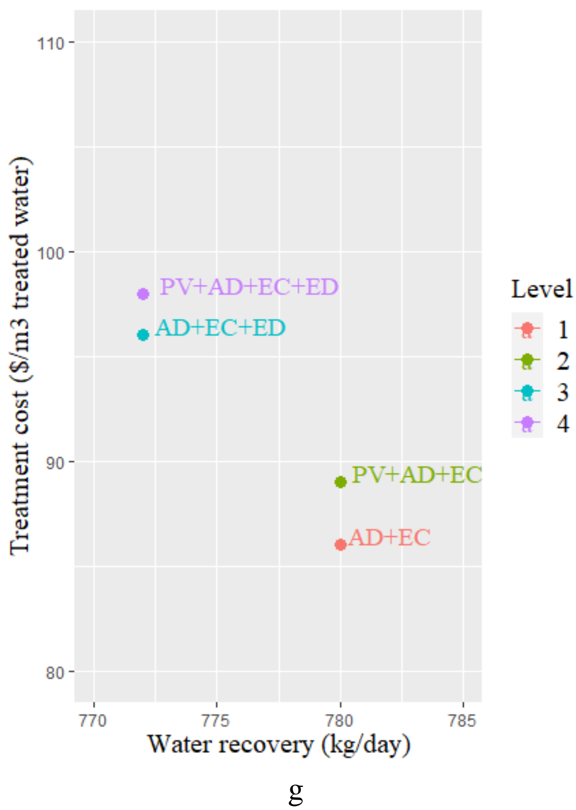


e

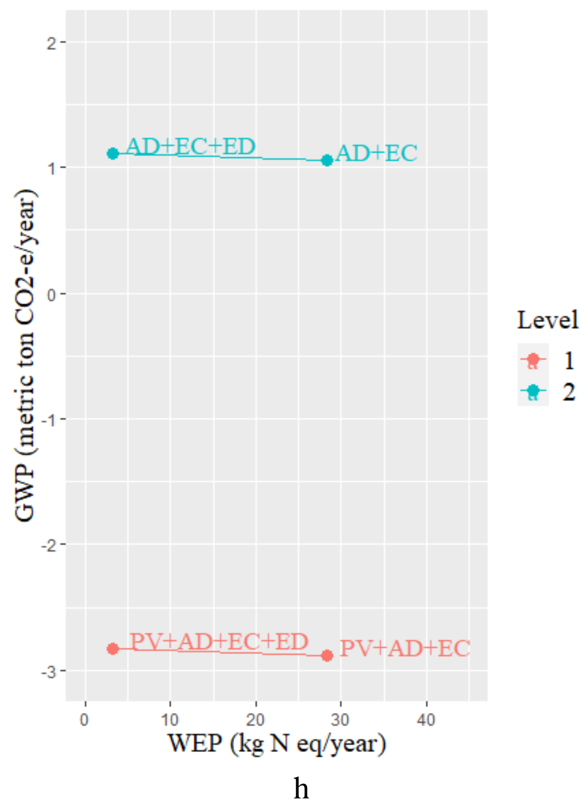
Figure 27 (cont'd)



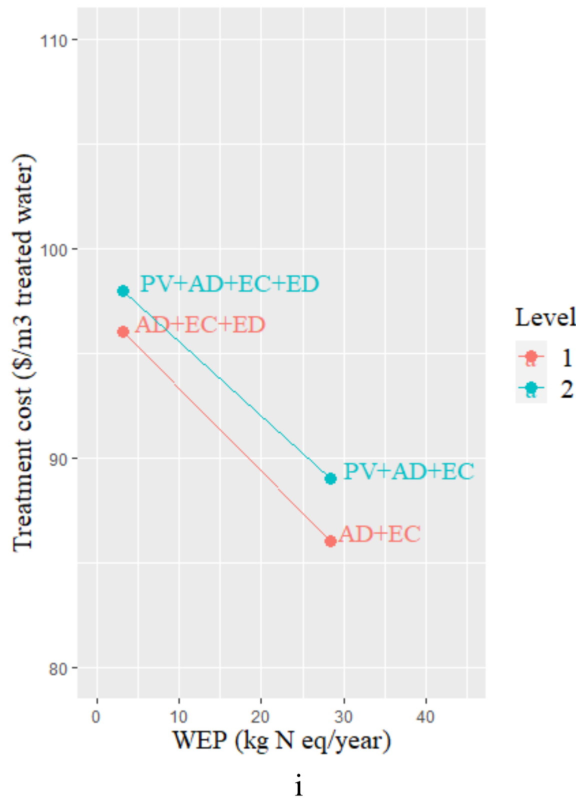
f



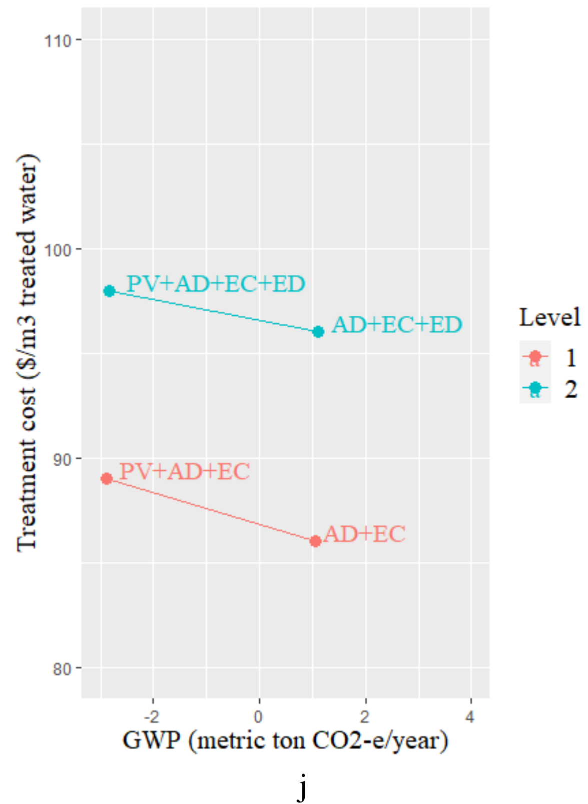
g



h



i



j

4. Conclusions

This study comprehensively analyzed and optimized combinations of PV, AD, EC, and ED technologies to develop decentralized blackwater and food waste co-treatment systems. The results concluded that synergistic integration of these technologies can conclude optimized treatment combinations with good water quality, carbon neutrality, and positive energy output. The multi-objective optimization concluded the preferred combinations to achieve five objective factors of water quality, water recovery, net energy output, GWP, WEP, and treatment cost as many as possible. If water quality, energy output, GWP, and WEP are the priorities, the combination of PV+AD+EC+ED is the preferred one to carry out the treatment. Meanwhile, this study also concluded that the multiple-objective optimization approach is a useful tool to integrate different treatment modules and conclude the best decentralized waste and wastewater treatment system.

CONCLUSIONS AND FUTURE WORK

1. Conclusions

The results of the studies conducted in this dissertation show that the decentralized wastewater management strategies investigated herein are effective, environmentally friendly, and economically feasible. The source separation of wastewaters into two streams: greywater and blackwater allows for the optimization of technology integration which resulted in better performance, improved energy efficiency, and a lower impact on the environment. The recycling of greywater has been proven to be a good strategy to reduce water demands in remote environments while minimizing the environmental impacts of discharging wastewater. It has been demonstrated that blackwater can be treated to be safely discharged into the environment with simple and effective biological treatment technologies. Utilizing AD for energy generation on the blackwater sludge has also proven to be economically effective and environmentally sound.

The LCIA of the decentralized wastewater treatment strategy demonstrated that integrating activated sludge, AD, and UF/RO filtration to separately treat blackwater and greywater led to the optimal treatment process with a water recovery efficiency of 99.9% and trace nutrient and PPCP concentrations in the recycled water (106 ug/L of PPCPs, 0.9 mg/L of TN, 0.04 mg/L of TP, and 3 mg/L of COD). This treatment strategy also resulted in the lowest net energy demand (4.2 kWh-e/m^3 recycled water). For global warming potential (GWP), Treatment B resulted in the lowest values of 0.16 and 2.08 kg O^3/m^3 recycled water for natural gas-based and diesel electricity, respectively, among all five treatment and control scenarios. Treatment B also has the lowest number of 24 g N eq/m^3 recycled water among all treatment and control scenarios. According to the distribution of TN and TP in the discharge water and sludge

of each treatment and control scenario, the discharge of the activated sludge and the digestion sludge had a much larger impact than the recycled and discharged water. The Eco-Toxicity analysis elucidates that biological treatments (activated sludge and anaerobic digestion) of greywater and blackwater can effectively remove PPCPs and lead to less eco-toxicity impact on the environment. The analysis also shows that the discharge water had a much larger impact on the eco-toxicity than the digestion sludge or the activated sludge. The economic analysis further underscores the feasibility of this approach, revealing varying treatment costs from different energy sources, \$3.31/m^3 wastewater and \$3.81/m^3 wastewater, for diesel electricity and natural gas electricity, respectively.

Focusing on the blackwater treatment component of the decentralized wastewater treatment strategy, the baffled bioreactor proved to be an efficient and effective treatment method for blackwater to produce water that could be discharged into the environment. The study concluded the baffled bioreactor enhanced microbial communities that facilitated the removal of total solids, and inorganic and organic nitrogen. Increasing feed amount in the range of 3000-4500 LPD improved the treatment performance. The microbial communities in the baffled bioreactor were analyzed to determine community differences at the different operational conditions. The NMDS analysis revealed that an increase in the feed amount enhanced the relative abundance of Verrucomicrobiaceae, unclassified Sphingomonadales, and unclassified Burkholderiales in the community, which also facilitated the removal of TS, TKN, and NO3-. The results demonstrate that the design of the reactor configuration increased the retention time of the activated sludge and further enabled and enhanced the treatment performance under higher feed amounts (higher organic loading). Based on the exergy destruction and other exergy values, universal exergy efficiencies were calculated using the Equations from section 2.7. Universal

exergy efficiency, which accounts for total mass inflows and outflows (the difference between them is the exergy destruction), increased from 51 to 61% with feed amount increasing from 3000 to 3750 LPD and did not show any considerable difference between 3750 and 4500 LPD. However, exergy rates of the treated water were increased with the increase in feed amount. Therefore, considering mass and energy balance and exergy efficiency, it is concluded that 3750 LPD is the preferred feed amount among the tested feed amounts to treat the blackwater.

In order to optimize greywater recycling for the decentralized strategy, an investigation into the performance and fouling of ultrafiltration membranes for direct filtration of greywater was conducted. The results elucidated that among the three membranes, PPG had the fastest flux for all three wastewaters. PPG also accumulated the least surface mass (fouling) compared to PVDF and PES on individual wastewaters, not including the combined shower/laundry. PES accumulated less surface mass than PPG and PVDF on the combined shower/laundry wastewater. While the PPG membrane demonstrated superior flux and fouling resistance across most wastewater types, PVDF and PES exhibited enhanced nutrient and chemical removal performance. The results of this investigation show that each specific water type can benefit from a unique treatment technology selection, even different materials of a certain class of membrane can have dramatic differences on different greywater sources. The employment of multiple objective optimization (MOO) was shown to be useful in the selection and optimization of treatment technologies for specific decentralized wastewater treatment operations.

A multi-objective optimization (MOO) approach was adopted in this study to carry out the optimization and selection of suitable treatment combinations for decentralized wastewater treatment including four modules: electrocoagulation (EC) treatment of blackwater, anaerobic digestion (AD) for treatment of food waste and EC sludge, electrodialysis (ED) membrane

treatment for final water treatment, electricity generation from biogas and photovoltaic (PV) solar energy for additional electricity generation. The combination of PV+AD+EC demonstrates the best performance on three vectors of water recovery (780 kg treated water/day), energy output (-16.4 kWh-e/day), and GWP (-2.88 Metric ton CO_2-e/year) (Figure 4e & h). Since the combination does not include ED, it had poor performance on the water quality (not satisfying the EPA discharging standards) and WEP (28 N-eq/year). The combination of PV+AD+EC+ED indicates the best performance on two vectors of water quality (satisfying the EPA discharging standards) and WEP (3.14 kg N-eq/year). It also performed well on GWP (-2.83 Metric ton CO2-e/year) and energy output (-15.8 kWh-e/day), even though they are slightly lower than the combination of PV+AD+EC. Due to the fact that all four modules are included in this combination, it had the highest treatment cost ($98/$m^3$ treated water) among the four combinations.

2. Future work

Despite the promising potential of the investigated decentralized wastewater treatment scenarios and technologies, additional investigations would benefit the future integration and adoption of decentralized wastewater treatment and utilization. Increasing the performance and decreasing the cost of these technologies represents an area where future research can have a large impact. While the greywater fouling study on ultrafiltration membranes showed the characteristics of the fouling layer, further research should be conducted to determine optimal cleaning protocols to remove this fouling and maintain optimal treatment performance of the membranes. There are also different operational conditions that could be utilized to minimize fouling on the ultrafilters.

Energy generation is a key component of decentralized wastewater management strategies, and this study focuses on carbon for renewable energy generation. Nitrogen is another potential resource for renewable energy that can be utilized from wastewater sources. Ammonia, as a carbon-free molecule, is a great green fuel candidate. It has several main advantages compared to its primary competitor - hydrogen, such as better energy density, easier and safer storage/distribution, and more versatile fuel applications. However, current green ammonia (carbon-free ammonia) production from the Haber-Bosch reaction has several major challenges: high energy demand (electrolysis of hydrogen production), low energy conversion rate, and high production cost. On the other hand, a large amount of anthropogenic nitrogen (urea, ammonia, and residual proteins) as waste is released into the environment. The nitrogen in wastewater can be converted to ammonia much easier than the ammonia synthesis from nitrogen and hydrogen. Current wastewater treatment practices apply biological processes of nitrification and denitrification to degrade those nitrogen compounds and release nitrogen gas and clean water into the environment. New pathways are needed to efficiently convert and utilize those nitrogen-based compounds in waste streams. The electro-dialysis (ED) treatment that was utilized in Chapter 4 has the potential to recover and concentrate ammonia producing a high ammonia stream (up to 10 -15 g/L) and generating clean water.

This research emphasizes the important potential of decentralized wastewater management, offering not only a method to address water scarcity and renewable energy resources, but also to mitigate environmental impacts associated with wastewater generation and treatment. By utilizing wastewater as a resource instead of a liability, this decentralized wastewater management system can contribute to sustainable water management practices in

remote locations and develop a path for a more resilient and environmentally focused approach to wastewater management for rural communities.

REFERENCES

1. Weerasekara, P., *The United Nations World Water Development Report 2017 Wastewater: The Untapped Resource.* Future of Food-Journal on Food Agriculture and Society, 2017. **5**(2): p. 80-81.

2. Group, A.R., Water and Waste Water Treatment Opportunity in India. 2011. p. 36.

3. Rockefeller, A.A. and R. Goodland, *What is environmental sustainability in sanitation?* Toward sustainable sanitation: What is sustainability in sanitation?, 2001: p. 7-16.

4. Metcalf & Eddy. et al., Wastewater Engineering: Treatment and Resource Recovery. 5 ed. 2013: McGraw Hill. 2018.

5. US EPA. (2016). Clean Watersheds Needs Survey 2012. In Report to Congress (EPA-830-R-15005).

6. Krop, R. A., Hernick, C., Frantz, C., 2008. Mayors Water Council, The United States Conference of Mayors.

7. Tchobanoglous, G., Stensel, H., Tsuchihashi, R., Burton, F., 2013. Wastewater Engineering: Treatment and Resource Recovery, 5 ed. McGraw Hill.

8. Mutamim, N.S.A., Noor, Z.Z., Abu Hassan, M.A., Olsson, G., 2012. Application of membrane bioreactor technology in treating high strengthindustrial wastewater: a performance review. Desalination305, 1–11. https://doi.org/10.1016/j.desal.2012.07.033.

9. Xu, S., Yao, J., Ainiwaer, M., Hong, Y., Zhang, Y., 2018. Analysis of bacterial community structure of activated sludge from wastewater treatment plants in winter. BioMedRes. Int., 2018. https://doi.org/10.1155/2018/8278970.

10. Seviour, R., Nielsen, P.H., 2010. Microbial Ecology of Activated Sludge: the International Water Association. (IWA) Publishing.

11. Oteng-Peprah, M., M.A. Acheampong, and N.K. deVries, *Greywater Characteristics, Treatment Systems, Reuse Strategies and User Perception-a Review.* Water Air and Soil Pollution, 2018. **229**(8): p. 16.

12. Hernandez-Leal, L., et al., *Grey water treatment concept integrating water and carbon recovery and removal of micropollutants.* Water Practice and Technology, 2011. **6**(2).

13. Nielsen, P.H., 2017. Microbial biotechnology and circulareconomy in wastewater treatment. Microb. Biotechnol. 10(5), 1102–1105. https://doi.org/10.1111/1751-7915.12821.

14. Corominas, L., Foley, J., Guest, J.S., Hospido, A., Larsen, H.F., Morera, S., Shaw, A., 2013. Life cycle assessment applied to wastewater treatment: State of the art. Water Research 47(15), 5480-5492.

15. Friedrich, E., Pillay, S., Buckley, C.A., 2007. The use of LCA in the water industry and the case for an environmental performance indicator. Water Sa 33(4), 443-451.

16. Kobayashi, Y., Ashbolt, N.J., Davies, E.G.R., Liu, Y., 2020. Life cycle assessment of decentralized greywater treatment systems with reuse at different scales in cold regions. Environment International 134.

17. Risch, E., Boutin, C., Roux, P., 2021. Applying life cycle assessment to assess the environmental performance of decentralised versus centralised wastewater systems. Water Research 196.

18. Sharvini, S.R., Noor, Z.Z., Kamaruddin, S.N., Sabli, N.S.M., Sabeen, A.H., Aris, A., Yong, E.L., Graham, D.W., 2022. Environmental impact evaluation of decentralized sewage treatment technologies: A life cycle assessment approach. Water and Environment Journal 36(2), 261-269.

19. APHA, 2012. Standard methods for the examination of water and wastewater, 22nd ed. American Public Health Association, American Water Works Association, Water Environment Federation, Washington, D.C.

20. Thomas, B.D., Uludag-Demirer, S., Frost, H., Liu, Y., Dusenbury, J.S., Liao, W., 2022. Decentralized high-strength wastewater treatment using a compact aerobic baffled bioreactor. Journal of Environmental Management 305.

21. Kanafin, Y.N., Kakimov, Y., Adamov, A., Makhatova, A., Yeshmuratov, A., Poulopoulos, S.G., Inglezakis, V.J., Arkhangelsky, E., 2021. The effect of caffeine, metronidazole, and ibuprofen on continuous flow activated sludge process. Journal of Chemical Technology and Biotechnology 96(5), 1370-1380.

22. Lim, J.-W., Tan, J.-Z., Seng, C.-E., 2013. Performance of phenol-acclimated activated sludge in the presence of various phenolic compounds. Applied Water Science 3(2), 515-525.

23. Santos, A., Reif, R., Hillis, P., Judd, S.J., 2011. Fate and removal of permethrin by conventional activated sludge treatment. Environmental Technology 32(12), 1367-1373.

24. Masunaga, S., Urushigawa, Y., Yonezawa, Y., 1986. BIODEGRADATION PATHWAY OF ORTHO-CRESOL BY HETEROGENEOUS CULTURE. Water Research 20(4), 477-484.

25. Dhulipala, R., Reddy, V.V., Reddy, G., 2016. Estimation and Biodegradation of Acetylsalicylic Acid in Contaminated Waste Water using ASP and UASB. International Journal of Science and Research (IJSR) 5(12), 1842-1847.

26. Wang, M., Yang, G., Min, H., Lv, Z., Jia, X., 2009. Bioaugmentation with the nicotine-degrading bacterium Pseudomonas sp HF-1 in a sequencing batch reactor treating tobacco wastewater: Degradation study and analysis of its mechanisms. Water Research 43(17), 4187-4196.

27. Sui, Q., Huang, J., Deng, S., Yu, G., Fan, Q., 2010. Occurrence and removal of pharmaceuticals, caffeine and DEET in wastewater treatment plants of Beijing, China. Water Research 44(2), 417-426.

28. Wang, D., Ji, M., Wang, C., 2014. DEGRADATION OF ORGANIC POLLUTANTS AND CHARACTERISTICS OF ACTIVATED SLUDGE IN AN ANAEROBIC/ANOXIC/OXIC REACTOR TREATING CHEMICAL INDUSTRIAL WASTEWATER. Brazilian Journal of Chemical Engineering 31(3), 703-713.

29. Sahledemessie, E., 2018. Comparative study on the performance of Anaerobic and Aerobic Biotrickling Filter for the Removal of Chloroform. Environmental Dataset Gateway.

30. Quesnel, D., Nakha, G., 2006. Removal kinetics of acetone and MIBK from a complex industrial wastewater by an acclimatized activated sludge. Journal of Hazardous Materials 132(2-3), 253-260.

31. Marttinen, S.K., Kettunen, R.H., Sormunen, K.M., Rintala, J.A., 2003. Removal of bis(2-ethylhexyl) phthalate at a sewage treatment plant. Water Research 37(6), 1385-1393.

32. Chen, R., Jiang, H., Li, Y.-Y., 2018. Caffeine degradation by methanogenesis: Efficiency in anaerobic membrane bioreactor and analysis of kinetic behavior. Chemical Engineering Journal 334, 444-452.

33. Ramakrishnan, A., Gupta, S.K., 2006. Anaerobic biogranulation in a hybrid reactor treating phenolic waste. Journal of Hazardous Materials 137(3), 1488-1495.

34. Kirk, P.W.W., Rogers, H.R., Lester, J.N., 1989. THE FATE OF CHLOROBENZENES AND PERMETHRINS DURING ANAEROBIC SEWAGE-SLUDGE DIGESTION. Chemosphere 18(9-10), 1771-1784.

35. Meher, K.K., Panchwagh, A.M., Rangrass, S., Gollakota, K.G., 1995. BIOMETHANATION OF TOBACCO WASTE. Environmental Pollution 90(2), 199-202.

36. Narumiya, M., Nakada, N., Yamashita, N., Tanaka, H., 2013. Phase distribution and removal of pharmaceuticals and personal care products during anaerobic sludge digestion. Journal of Hazardous Materials 260, 305-312.

37. Yumihara, K., Shigematsu, T., Hamada, K., Morimura, S., Kida, K., 2002. Anaerobic Degradation of Terephthalic Acid and Aniline by Methanogenic Consortia. Japanese Journal of Water Treatment Biology 38(1), 1-9.

38. Carballa, M., Omil, F., Ternes, T., Lema, J.M., 2007. Fate of pharmaceutical and personal care products (PPCPs) during anaerobic digestion of sewage sludge. Water Research 41(10), 2139-2150.

39. Zhang, J., Wei, Y., Xiao, W., Zhou, Z., Yan, X., 2011. Performance and spatial community succession of an anaerobic baffled reactor treating acetone-butanol-ethanol fermentation wastewater. Bioresource Technology 102(16), 7407-7414.

40. Yang, J., Speece, R.E., 1986. THE EFFECTS OF CHLOROFORM TOXICITY ON METHANE FERMENTATION. Water Research 20(10), 1273-1279.

41. Urtiaga, A.M., Perez, G., Ibanez, R., Ortiz, I., 2013. Removal of pharmaceuticals from a WWTP secondary effluent by ultrafiltration/reverse osmosis followed by electrochemical oxidation of the RO concentrate. Desalination 331, 26-34.

42. Amat, N.A.A., Tan, Y.H., Lau, W.J., Lai, G.S., Ong, C.S., Mokhtar, N.M., Sani, N.A.A., Ismail, A.F., Goh, P.S., Chong, K.C., Lai, S.O., 2015. Tackling colour issue of anaerobically-treated palm oil mill effluent using membrane technology. Journal of Water Process Engineering 8, 221-226.

43. Cailean, D., Barjoveanu, G., Teodosiu, C., Pintilie, L., Dascalescu, I.G., Paduraru, C., 2015. Technical performances of ultrafiltration applied to municipal wastewater treatment plant effluents. Desalination and Water Treatment 56(6), 1476-1488.

44. Acero, J.L., Javier Benitez, F., Leal, A.I., Real, F.J., Teva, F., 2010. Membrane filtration technologies applied to municipal secondary effluents for potential reuse. Journal of Hazardous Materials 177(1-3), 390-398.

45. Bloetscher, F., Stambaugh, D., Hart, J., Cooper, J., Kennedy, K., Burack, L.S., Ruffini, A.P., Cicala, A., Cimenello, S., 2011. Evaluating Membrane Options for Aquifer Recharge in Southeast Florida. IDA Journal of Desalination and Water Reuse 3(4), 46-57.

46. Fang, C., Chu, Y., Jiang, L., Wang, H., Long, Y., Shen, D., 2018. Removal of phthalic acid diesters through a municipal solid waste landfill leachate treatment process. Journal of Material Cycles and Waste Management 20(1), 585-591.

47. Al-Rifai, J.H., Khabbaz, H., Schaefer, A.I., 2011. Removal of pharmaceuticals and endocrine disrupting compounds in a water recycling process using reverse osmosis systems. Separation and Purification Technology 77(1), 60-67.

48. Comerton, A.M., Andrews, R.C., Bagley, D.M., Hao, C., 2008. The rejection of endocrine disrupting and pharmaceutically active compounds by NF and RO membranes as a function of compound a water matrix properties. Journal of Membrane Science 313(1-2), 323-335.

49. Abdel-Karim, A., Gad-Allah, T.A., Badawy, M.I., Khalil, A.S.G., Ulbricht, M., 2017. Removal of humic acid and chloroform from drinking water by using commercial nanofiltration and reverse osmosis membranes. Desalination and Water Treatment 59, 48-54.

50. Speece, R.E., 1983. ANAEROBIC BIOTECHNOLOGY FOR INDUSTRIAL WASTEWATER-TREATMENT. Environmental Science & Technology 17(9), A416-A427.

51. Bare, J., 2011. TRACI 2.0: the tool for the reduction and assessment of chemical and other environmental impacts 2.0. Clean Technologies and Environmental Policy 13(5), 687-696.

52. Henze, M., & Comeau, Y. (2008). Wastewater Characterization. Biological Wastewater Treatment: Principles, Modelling and Design, 33-52.

53. Mutamim, N. S. A., Noor, Z. Z., Abu Hassan, M. A., & Olsson, G. (2012). Application of membrane bioreactor technology in treating high strength industrial wastewater: a performance review. Desalination, 305, 1-11. doi:10.1016/j.desal.2012.07.033

54. Kalemba, K. (2020). CIRCULAR ECONOMY IN WASTEWATER TREATMENT PLANT. Architecture Civil Engineering Environment, 13(4), 93-97. doi:10.21307/acee-2020-033

55. Nielsen, P. H. (2017). Microbial biotechnology and circular economy in wastewater treatment. Microbial Biotechnology, 10(5), 1102-1105. doi:10.1111/1751-7915.12821

56. Xu, S., Yao, J., Ainiwaer, M., Hong, Y., & Zhang, Y. (2018). Analysis of Bacterial Community Structure of Activated Sludge from Wastewater Treatment Plants in Winter. BioMed Research International, 2018. doi:10.1155/2018/8278970

57. Seviour, R., & Nielsen, P. H. (2010). Microbial Ecology of Activated Sludge: The International Water Association (IWA) Publishing.

58. Chan, Y. J., Chong, M. F., Law, C. L., & Hassell, D. G. (2009). A review on anaerobic-aerobic treatment of industrial and municipal wastewater. Chemical Engineering Journal, 155(1-2), 1-18. doi:10.1016/j.cej.2009.06.041

59. Liu, G. Q., Canter, T., Wang, D. M., Wang, J. M., Fitch, M. W., & Burken, J. (2012). Baffled Bioreactor for Municipal Wastewater Treatment. Journal of Environmental Engineering-Asce, 138(3), 239-247. doi:10.1061/(asce)ee.1943-7870.0000410

60. Tucker, R. E., Liu, G., & Wang, J. (May-August 2016). Wasting Less Water: Deployable Wastewater Treatment System Tested at Fort Leonard Wood. In (Vol. Engineer: The Professional Bulletin for Army Engineers): US Army.

61. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N., Abnet, C. C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852-857. doi:10.1038/s41587-019-0209-9

62. Fitzsimons, L., Horrigan, M., McNamara, G., Doherty, E., Phelan, T., Corcoran, B., . . . Clifford, E. (2016). Assessing the thermodynamic performance of Irish municipal wastewater treatment plants using exergy analysis: a potential benchmarking approach. Journal of Cleaner Production, 131, 387-398. doi:10.1016/j.jclepro.2016.05.016

63. Tenuta, M., Bergstrom, D. W., & Beauchamp, E. G. (2000). Denitrifying enzyme activity and carbon availability for denitrification following manure application. Communications in Soil Science and Plant Analysis, 31(7-8), 861-876.

64. Zielinska, M., & Wojnowska-Baryla, I. (2006). Effect of volumetric organic loading on the nitrogen removal rate by immobilised activated sludge. Environmental Technology, 27(5), 559-564. doi:10.1080/09593332808618668

65. Shen, N., & Zhou, Y. (2016). Enhanced biological phosphorus removal with different carbon sources. Applied Microbiology and Biotechnology, 100(11), 4735-4745. doi:10.1007/s00253-016-7518-4

66. Secondary Treatment Regulation. (40 CFR 133). (2011).

67. Balkwill, D. L., Drake, G. R., Reeves, R. H., Fredrickson, J. K., White, D. C., Ringelberg, D. B., . . . Spadoni, C. M. (1997). Taxonomic study of aromatic-degrading bacteria from deep-terrestrial-subsurface sediments and description of Sphingomonas aromaticivorans sp nov, Sphingomonas subterranea sp nov, and Sphingomonas stygia sp nov. International Journal of Systematic Bacteriology, 47(1), 191-201. doi:10.1099/00207713-47-1-191

68. Pang, C. M., & Liu, W.-T. (2007). Community structure analysis of reverse osmosis membrane biofilms and the significance of Rhizobiales bacteria in biofouling. Environmental Science & Technology, 41(13), 4728-4734. doi:10.1021/es0701614

69. White, C. P., DeBry, R. W., & Lytle, D. A. (2012). Microbial Survey of a Full-Scale, Biologically Active Filter for Treatment of Drinking Water. Applied and Environmental Microbiology, 78(17), 6390-6394. doi:10.1128/aem.00308-12

70. Ishii, S., Yamamoto, M., Kikuchi, M., Oshima, K., Hattori, M., Otsuka, S., & Senoo, K. (2009). Microbial Populations Responsive to Denitrification-Inducing Conditions in Rice Paddy Soil, as Revealed by Comparative 16S rRNA Gene Analysis. Applied and Environmental Microbiology, 75(22), 7070-7078. doi:10.1128/aem.01481-09

71. Juretschko, S., Loy, A., Lehner, A., & Wagner, M. (2002). The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. Systematic and Applied Microbiology, 25(1), 84-99. doi:10.1078/0723-2020-00093

72. Duan, Z.-h., Pan, L.-m., Chen, X.-o., Wang, X.-d., Zhao, L.-j., & Tian, L.-q. (2016). Changes of Microbial Community Structure in Activated Sludge Bulking at Low Temperature. Huanjing Kexue, 37(3), 1070-1074.

73. Yoon, J. (2014). The Family Verrucomicrobiaceae. In E. Rosenberg, E. F. Delong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), The Prokaryotes. Berlin, Heidelberg: Springer.

74. Wertz, J. T., Kim, E., Breznak, J. A., Schmidt, T. M., & Rodrigues, J. L. M. (2012). Genomic and Physiological Characterization of the Verrucomicrobia Isolate Diplosphaera colitermitum gen. nov., sp nov., Reveals Microaerophily and Nitrogen Fixation Genes. Applied and Environmental Microbiology, 78(5), 1544-1555. doi:10.1128/aem.06466-11

75. Gikas, P. (2017). Towards energy positive wastewater treatment plants. In (Vol. 203, pp. 621-629). Journal of environmental management.

76. Mizuta, K., & Shimada, M. (2010). Benchmarking energy consumption in municipal wastewater treatment plants in Japan. Water Science and Technology, 62(10), 2256-2262. doi:10.2166/wst.2010.510

77. Hellstrom, D. (1997). An exergy analysis for a wastewater treatment plant - An estimation of the consumption of physical resources. Water Environment Research, 69(1), 44-51. doi:10.2175/106143097x125173

78. Seckin, C., & Bayulken, A. R. (2013). Extended Exergy Accounting (EEA) analysis of municipal wastewater treatment - Determination of environmental remediation cost for municipal wastewater. Applied Energy, 110, 55-64. doi:10.1016/j.apenergy.2013.04.042

79. Salehi, M. (2022). Global water shortage and potable water safety; Today's concern and tomorrow's crisis. Environment International, 158, 106936. https://doi.org/10.1016/j.envint.2021.106936

80. Josefine Lund Schlamovitz & Per Becker (2021) Differentiated vulnerabilities and capacities for adaptation to water shortage in Gaborone, Botswana, International Journal

of Water Resources Development, 37:2, 278-299, DOI: 10.1080/07900627.2020.1756752

81. Casanova, L. M., Gerba, C. P., & Karpiscak, M. (2001). Chemical and microbial characterization of household greywater. Journal of environmental science and health, Part A. Toxic/Hazardous Substances and Environmental Engineering, 34, 395–401.

82. Nghiem, L. D., Oschmann, N., & Schäfer, A. (2006). Fouling in greywater recycling by direct ultrafiltration. Desalination, 187(1–3), 283–290. https://doi.org/10.1016/j.desal.2005.04.087

83. Wu, B. (2019). Membrane-based technology in greywater reclamation: A review. Science of the Total Environment, 656, 184–200. https://doi.org/10.1016/j.scitotenv.2018.11.347

84. Oschmann, N., Nghiem, L. D., & Schäfer, A. (2005). Fouling mechanisms of submerged ultrafiltration membranes in greywater recycling. Desalination, 179(1–3), 215–223. https://doi.org/10.1016/j.desal.2004.12.023

85. Schäfer, A., Nghiem, L. D., & Oschmann, N. (2006). Bisphenol A retention in the direct ultrafiltration of greywater. Journal of Membrane Science, 283(1–2), 233–243. https://doi.org/10.1016/j.memsci.2006.06.035

86. https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/spectro.htm#contnt

87. Hussain, J., Wang, X., Sousa, L., Ali, R., Rittmann, B. E., Liao, W. 2020. Using non-metric multi-dimensional scaling analysis and multi-objective optimization to evaluate green algae for production of proteins, carbohydrates, lipids, and simultaneously fix carbon dioxide. Biomass and Bioenergy 141, 105711.

88. Tušar T, Filipič B. Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. IEEE Trans Evol Comput. 2014;19(2):225–245.

89. Massoud, M. A., Tarhini, A., & Nasr, J. (2009). Decentralized approaches to wastewater treatment and management: Applicability in developing countries. Journal of Environmental Management, 90(1), 652–659.

90. Aviraj, D., Mukund, P., & Wani, S. P. (2018). Improving rural wastewater management. Corporate social responsibility: win-win propositions for communities, corporates and agriculture, 213-228. https://doi.org/10.1079/9781786394514.0213

91. California Regional Water Quality Control Board. (2018). Draft General Waste Discharge Requrements Order No. R1-2019-0001 for dairies in the North Coast region. In www.waterboards.ca.gov (RS-2018-0001). State of Claifornia. Retrieved April 5, 2024, from

https://www.waterboards.ca.gov/northcoast/water_issues/programs/dairies/pdf/190625/Public%20comments%20on%20May%2014,%202019%20new%20Draft%20Order%20package.pdf

92. US EPA. (2000). Decentralized Systems Technology Fact Sheet - Aerobic treatment. In www.epa.gov (EPA 832-F-00-031). United States Environmental Protection Agency. Retrieved April 5, 2024, from https://www.epa.gov/sites/default/files/2015-06/documents/aerobic_treatment_0.pdf

93. R.B. Dean, PROCESSES FOR WATER RECLAMATION, Waste Manage. Res. 9(5) (1991) 425-430.

94. V.K. Tyagi, A.A. Khan, A.A. Kazmi, A.K. Chopra, Enhancement of Coagulation Flocculation Process Using Anionic Polymer for the Post Treatment of UASB Reactor Effluent, Sep. Sci. Technol. 45(5) (2010) 626-634.

95. N. Jaafarzadeh, M. Omidinasab, F. Ghanbari, Combined electrocoagulation and UV-based sulfate radical oxidation processes for treatment of pulp and paper wastewater, Process Safety and Environmental Protection 102 (2016) 462-472.

96. K. Yetilmezsoy, F. Ilhan, Z. Sapci-Zengin, S. Sakar, M.T. Gonullu, Decolorization and COD reduction of UASB pretreated poultry manure wastewater by electrocoagulation process: A post-treatment study, Journal of Hazardous Materials 162(1) (2009) 120-132.

97. M. Kobya, E. Senturk, M. Bayramoglu, Treatment of poultry slaughterhouse wastewaters by electrocoagulation, Journal of Hazardous Materials 133(1-3) (2006) 172-176.

98. C. Ricordel, A. Darchen, D. Hadjiev, Electrocoagulation-electroflotation as a surface water treatment for industrial uses, Separation and Purification Technology 74(3) (2010) 342-347.

99. M.M. Emamjomeh, M. Sivakumar, Review of pollutants removed by electrocoagulation and electrocoagulation/flotation processes, Journal of Environmental Management 90(5) (2009) 1663-1679.

100. D. Ghernaout, A. Badis, A. Kellil, B. Ghernaout, Application of electrocoagulation in Escherichia coli culture and two surface waters, Desalination 219(1-3) (2008) 118-125.

101. E. Butler, Y.-T. Hung, R.Y.-L. Yeh, M.S. Al Ahmad, Electrocoagulation in Wastewater Treatment, Water 3 (2011) 495-525.

102. Z.G. Liu, D. Stromberg, X.M. Liu, W. Liao, Y. Liu, A new multiple-stage electrocoagulation process on anaerobic digestion effluent to simultaneously reclaim water and clean up biogas, Journal of Hazardous Materials 285 (2015) 483-490.

103. F. Monnet, An introduction to anaerobic digestion of organic wastes, final report., Remade Scotland, 2003.

104. A. Singh, S. Bhattacharya, C. Thinaharan, D.K. Aswal, S.K. Gupta, J.V. Yakhmi, K. Bhanumurthy, Development of low resistance electrical contacts for thermoelectric devices based on n-type PbTe and p-type TAGS-85 ((AgSbTe2)(0.15)(GeTe)(0.85)), Journal of Physics D-Applied Physics 42(1) (2009).

105. R. International, Greenhouse Gas Emissions Estimation Methodologies for Biogenic Emissions from Selected Source Categories: Solid Waste Disposal, Wastewater Treatment, Ethanol Fermentation, RTI International, EPA, 2010.

106. EPA, AP-42: Compilation of air emissions factors, United States Environmental Protection Agency, 1995.

APPENDIX A: ORIGINAL R SOFTWARE CODE

CHAPTER 2

Results of the statistical analysis of blackwater treatment – Frontier unit

1. Turbidity

```
> ## Normality check on effluent
> shapiro.test(data1$Turbidity)

        Shapiro-Wilk normality test

data:  data1$Turbidity
W = 0.89425, p-value = 0.0006307

> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$Turbidity)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.95538, p-value = 0.08139

>
> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_Turbidity")
> data111
   Feed_amount sqrt_Turbidity
1          800       3.391165
2          800       2.677686
3          800       3.500000
4          800       4.511097
5          800       4.398863
6          800       4.110961
7          800       4.549725
8          800       3.263434
9          800       3.834058
10         800       7.049823
11         800       2.505993
12         800       2.362202
13         800       5.545268
14         800       4.888763
15         800       5.877925
16         800       7.095773
17         800       6.674579
18         800       8.354639
19         800       6.220932
20         800       5.516339
21         800       1.996246
22         800       2.620115
23         800       3.464102
24         800       4.117038
25        1000       7.690904
26        1000       8.485281
27        1000       4.364631
28        1000       5.766281
29        1000       6.606815
```

```
30        1000       8.077747
31        1000       7.141428
32        1000       7.120393
33        1000       4.979960
34        1000       3.605551
35        1000       4.449719
36        1000       3.076524
37        1000       4.031129
38        1200       3.911521
39        1200       4.024922
40        1200       5.403702
41        1200       3.224903
42        1200       2.833725
43        1200       3.324154
44        1200       5.839521
45        1200       5.692100
>
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_Turbidity, data1112$sqrt_Turbidity)

        F test to compare two variances

data:  data1111$sqrt_Turbidity and data1112$sqrt_Turbidity
F = 0.87472, num df = 23, denom df = 12, p-value = 0.7514
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.288631 2.247975
sample estimates:
ratio of variances
         0.8747183

> var.test(data1111$sqrt_Turbidity, data1113$sqrt_Turbidity)

        F test to compare two variances

data:  data1111$sqrt_Turbidity and data1113$sqrt_Turbidity
F = 2.0438, num df = 23, denom df = 7, p-value = 0.335
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4617325 5.9317457
sample estimates:
ratio of variances
          2.043775

> var.test(data1112$sqrt_Turbidity, data1113$sqrt_Turbidity)

        F test to compare two variances

data:  data1112$sqrt_Turbidity and data1113$sqrt_Turbidity
F = 2.3365, num df = 12, denom df = 7, p-value = 0.2667
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5007674 8.4266045
sample estimates:
ratio of variances
          2.336495

>
> ## Normality check on data2
> shapiro.test(data2$Turbidity)
```

```
        Shapiro-Wilk normality test

data:  data2$Turbidity
W = 0.98451, p-value = 0.8011

>
> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$Turbidity, data22$Turbidity)

        F test to compare two variances

data:  data21$Turbidity and data22$Turbidity
F = 1.2266, num df = 23, denom df = 12, p-value = 0.7324
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4047537 3.1523852
sample estimates:
ratio of variances
          1.226637

> var.test(data21$Turbidity, data23$Turbidity)

        F test to compare two variances

data:  data21$Turbidity and data23$Turbidity
F = 1.752, num df = 23, denom df = 7, p-value = 0.456
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3958128 5.0848939
sample estimates:
ratio of variances
          1.751994

> var.test(data22$Turbidity, data23$Turbidity)

        F test to compare two variances

data:  data22$Turbidity and data23$Turbidity
F = 1.4283, num df = 12, denom df = 7, p-value = 0.6549
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3061171 5.1511493
sample estimates:
ratio of variances
           1.42829

>
> # One-way ANOVA
> ## Data 1 - Effluent
> fit1 <- aov(sqrt_Turbidity~Feed_amount, data111)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2  16.89   8.445   3.024 0.0593 .
Residuals   42 117.30   2.793
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

```
Fit: aov(formula = sqrt_Turbidity ~ Feed_amount, data = data111)

$Feed_amount
               diff        lwr      upr     p adj
1000-800   1.2777733 -0.1204284 2.675975 0.0793524
1200-800  -0.2401283 -1.8976936 1.417437 0.9341059
1200-1000 -1.5179016 -3.3423823 0.306579 0.1195709

>
> ## Data 2 - Feed
> fit2 <- aov(Turbidity~Feed_amount, data2)
> summary(fit2)
            Df   Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   455597  227798   0.625   0.54
Residuals   42 15318321  364722
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Turbidity ~ Feed_amount, data = data2)

$Feed_amount
              diff       lwr      upr     p adj
1000-800   155.4696 -349.7967 660.7358 0.7367588
1200-800  -141.0208 -740.0129 457.9713 0.8357106
1200-1000 -296.4904 -955.8004 362.8197 0.5239401

> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$Turbidity)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.95538, p-value = 0.08139

> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_Turbidity")
> data111
   Feed_amount sqrt_Turbidity
1          800       3.391165
2          800       2.677686
3          800       3.500000
4          800       4.511097
5          800       4.398863
6          800       4.110961
7          800       4.549725
8          800       3.263434
9          800       3.834058
10         800       7.049823
11         800       2.505993
12         800       2.362202
13         800       5.545268
14         800       4.888763
15         800       5.877925
16         800       7.095773
17         800       6.674579
18         800       8.354639
19         800       6.220932
20         800       5.516339
21         800       1.996246
22         800       2.620115
```

```
23         800       3.464102
24         800       4.117038
25        1000       7.690904
26        1000       8.485281
27        1000       4.364631
28        1000       5.766281
29        1000       6.606815
30        1000       8.077747
31        1000       7.141428
32        1000       7.120393
33        1000       4.979960
34        1000       3.605551
35        1000       4.449719
36        1000       3.076524
37        1000       4.031129
38        1200       3.911521
39        1200       4.024922
40        1200       5.403702
41        1200       3.224903
42        1200       2.833725
43        1200       3.324154
44        1200       5.839521
45        1200       5.692100
>
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_Turbidity, data1112$sqrt_Turbidity)

        F test to compare two variances

data:  data1111$sqrt_Turbidity and data1112$sqrt_Turbidity
F = 0.87472, num df = 23, denom df = 12, p-value = 0.7514
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.288631 2.247975
sample estimates:
ratio of variances
         0.8747183

> var.test(data1111$sqrt_Turbidity, data1113$sqrt_Turbidity)

        F test to compare two variances

data:  data1111$sqrt_Turbidity and data1113$sqrt_Turbidity
F = 2.0438, num df = 23, denom df = 7, p-value = 0.335
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4617325 5.9317457
sample estimates:
ratio of variances
          2.043775

> var.test(data1112$sqrt_Turbidity, data1113$sqrt_Turbidity)

        F test to compare two variances

data:  data1112$sqrt_Turbidity and data1113$sqrt_Turbidity
F = 2.3365, num df = 12, denom df = 7, p-value = 0.2667
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5007674 8.4266045
sample estimates:
```

```
ratio of variances
          2.336495

>
> ## Normality check on data2
> shapiro.test(data2$Turbidity)

        Shapiro-Wilk normality test

data:  data2$Turbidity
W = 0.98451, p-value = 0.8011

>
> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$Turbidity, data22$Turbidity)

        F test to compare two variances

data:  data21$Turbidity and data22$Turbidity
F = 1.2266, num df = 23, denom df = 12, p-value = 0.7324
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4047537 3.1523852
sample estimates:
ratio of variances
          1.226637

> var.test(data21$Turbidity, data23$Turbidity)

        F test to compare two variances

data:  data21$Turbidity and data23$Turbidity
F = 1.752, num df = 23, denom df = 7, p-value = 0.456
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3958128 5.0848939
sample estimates:
ratio of variances
          1.751994

> var.test(data22$Turbidity, data23$Turbidity)

        F test to compare two variances

data:  data22$Turbidity and data23$Turbidity
F = 1.4283, num df = 12, denom df = 7, p-value = 0.6549
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3061171 5.1511493
sample estimates:
ratio of variances
           1.42829

>
> # One-way ANOVA
> ## Data 1 - Effluent
> fit1 <- aov(sqrt_Turbidity~Feed_amount, data111)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2  16.89   8.445   3.024 0.0593 .
Residuals   42 117.30   2.793
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sqrt_Turbidity ~ Feed_amount, data = data111)

$Feed_amount
                diff        lwr      upr     p adj
1000-800   1.2777733 -0.1204284 2.675975 0.0793524
1200-800  -0.2401283 -1.8976936 1.417437 0.9341059
1200-1000 -1.5179016 -3.3423823 0.306579 0.1195709

>
> ## Data 2 - Feed
> fit2 <- aov(Turbidity~Feed_amount, data2)
> summary(fit2)
            Df   Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   455597  227798   0.625   0.54
Residuals   42 15318321  364722
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Turbidity ~ Feed_amount, data = data2)

$Feed_amount
               diff       lwr      upr     p adj
1000-800   155.4696 -349.7967 660.7358 0.7367588
1200-800  -141.0208 -740.0129 457.9713 0.8357106
1200-1000 -296.4904 -955.8004 362.8197 0.5239401



> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=Turbidity)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent Turbidity (NTU)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="Turbidity",
+                            groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount Turbidity       sd
1         800  23.24833 17.13784
2        1000  36.72038 21.38197
3        1200  19.58500 10.60933
```

```
> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=Turbidity)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent Turbidity (NTU)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="Turbidity",
+                            groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount Turbidity       sd
1         800  1666.646 645.3715
2        1000  1822.115 582.7091
3        1200  1525.625 487.5774
```

2. TS

```
        Shapiro-Wilk normality test

data:  data1$TS
W = 0.94528, p-value = 0.01431

> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$TS)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.959, p-value = 0.05822

> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_TS")
> data111
   Feed_amount  sqrt_TS
1          800 27.47726
2          800 28.01785
3          800 25.69047
4          800 27.11088
5          800 34.05877
6          800 29.24038
7          800 24.28992
8          800 30.16621
9          800 28.98275
10         800 28.98275
11         800 29.66479
12         800 27.20294
13         800 26.26785
14         800 30.90307
15         800 31.62278
16         800 30.49590
17         800 30.90307
18         800 32.86335
```

```
19          800 29.83287
20          800 29.15476
21          800 33.31666
22          800 30.74085
23          800 30.16621
24          800 29.83287
25          800 29.66479
26          800 26.36285
27          800 29.06888
28          800 25.88436
29          800 31.54362
30          800 26.07681
31          800 27.65863
32          800 26.07681
33          800 27.74887
34          800 26.36285
35          800 25.88436
36          800 26.36285
37         1000 27.38613
38         1000 28.54820
39         1000 27.74887
40         1000 26.07681
41         1000 27.01851
42         1000 28.63564
43         1000 27.29469
44         1000 28.28427
45         1000 27.83882
46         1000 26.55184
47         1000 27.74887
48         1000 26.26785
49         1200 32.93934
50         1200 26.73948
51         1200 31.78050
52         1200 25.39685
53         1200 28.10694
54         1200 25.88436
55         1200 25.00000
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_TS, data1112$sqrt_TS)

        F test to compare two variances

data:  data1111$sqrt_TS and data1112$sqrt_TS
F = 7.8982, num df = 35, denom df = 11, p-value = 0.0008536
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  2.55967 18.84970
sample estimates:
ratio of variances 
          7.898195 

> var.test(data1111$sqrt_TS, data1113$sqrt_TS)

        F test to compare two variances

data:  data1111$sqrt_TS and data1113$sqrt_TS
F = 0.56517, num df = 35, denom df = 6, p-value = 0.2684
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1122435 1.5802822
sample estimates:
```

```
ratio of variances
         0.5651662

> var.test(data1112$sqrt_TS, data1113$sqrt_TS)

        F test to compare two variances

data:  data1112$sqrt_TS and data1113$sqrt_TS
F = 0.071556, num df = 11, denom df = 6, p-value = 0.0002834
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01322727 0.27768530
sample estimates:
ratio of variances
        0.07155637

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$TS, data22$TS)

        F test to compare two variances

data:  data21$TS and data22$TS
F = 2.1201, num df = 38, denom df = 15, p-value = 0.1187
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.817185 4.667131
sample estimates:
ratio of variances
          2.120112

> var.test(data21$TS, data23$TS)

        F test to compare two variances

data:  data21$TS and data23$TS
F = 2.7344, num df = 38, denom df = 5, p-value = 0.2624
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4422203 7.9918609
sample estimates:
ratio of variances
          2.734372

> var.test(data22$TS, data23$TS)

        F test to compare two variances

data:  data22$TS and data23$TS
F = 1.2897, num df = 15, denom df = 5, p-value = 0.8353
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.200651 4.612610
sample estimates:
ratio of variances
           1.28973

> # t-test
> ## Data 1 - Effluent
> t.test(data1111$sqrt_TS, data1112$sqrt_TS, var.equal = FALSE)

        Welch Two Sample t-test
```

```
data:  data1111$sqrt_TS and data1112$sqrt_TS
t = 2.8211, df = 45.672, p-value = 0.007061
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3776414 2.2601025
sample estimates:
mean of x mean of y 
 28.76891  27.45004 

> t.test(data1111$sqrt_TS, data1113$sqrt_TS, var.equal = TRUE)

        Two Sample t-test

data:  data1111$sqrt_TS and data1113$sqrt_TS
t = 0.75995, df = 41, p-value = 0.4516
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.310572  2.891983
sample estimates:
mean of x mean of y 
 28.76891  27.97821 

> t.test(data1112$sqrt_TS, data1113$sqrt_TS, var.equal = TRUE)

        Two Sample t-test

data:  data1112$sqrt_TS and data1113$sqrt_TS
t = -0.55332, df = 17, p-value = 0.5872
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.542078  1.485745
sample estimates:
mean of x mean of y 
 27.45004  27.97821 

> 
> ## Data 2 - Feed
> t.test(data21$TS, data22$TS, var.equal = FALSE)

        Welch Two Sample t-test

data:  data21$TS and data22$TS
t = 0.044359, df = 40.382, p-value = 0.9648
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -245.1588  256.1652
sample estimates:
mean of x mean of y 
 1891.128  1885.625 

> t.test(data21$TS, data22$TS, var.equal = TRUE)

        Two Sample t-test

data:  data21$TS and data22$TS
t = 0.038038, df = 53, p-value = 0.9698
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -284.6801  295.6865
sample estimates:
mean of x mean of y 
 1891.128  1885.625 
```

```
> t.test(data22$TS, data23$TS, var.equal = TRUE)

        Two Sample t-test

data:  data22$TS and data23$TS
t = -0.90477, df = 20, p-value = 0.3764
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -504.7784  199.3617
sample estimates:
mean of x mean of y 
 1885.625  2038.333 

> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=TS)) + 
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent TS (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
> 
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="TS", 
+                             groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount       TS        sd
1         800 833.1944 138.98177
2        1000 754.1667  46.50676
3        1200 791.4286 184.24880

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=TS)) + 
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent TS (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
> 
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="TS", 
+                            groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount       TS       sd
1         800 1891.128 528.4174
2        1000 1885.625 362.9090
3        1200 2038.333 319.5570
```

3. TSS

```
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
> head(data1)
       Date Feed_amount Water_type TSS
1 8/28/2019         800   Effluent  13
2 8/29/2019         800   Effluent  14
3 8/30/2019         800   Effluent  18
4 8/31/2019         800   Effluent  45
5  9/1/2019         800   Effluent  14
6  9/2/2019         800   Effluent   9
>
> TSS_effluent_data <- data_summary(data1, varname="TSS",
+                                         groupnames=c("Feed_amount"))
> TSS_effluent_data
  Feed_amount      TSS       sd
1         800 31.95000 22.54205
2        1000 54.23077 35.30086
3        1200 36.50000 34.03884
> #write.csv(TSS_effluent, "TSS_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
> head(data2)
        Date Feed_amount Water_type  TSS
54 8/24/2019         800       Feed 2680
55 8/28/2019         800       Feed  985
56 8/29/2019         800       Feed 1230
57 8/30/2019         800       Feed  715
58 8/31/2019         800       Feed 1430
59  9/1/2019         800       Feed 2050
>
> TSS_feed_data <- data_summary(data2, varname="TSS",
+                                         groupnames=c("Feed_amount"))
> TSS_feed_data
  Feed_amount      TSS       sd
1         800 1135.000 490.9071
2        1000 1227.647 416.8787
3        1200 1208.750 540.0182
> #write.csv(TSS_feed, "TSS_feed.csv")

> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$TSS)

        Shapiro-Wilk normality test

data:  data1$TSS
W = 0.92777, p-value = 0.004092


>
> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$TSS)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.9771, p-value = 0.4239



> ### New data structure for the data1 (Effluent data)
```

```
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_TSS")
> data111
   Feed_amount  sqrt_TSS
1          800  3.605551
2          800  3.741657
3          800  4.242641
4          800  6.708204
5          800  3.741657
6          800  3.000000
7          800  3.741657
8          800  4.415880
9          800  4.949747
10         800  4.358899
11         800  7.106335
12         800  6.670832
13         800  4.795832
14         800  6.819091
15         800  5.291503
16         800  5.099020
17         800  3.162278
18         800  3.674235
19         800  4.582576
20         800  6.324555
21         800  7.874008
22         800  9.591663
23         800  9.486833
24         800  6.892024
25         800  5.477226
26         800  8.215838
27         800  2.828427
28         800  5.385165
29         800  4.062019
30         800  4.795832
31        1000  6.519202
32        1000  8.306624
33        1000 10.723805
34        1000  7.745967
35        1000  8.660254
36        1000  3.605551
37        1000 10.000000
38        1000  7.106335
39        1000  0.000000
40        1000  8.062258
41        1000  6.324555
42        1000  0.000000
43        1000  8.660254
44        1200  7.582875
45        1200  9.617692
46        1200  8.366600
47        1200  2.236068
48        1200  2.449490
49        1200  0.000000
50        1200  6.082763
51        1200  4.898979
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_TSS, data1112$sqrt_TSS)

        F test to compare two variances

data:  data1111$sqrt_TSS and data1112$sqrt_TSS
```

```
F = 0.29091, num df = 29, denom df = 12, p-value = 0.006399
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.09791498 0.70676407
sample estimates:
ratio of variances
         0.2909064

> var.test(data1111$sqrt_TSS, data1113$sqrt_TSS)

        F test to compare two variances

data:  data1111$sqrt_TSS and data1113$sqrt_TSS
F = 0.29863, num df = 29, denom df = 7, p-value = 0.01943
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.0683419 0.8252196
sample estimates:
ratio of variances
         0.2986338

> var.test(data1112$sqrt_TSS, data1113$sqrt_TSS)

        F test to compare two variances

data:  data1112$sqrt_TSS and data1113$sqrt_TSS
F = 1.0266, num df = 12, denom df = 7, p-value = 0.9825
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2200173 3.7023149
sample estimates:
ratio of variances
          1.026563

> ## Normality check on data2
> shapiro.test(data2$TSS)

        Shapiro-Wilk normality test

data:  data2$TSS
W = 0.94762, p-value = 0.00808

>
> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$TSS, data22$TSS)

        F test to compare two variances

data:  data21$TSS and data22$TSS
F = 1.3867, num df = 39, denom df = 16, p-value = 0.4879
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5517704 3.0003587
sample estimates:
ratio of variances
          1.386689

> var.test(data21$TSS, data23$TSS)

        F test to compare two variances
```

```
data:  data21$TSS and data23$TSS
F = 0.82638, num df = 39, denom df = 7, p-value = 0.6408
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1916018 2.1758639
sample estimates:
ratio of variances
         0.8263838

> var.test(data22$TSS, data23$TSS)

        F test to compare two variances

data:  data22$TSS and data23$TSS
F = 0.59594, num df = 16, denom df = 7, p-value = 0.3698
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1311829 1.9185885
sample estimates:
ratio of variances
         0.5959402


> # t-test and ANOVA
> ## Data 1 - Effluent
> t.test(data1111$sqrt_TSS, data1112$sqrt_TSS, var.equal = FALSE)

        Welch Two Sample t-test

data:  data1111$sqrt_TSS and data1112$sqrt_TSS
t = -1.2329, df = 15.117, p-value = 0.2364
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.3787877  0.9013068
sample estimates:
mean of x mean of y
 5.354706  6.593447

> t.test(data1111$sqrt_TSS, data1113$sqrt_TSS, var.equal = FALSE)

        Welch Two Sample t-test

data:  data1111$sqrt_TSS and data1113$sqrt_TSS
t = 0.16191, df = 8.1468, p-value = 0.8753
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.644924  3.045720
sample estimates:
mean of x mean of y
 5.354706  5.154308

> t.test(data1112$sqrt_TSS, data1113$sqrt_TSS, var.equal = TRUE)

        Two Sample t-test

data:  data1112$sqrt_TSS and data1113$sqrt_TSS
t = 0.94267, df = 19, p-value = 0.3577
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.756212  4.634488
sample estimates:
mean of x mean of y
 6.593447  5.154308
```

```
> ## Data 2 - Feed
> fit2 <- aov(TSS~Feed_amount, data2)
> summary(fit2)
            Df   Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   117320   58660   0.256  0.775
Residuals   62 14220543  229364
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TSS ~ Feed_amount, data = data2)

$Feed_amount
               diff       lwr      upr     p adj
1000-800   92.64706 -240.3067 425.6008 0.7827561
1200-800   73.75000 -371.6463 519.1463 0.9166695
1200-1000 -18.89706 -511.9591 474.1649 0.9953418

> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=TSS)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent TSS (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="TSS",
+                              groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount      TSS       sd
1         800 31.95000 22.54205
2        1000 54.23077 35.30086
3        1200 36.50000 34.03884



> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=TSS)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent TSS (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="TSS",
```

```
+                             groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount      TSS       sd
1         800 1135.000 490.9071
2        1000 1227.647 416.8787
3        1200 1208.750 540.0182
```

4. COD

```
> # Choose data file COD.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type   COD
1 8/29/2019         800   Effluent 174.5
2 8/30/2019         800   Effluent 180.0
3 8/31/2019         800   Effluent 166.5
4  9/1/2019         800   Effluent 153.0
5  9/2/2019         800   Effluent 109.5
6  9/4/2019         800   Effluent 101.0

> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> COD_effluent_data <- data_summary(data1, varname="COD",
+                                           groupnames=c("Feed_amount"))
> COD_effluent_data
  Feed_amount      COD       sd
1         800 139.6129 62.41043
2        1000 147.9412 36.16753
3        1200 166.8571 74.32570
> #write.csv(COD_effluent, "COD_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> COD_feed_data <- data_summary(data2, varname="COD",
+                                          groupnames=c("Feed_amount"))
> COD_feed_data
  Feed_amount      COD       sd
1         800 2753.314 846.5639
2        1000 2973.235 848.5004
3        1200 2729.375 599.0495
> #write.csv(COD_feed, "COD_feed.csv")

> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$COD)

        Shapiro-Wilk normality test

data:  data1$COD
W = 0.93013, p-value = 0.003332

> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$COD)
> shapiro.test(data11)
```

```
        Shapiro-Wilk normality test

data:  data11
W = 0.96681, p-value = 0.1321

> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_COD")
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_COD, data1112$sqrt_COD)

        F test to compare two variances

data:  data1111$sqrt_COD and data1112$sqrt_COD
F = 2.6219, num df = 30, denom df = 16, p-value = 0.04539
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.021066 5.977657
sample estimates:
ratio of variances
          2.621907

> var.test(data1111$sqrt_COD, data1113$sqrt_COD)

        F test to compare two variances

data:  data1111$sqrt_COD and data1113$sqrt_COD
F = 0.79222, num df = 30, denom df = 6, p-value = 0.6081
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1564038 2.2710558
sample estimates:
ratio of variances
         0.7922206

> var.test(data1112$sqrt_COD, data1113$sqrt_COD)

        F test to compare two variances

data:  data1112$sqrt_COD and data1113$sqrt_COD
F = 0.30215, num df = 16, denom df = 6, p-value = 0.05177
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.0576206 1.0093862
sample estimates:
ratio of variances
         0.3021544

> ## Normality check on data2
> shapiro.test(data2$COD)

        Shapiro-Wilk normality test

data:  data2$COD
W = 0.97724, p-value = 0.2479

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
```

```
> var.test(data21$COD, data22$COD)

        F test to compare two variances

data:  data21$COD and data22$COD
F = 0.99544, num df = 42, denom df = 16, p-value = 0.9415
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.398194 2.126865
sample estimates:
ratio of variances
         0.9954407

> var.test(data21$COD, data23$COD)

        F test to compare two variances

data:  data21$COD and data23$COD
F = 1.9971, num df = 42, denom df = 7, p-value = 0.3428
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4643112 5.2058775
sample estimates:
ratio of variances
          1.997074

> var.test(data22$COD, data23$COD)

        F test to compare two variances

data:  data22$COD and data23$COD
F = 2.0062, num df = 16, denom df = 7, p-value = 0.3556
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4416248 6.4588901
sample estimates:
ratio of variances
          2.006221

> # t-test and ANOVA
> ## Data 1 - Effluent
> t.test(data1111$sqrt_COD, data1112$sqrt_COD, var.equal = FALSE)

        Welch Two Sample t-test

data:  data1111$sqrt_COD and data1112$sqrt_COD
t = -0.87851, df = 45.224, p-value = 0.3843
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.6899181  0.6633378
sample estimates:
mean of x mean of y
 11.55736  12.07065

> t.test(data1111$sqrt_COD, data1113$sqrt_COD, var.equal = FALSE)

        Welch Two Sample t-test

data:  data1111$sqrt_COD and data1113$sqrt_COD
t = -0.95135, df = 8.2856, p-value = 0.3683
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.736282  1.544563
sample estimates:
```

```
mean of x mean of y
 11.55736  12.65322

> t.test(data1112$sqrt_COD, data1113$sqrt_COD, var.equal = TRUE)

	Two Sample t-test

data:  data1112$sqrt_COD and data1113$sqrt_COD
t = -0.65856, df = 22, p-value = 0.517
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.417133  1.251995
sample estimates:
mean of x mean of y
 12.07065  12.65322


>
> ## Data 2 - Feed
> fit2 <- aov(COD~Feed_amount, data2)
> summary(fit2)
            Df   Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   641763  320881   0.473  0.625
Residuals   65 44131428  678945
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = COD ~ Feed_amount, data = data2)

$Feed_amount
                diff        lwr      upr     p adj
1000-800   219.92134  -346.2957 786.1384 0.6224216
1200-800   -23.93895  -784.9166 737.0387 0.9968662
1200-1000 -243.86029 -1091.2176 603.4970 0.7699916


> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=COD)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent COD (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="COD",
+                               groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount      COD       sd
1         800 139.6129 62.41043
2        1000 147.9412 36.16753
3        1200 166.8571 74.32570
```

```
> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=COD)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent COD (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="COD",
+                            groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount      COD       sd
1         800 2753.314 846.5639
2        1000 2973.235 848.5004
3        1200 2729.375 599.0495
```

5. TOC

```
> # Choose data file TOC.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type  TOC
1 8/28/2019         800   Effluent 97.5
2 8/29/2019         800   Effluent 68.5
3  9/4/2019         800   Effluent 44.5
4  9/6/2019         800   Effluent 54.0
5  9/9/2019         800   Effluent 52.0
6 9/11/2019         800   Effluent 46.0
>
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)

> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> TOC_effluent_data <- data_summary(data1, varname="TOC",
+                                   groupnames=c("Feed_amount"))
> TOC_effluent_data
  Feed_amount      TOC       sd
1         800 65.40357 25.42590
2        1000 61.50000 31.60222
3        1200 40.25000 16.93369
> #write.csv(TOC_effluent, "TOC_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> TOC_feed_data <- data_summary(data2, varname="TOC",
+                               groupnames=c("Feed_amount"))
> TOC_feed_data
  Feed_amount      TOC       sd
1         800 910.6786 448.7026
2        1000 578.5625 139.9003
3        1200 693.7500 210.8070
> #write.csv(TOC_feed, "TOC_feed.csv")
```

```
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$TOC)

        Shapiro-Wilk normality test

data:  data1$TOC
W = 0.90973, p-value = 0.03479

> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$TOC)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.93494, p-value = 0.1257


> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_TOC")
>
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_TOC, data1112$sqrt_TOC)

        F test to compare two variances

data:  data1111$sqrt_TOC and data1112$sqrt_TOC
F = 0.68122, num df = 13, denom df = 5, p-value = 0.5317
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1050042 2.5659455
sample estimates:
ratio of variances
         0.6812231

> var.test(data1111$sqrt_TOC, data1113$sqrt_TOC)

        F test to compare two variances

data:  data1111$sqrt_TOC and data1113$sqrt_TOC
F = 1.2718, num df = 13, denom df = 3, p-value = 0.9546
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.08890973 5.52877365
sample estimates:
ratio of variances
          1.271807

> var.test(data1112$sqrt_TOC, data1113$sqrt_TOC)

        F test to compare two variances

data:  data1112$sqrt_TOC and data1113$sqrt_TOC
F = 1.8669, num df = 5, denom df = 3, p-value = 0.6442
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
```

```
  0.1254262 14.4942103
sample estimates:
ratio of variances
          1.866947


> ## Normality check on data2
> shapiro.test(data2$TOC)

        Shapiro-Wilk normality test

data:  data2$TOC
W = 0.88577, p-value = 0.01088

>
> ### the data are not normal, square root transformation is needed.
> data21<-sqrt(data2$TOC)
> shapiro.test(data21)

        Shapiro-Wilk normality test

data:  data21
W = 0.91979, p-value = 0.05779


> ### New data structure for the data1 (Feed data)
> data211<-data.frame(data1$Feed_amount, data21)
> colnames(data211)<-c("Feed_amount","sqrt_TOC")
>
> ### Equal variance check for data2
> data2111<-data211[which(data211$Feed_amount=="800"),]
> data2112<-data211[which(data211$Feed_amount=="1000"),]
> data2113<-data211[which(data211$Feed_amount=="1200"),]
> var.test(data2111$sqrt_TOC, data2112$sqrt_TOC)

        F test to compare two variances

data:  data2111$sqrt_TOC and data2112$sqrt_TOC
F = 3.4585, num df = 13, denom df = 5, p-value = 0.1788
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.5331008 13.0271694
sample estimates:
ratio of variances
          3.458534

> var.test(data2111$sqrt_TOC, data2113$sqrt_TOC)

        F test to compare two variances

data:  data2111$sqrt_TOC and data2113$sqrt_TOC
F = 1.7537, num df = 13, denom df = 3, p-value = 0.7111
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1225981 7.6236550
sample estimates:
ratio of variances
          1.753702

> var.test(data2112$sqrt_TOC, data2113$sqrt_TOC)

        F test to compare two variances

data:  data2112$sqrt_TOC and data2113$sqrt_TOC
```

```
F = 0.50707, num df = 5, denom df = 3, p-value = 0.4733
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.03406592 3.93664586
sample estimates:
ratio of variances
         0.5070652


> # ANOVA
> ## Data 1 - Effluent
> fit1 <- aov(sqrt_TOC~Feed_amount, data111)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   9.16   4.582   1.593  0.227
Residuals   21  60.41   2.877
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sqrt_TOC ~ Feed_amount, data = data111)

$Feed_amount
                diff       lwr       upr     p adj
1000-800  -0.3011919 -2.387275 1.7848907 0.9298392
1200-800  -1.7136164 -4.137425 0.7101926 0.1998486
1200-1000 -1.4124245 -4.172052 1.3472034 0.4161351

> ## Data 2 - Feed
> fit2 <- aov(sqrt_TOC~Feed_amount, data211)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2   34.3   17.14   0.683  0.516
Residuals   21  527.3   25.11
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sqrt_TOC ~ Feed_amount, data = data211)

$Feed_amount
                diff       lwr       upr     p adj
1000-800  -2.8571036 -9.020369  3.306162 0.4843540
1200-800  -0.8200824 -7.981150  6.340985 0.9552057
1200-1000  2.0370212 -6.116213 10.190255 0.8055676


> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=TOC)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent TOC (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
```

```
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="TOC",
+                             groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount      TOC       sd
1         800 65.40357 25.42590
2        1000 61.50000 31.60222
3        1200 40.25000 16.93369
```

```
> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=TOC)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent TOC (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="TOC",
+                            groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount      TOC       sd
1         800 786.2083 336.5103
2        1000 578.5625 139.9003
3        1200 693.7500 210.8070
```

6. BOD

```
> # Choose data file TOC.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type     BOD
1 8/27/2019         800   Effluent 215.220
2 8/30/2019         800   Effluent  71.740
3  9/5/2019         800   Effluent 132.950
4 9/11/2019         800   Effluent  25.905
5 9/16/2019         800   Effluent  81.235
6 9/24/2019         800   Effluent 408.285
>
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> BOD_effluent_data <- data_summary(data1, varname="BOD",
+                                   groupnames=c("Feed_amount"))
> BOD_effluent_data
  Feed_amount     BOD       sd
1         800 132.405 116.8274
2        1000 237.800       NA
> #write.csv(TOC_effluent, "TOC_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
```

```
> BOD_feed_data <- data_summary(data2, varname="BOD",
+                                   groupnames=c("Feed_amount"))
> BOD_feed_data
  Feed_amount      BOD       sd
1         800 1504.669 465.2271
2        1000 1732.310       NA
> #write.csv(TOC_feed, "TOC_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$BOD)

        Shapiro-Wilk normality test

data:  data1$BOD
W = 0.88726, p-value = 0.1087

>
> ## Normality check on data2
> shapiro.test(data2$BOD)

        Shapiro-Wilk normality test

data:  data2$BOD
W = 0.95396, p-value = 0.6594

>
>
> # ANOVA
> ## Data 1 - Effluent
> fit1 <- aov(BOD~Feed_amount, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  1  10182   10182   0.746  0.408
Residuals   10 136486   13649
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = BOD ~ Feed_amount, data = data1)

$Feed_amount
            diff       lwr      upr     p adj
1000-800 105.395 -166.4875 377.2775 0.4079724

>
> ## Data 2 - Feed
> fit2 <- aov(BOD~Feed_amount, data2)
> summary(fit2)
            Df  Sum Sq Mean Sq F value Pr(>F)
Feed_amount  1   47834   47834   0.221  0.647
Residuals   11 2380799  216436
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = BOD ~ Feed_amount, data = data2)

$Feed_amount
             diff       lwr     upr     p adj
```

```
1000-800 227.6408 -838.1285 1293.41 0.6474557

>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=BOD)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent BOD (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="BOD",
+                            groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount     BOD       sd
1         800 132.405 116.8274
2        1000 237.800       NA
>


> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=BOD)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent BOD (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="BOD",
+                           groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount      BOD       sd
1         800 1504.669 465.2271
2        1000 1732.310       NA
```

7. NH3

```
> # Choose data file NH3.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type   NH3
1 8/28/2019         800   Effluent 10.00
2 8/29/2019         800   Effluent  7.50
3 8/30/2019         800   Effluent  6.50
4  9/2/2019         800   Effluent  4.20
```

```
5  9/3/2019         800   Effluent  5.70
6  9/4/2019         800   Effluent  4.55
>
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> NH3_effluent_data <- data_summary(data1, varname="NH3",
+                                           groupnames=c("Feed_amount"))
> NH3_effluent_data
  Feed_amount      NH3        sd
1         800 6.738095 2.8422660
2        1000 4.955882 1.8107217
3        1200 1.887500 0.9034655
> #write.csv(NH3_effluent, "NH3_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> NH3_feed_data <- data_summary(data2, varname="NH3",
+                                        groupnames=c("Feed_amount"))
> NH3_feed_data
  Feed_amount     NH3       sd
1         800 42.7600 9.189354
2        1000 38.0875 4.784715
3        1200 35.2750 4.657329
> #write.csv(NH3_feed, "NH3_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$NH3)

        Shapiro-Wilk normality test

data:  data1$NH3
W = 0.93147, p-value = 0.00952

>
> ### the data are not normal, square root transformation is needed.
> data11<-sqrt(data1$NH3)
> shapiro.test(data11)

        Shapiro-Wilk normality test

data:  data11
W = 0.97455, p-value = 0.4042

>
> ### New data structure for the data1 (Effluent data)
> data111<-data.frame(data1$Feed_amount, data11)
> colnames(data111)<-c("Feed_amount","sqrt_NH3")
>
> ### Equal variance check for data1
> data1111<-data111[which(data111$Feed_amount=="800"),]
> data1112<-data111[which(data111$Feed_amount=="1000"),]
> data1113<-data111[which(data111$Feed_amount=="1200"),]
> var.test(data1111$sqrt_NH3, data1112$sqrt_NH3)

        F test to compare two variances
```

```
data:  data1111$sqrt_NH3 and data1112$sqrt_NH3
F = 1.5361, num df = 20, denom df = 16, p-value = 0.3873
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5730112 3.9118019
sample estimates:
ratio of variances
          1.536124

> var.test(data1111$sqrt_NH3, data1113$sqrt_NH3)

        F test to compare two variances

data:  data1111$sqrt_NH3 and data1113$sqrt_NH3
F = 2.6651, num df = 20, denom df = 7, p-value = 0.1879
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5966585 8.0151201
sample estimates:
ratio of variances
          2.665118

> var.test(data1112$sqrt_NH3, data1113$sqrt_NH3)

        F test to compare two variances

data:  data1112$sqrt_NH3 and data1113$sqrt_NH3
F = 1.735, num df = 16, denom df = 7, p-value = 0.4699
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3819133 5.5855931
sample estimates:
ratio of variances
          1.734963

>
> ## Normality check on data2
> shapiro.test(data2$NH3)

        Shapiro-Wilk normality test

data:  data2$NH3
W = 0.97456, p-value = 0.2516

>
> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$NH3, data22$NH3)

        F test to compare two variances

data:  data21$NH3 and data22$NH3
F = 3.6886, num df = 34, denom df = 15, p-value = 0.009319
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.409821 8.290459
sample estimates:
ratio of variances
          3.688568

> var.test(data21$NH3, data23$NH3)
```

```
        F test to compare two variances

data:  data21$NH3 and data23$NH3
F = 3.8931, num df = 34, denom df = 7, p-value = 0.06741
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.8975821 10.4639951
sample estimates:
ratio of variances
          3.893105

> var.test(data22$NH3, data23$NH3)

        F test to compare two variances

data:  data22$NH3 and data23$NH3
F = 1.0555, num df = 15, denom df = 7, p-value = 0.9987
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2310641 3.4759820
sample estimates:
ratio of variances
          1.055452

>
> # t-test and ANOVA
> ## Data 1 - Effluent
> fit1 <- aov(NH3~Feed_amount, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Feed_amount  2  138.4   69.21   13.54 2.75e-05 ***
Residuals   43  219.7    5.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = NH3 ~ Feed_amount, data = data1)

$Feed_amount
               diff       lwr         upr     p adj
1000-800  -1.782213 -3.572527  0.00810091 0.0512683
1200-800  -4.850595 -7.130495 -2.57069509 0.0000174
1200-1000 -3.068382 -5.421112 -0.71565274 0.0078397

>
> ## Data 2 - Feed
> t.test(data21$NH3, data22$NH3, var.equal=FALSE)

        Welch Two Sample t-test

data:  data21$NH3 and data22$NH3
t = 2.3833, df = 48.011, p-value = 0.02116
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7306861 8.6143139
sample estimates:
mean of x mean of y
  42.7600   38.0875

> t.test(data21$NH3, data23$NH3, var.equal = TRUE)
```

```
        Two Sample t-test

data:  data21$NH3 and data23$NH3
t = 2.2244, df = 41, p-value = 0.03169
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.6893556 14.2806444
sample estimates:
mean of x mean of y
   42.760    35.275

> t.test(data22$NH3, data23$NH3, var.equal = TRUE)

        Two Sample t-test

data:  data22$NH3 and data23$NH3
t = 1.369, df = 22, p-value = 0.1848
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.448173  7.073173
sample estimates:
mean of x mean of y
  38.0875   35.2750

>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=NH3)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent NH3 (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="NH3",
+                             groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount      NH3        sd
1         800 6.738095 2.8422660
2        1000 4.955882 1.8107217
3        1200 1.887500 0.9034655
>

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=NH3)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent NH3 (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
```

```
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="NH3",
+                              groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount     NH3       sd
1         800 42.7600 9.189354
2        1000 38.0875 4.784715
3        1200 35.2750 4.657329
```

8. Nitrite

```
> # Choose data file NO2.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type    NO2
1  9/6/2019         800   Effluent 0.2380
2 9/10/2019         800   Effluent 0.2615
3 9/11/2019         800   Effluent 0.2245
4 9/12/2019         800   Effluent 0.0265
5 9/13/2019         800   Effluent 0.0120
6 9/16/2019         800   Effluent 0.0820
>
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> NO2_effluent_data <- data_summary(data1, varname="NO2",
+                                           groupnames=c("Feed_amount"))
> NO2_effluent_data
  Feed_amount        NO2         sd
1         800 0.09388000 0.07202304
2        1000 0.06485294 0.04916197
3        1200 0.03310000 0.01432393
> #write.csv(NO2_effluent, "NO2_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> NO2_feed_data <- data_summary(data2, varname="NO2",
+                                          groupnames=c("Feed_amount"))
> NO2_feed_data
  Feed_amount       NO2         sd
1         800 0.1878286 0.08829912
2        1000 0.1677143 0.07783358
3        1200 0.1715000 0.03961481
> #write.csv(NO2_feed, "NO2_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$NO2)

        Shapiro-Wilk normality test

data:  data1$NO2
W = 0.85848, p-value = 4.35e-05
```

```
> # Variance
> infer_levene_test(data1, Feed_amount, NO2)
           Summary Statistics
Levels    Frequency    Mean    Std. Dev
----------------------------------------
  0          47         1.57      0.68
  1          47         0.08      0.06
----------------------------------------
Total        94         0.83      0.89
----------------------------------------

                              Test Statistics
---------------------------------------------------------------------------
Statistic                             Num DF    Den DF           F    Pr > F
---------------------------------------------------------------------------
Brown and Forsythe                         1        92    169.8704         0
Levene                                     1        92     27.8173         0
Brown and Forsythe (Trimmed Mean)          1        92    142.9221         0
---------------------------------------------------------------------------
> # Significance
> kruskal.test(NO2 ~ Feed_amount, data = data1)

        Kruskal-Wallis rank sum test

data:  NO2 by Feed_amount
Kruskal-Wallis chi-squared = 4.246, df = 2, p-value = 0.1197

>
> ## Normality check on data2
> shapiro.test(data2$NO2)

        Shapiro-Wilk normality test

data:  data2$NO2
W = 0.95528, p-value = 0.03683

> # Variance
> infer_levene_test(data2, Feed_amount, NO2)
           Summary Statistics
Levels    Frequency    Mean    Std. Dev
----------------------------------------
  0          56         1.5       0.71
  1          56         0.18      0.08
----------------------------------------
Total        112        0.84      0.83
----------------------------------------

                              Test Statistics
---------------------------------------------------------------------------
Statistic                             Num DF    Den DF           F    Pr > F
---------------------------------------------------------------------------
Brown and Forsythe                         1       110    154.8742         0
Levene                                     1       110     20.8045         0
Brown and Forsythe (Trimmed Mean)          1       110     99.9614         0
---------------------------------------------------------------------------
> # Significance
> kruskal.test(NO2 ~ Feed_amount, data = data2)

        Kruskal-Wallis rank sum test

data:  NO2 by Feed_amount
Kruskal-Wallis chi-squared = 0.37615, df = 2, p-value = 0.8286
```

```
>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=NO2)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent nitrite (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="NO2",
+                             groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount        NO2         sd
1         800 0.09388000 0.07202304
2        1000 0.06485294 0.04916197
3        1200 0.03310000 0.01432393
>

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=NO2)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent nitrite (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="NO2",
+                            groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount       NO2         sd
1         800 0.1878286 0.08829912
2        1000 0.1677143 0.07783358
3        1200 0.1715000 0.03961481
```

9. Nitrate

```
> # Choose data file NO3.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type    NO3
1  9/2/2019         800   Effluent 0.8815
2  9/3/2019         800   Effluent 0.4455
3  9/4/2019         800   Effluent 0.6290
4  9/5/2019         800   Effluent 0.8550
5  9/6/2019         800   Effluent 0.9195
6 9/10/2019         800   Effluent 0.6100
>
```

```
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> NO3_effluent_data <- data_summary(data1, varname="NO3",
+                                   groupnames=c("Feed_amount"))
> NO3_effluent_data
  Feed_amount       NO3         sd
1         800 0.5232586 0.21577194
2        1000 0.4332308 0.06401713
3        1200 0.3395714 0.09573115
> #write.csv(NO3_effluent, "NO3_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> NO3_feed_data <- data_summary(data2, varname="NO3",
+                                  groupnames=c("Feed_amount"))
> NO3_feed_data
  Feed_amount       NO3        sd
1         800 0.7355000 0.2250017
2        1000 0.6503846 0.1744834
3        1200 0.6353125 0.1605074
> #write.csv(NO3_feed, "NO3_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$NO3)

        Shapiro-Wilk normality test

data:  data1$NO3
W = 0.92606, p-value = 0.004385

> # Variance
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> var.test(data11$NO3, data12$NO3)

        F test to compare two variances

data:  data11$NO3 and data12$NO3
F = 11.361, num df = 28, denom df = 12, p-value = 8.314e-05
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  3.813193 27.814773
sample estimates:
ratio of variances
           11.3605

> var.test(data11$NO3, data13$NO3)

        F test to compare two variances

data:  data11$NO3 and data13$NO3
F = 5.0802, num df = 28, denom df = 6, p-value = 0.05
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.000015 14.746156
```

```
sample estimates:
ratio of variances
           5.08023

> var.test(data12$NO3, data13$NO3)

        F test to compare two variances

data:  data12$NO3 and data13$NO3
F = 0.44718, num df = 12, denom df = 6, p-value = 0.2219
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.08333269 1.66723089
sample estimates:
ratio of variances
         0.4471835

>
> # Significance
> kruskal.test(NO3 ~ Feed_amount, data = data1)

        Kruskal-Wallis rank sum test

data:  NO3 by Feed_amount
Kruskal-Wallis chi-squared = 6.8898, df = 2, p-value = 0.03191

>
> # t-test
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> t.test(data11$NO3, data12$NO3, var.equal=FALSE)

        Welch Two Sample t-test

data:  data11$NO3 and data12$NO3
t = 2.0542, df = 36.768, p-value = 0.04711
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.001209958 0.178845745
sample estimates:
mean of x mean of y
0.5232586 0.4332308

> t.test(data11$NO3, data13$NO3, var.equal = TRUE)

        Two Sample t-test

data:  data11$NO3 and data13$NO3
t = 2.1821, df = 34, p-value = 0.03611
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01261275 0.35476164
sample estimates:
mean of x mean of y
0.5232586 0.3395714

> t.test(data12$NO3, data13$NO3, var.equal = TRUE)

        Two Sample t-test

data:  data12$NO3 and data13$NO3
t = 2.6262, df = 18, p-value = 0.01713
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
 0.01873396 0.16858472
sample estimates:
mean of x mean of y
0.4332308 0.3395714

>
> ## Normality check on data2
> shapiro.test(data2$NO3)

        Shapiro-Wilk normality test

data:  data2$NO3
W = 0.96393, p-value = 0.08713

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$NO3, data22$NO3)

        F test to compare two variances

data:  data21$NO3 and data22$NO3
F = 1.6629, num df = 35, denom df = 12, p-value = 0.3479
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5673575 3.8922042
sample estimates:
ratio of variances
           1.66289

> var.test(data21$NO3, data23$NO3)

        F test to compare two variances

data:  data21$NO3 and data23$NO3
F = 1.9651, num df = 35, denom df = 7, p-value = 0.357
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4536285 5.2576109
sample estimates:
ratio of variances
          1.965085

> var.test(data22$NO3, data23$NO3)

        F test to compare two variances

data:  data22$NO3 and data23$NO3
F = 1.1817, num df = 12, denom df = 7, p-value = 0.8572
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2532731 4.2619240
sample estimates:
ratio of variances
          1.181729

>
> # Significance
> fit2 <- aov(NO3~Feed_amount, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2 0.1106 0.05531   1.289  0.284
```

```
Residuals   54 2.3176 0.04292
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = NO3 ~ Feed_amount, data = data2)

$Feed_amount
                 diff        lwr        upr     p adj
1000-800  -0.08511538 -0.2466662 0.07643546 0.4182539
1200-800  -0.10018750 -0.2953355 0.09496055 0.4367220
1200-1000 -0.01507212 -0.2394226 0.20927841 0.9856564

>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=NO3)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Effluent nitrate (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="NO3",
+                              groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount       NO3         sd
1         800 0.5232586 0.21577194
2        1000 0.4332308 0.06401713
3        1200 0.3395714 0.09573115
>

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=NO3)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Influent nitrate (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="NO3",
+                             groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount       NO3        sd
1         800 0.7355000 0.2250017
2        1000 0.6503846 0.1744834
3        1200 0.6353125 0.1605074
```

10. TKN

```
> # Choose data file TKN.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
       Date Feed_amount Water_type    TKN
1 8/24/2019         800   Effluent 21.600
2 8/26/2019         800   Effluent 12.400
3 8/27/2019         800   Effluent  3.020
4  9/1/2019         800   Effluent 10.015
5  9/2/2019         800   Effluent  4.740
6  9/3/2019         800   Effluent  3.870
>
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> TKN_effluent_data <- data_summary(data1, varname="TKN",
+                                   groupnames=c("Feed_amount"))
> TKN_effluent_data
  Feed_amount       TKN        sd
1         800 14.028919 13.129832
2        1000  8.791250  2.719813
3        1200  5.262143  1.517841
> #write.csv(TKN_effluent, "TKN_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> TKN_feed_data <- data_summary(data2, varname="TKN",
+                                groupnames=c("Feed_amount"))
> TKN_feed_data
  Feed_amount      TKN       sd
1         800 99.47625 25.67539
2        1000 96.56667 21.56670
3        1200 93.82143 15.26562
> #write.csv(TKN_feed, "TKN_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$TKN)

        Shapiro-Wilk normality test

data:  data1$TKN
W = 0.72939, p-value = 3.417e-09

> # Variance
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> var.test(data11$TKN, data12$TKN)

        F test to compare two variances

data:  data11$TKN and data12$TKN
```

```
F = 23.305, num df = 36, denom df = 15, p-value = 6.487e-08
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  8.946766 51.808687
sample estimates:
ratio of variances
          23.30452

> var.test(data11$TKN, data13$TKN)

        F test to compare two variances

data:  data11$TKN and data13$TKN
F = 74.828, num df = 36, denom df = 6, p-value = 2.433e-05
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  14.87597 208.36321
sample estimates:
ratio of variances
          74.82831

> var.test(data12$TKN, data13$TKN)

        F test to compare two variances

data:  data12$TKN and data13$TKN
F = 3.2109, num df = 15, denom df = 6, p-value = 0.1577
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.6094318 10.9641225
sample estimates:
ratio of variances
          3.210893

>
> # Significance
> kruskal.test(TKN ~ Feed_amount, data = data1)

        Kruskal-Wallis rank sum test

data:  TKN by Feed_amount
Kruskal-Wallis chi-squared = 5.1976, df = 2, p-value = 0.07436

>
> ## Normality check on data2
> shapiro.test(data2$TKN)

        Shapiro-Wilk normality test

data:  data2$TKN
W = 0.98145, p-value = 0.4701

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$TKN, data22$TKN)

        F test to compare two variances

data:  data21$TKN and data22$TKN
F = 1.4173, num df = 39, denom df = 14, p-value = 0.4894
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
```

```
 0.5290941 3.1498598
sample estimates:
ratio of variances
          1.417317

> var.test(data21$TKN, data23$TKN)

        F test to compare two variances

data:  data21$TKN and data23$TKN
F = 2.8288, num df = 39, denom df = 6, p-value = 0.193
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5638968 7.7894381
sample estimates:
ratio of variances
          2.828821

> var.test(data22$TKN, data23$TKN)

        F test to compare two variances

data:  data22$TKN and data23$TKN
F = 1.9959, num df = 14, denom df = 6, p-value = 0.405
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3768114 6.9883705
sample estimates:
ratio of variances
          1.995899

>
> # Significance
> fit2 <- aov(TKN~Feed_amount, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2    239   119.6    0.21  0.811
Residuals   59  33620   569.8
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TKN ~ Feed_amount, data = data2)

$Feed_amount
               diff       lwr      upr     p adj
1000-800  -2.909583 -20.28583 14.46666 0.9146732
1200-800  -5.654821 -29.16848 17.85884 0.8322590
1200-1000 -2.745238 -29.01566 23.52518 0.9658218

>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=TKN)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("TN (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
```

```
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="TKN",
+                                     groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount       TKN        sd
1         800 14.028919 13.129832
2        1000  8.791250  2.719813
3        1200  5.262143  1.517841
>


> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=TKN)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("TN (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="TKN",
+                                    groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount      TKN       sd
1         800 99.47625 25.67539
2        1000 96.56667 21.56670
3        1200 93.82143 15.26562
```

11. TP

```
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> TP_effluent_data <- data_summary(data1, varname="TP",
+                                                  groupnames=c("Feed_amount"))
> TP_effluent_data
  Feed_amount        TP        sd
1         800 2.0148571 1.4887428
2        1000 1.7088235 1.0942456
3        1200 0.9883333 0.5207463
> #write.csv(TP_effluent, "TP_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> TP_feed_data <- data_summary(data2, varname="TP",
+                                                 groupnames=c("Feed_amount"))
> TP_feed_data
  Feed_amount       TP        sd
1         800 30.62718 11.574890
```

```
2        1000 32.95000 17.150848
3        1200 26.73571  6.222119
> #write.csv(TP_feed, "TP_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$TP)

        Shapiro-Wilk normality test

data:  data1$TP
W = 0.86448, p-value = 1.13e-05

> # Variance
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> var.test(data11$TP, data12$TP)

        F test to compare two variances

data:  data11$TP and data12$TP
F = 1.851, num df = 34, denom df = 16, p-value = 0.19
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7286972 4.1095093
sample estimates:
ratio of variances
          1.851014

> var.test(data11$TP, data13$TP)

        F test to compare two variances

data:  data11$TP and data13$TP
F = 8.1731, num df = 34, denom df = 5, p-value = 0.02708
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.317696 24.257961
sample estimates:
ratio of variances
          8.173104

> var.test(data12$TP, data13$TP)

        F test to compare two variances

data:  data12$TP and data13$TP
F = 4.4155, num df = 16, denom df = 5, p-value = 0.1088
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.6895772 15.4635037
sample estimates:
ratio of variances
          4.415474

>
> # Significance
> kruskal.test(TP ~ Feed_amount, data = data1)

        Kruskal-Wallis rank sum test
```

```
data:  TP by Feed_amount
Kruskal-Wallis chi-squared = 2.8036, df = 2, p-value = 0.2462

>
> ## Normality check on data2
> shapiro.test(data2$TP)

        Shapiro-Wilk normality test

data:  data2$TP
W = 0.85324, p-value = 2.399e-06

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$TP, data22$TP)

        F test to compare two variances

data:  data21$TP and data22$TP
F = 0.45547, num df = 38, denom df = 16, p-value = 0.04713
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1808848 0.9900694
sample estimates:
ratio of variances
         0.4554729

> var.test(data21$TP, data23$TP)

        F test to compare two variances

data:  data21$TP and data23$TP
F = 3.4606, num df = 38, denom df = 6, p-value = 0.1229
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6892542 9.5629235
sample estimates:
ratio of variances
          3.460646

> var.test(data22$TP, data23$TP)

        F test to compare two variances

data:  data22$TP and data23$TP
F = 7.5979, num df = 16, denom df = 6, p-value = 0.01946
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.448917 25.381839
sample estimates:
ratio of variances
          7.597918

>
> # Significance
> kruskal.test(TP ~ Feed_amount, data = data2)

        Kruskal-Wallis rank sum test

data:  TP by Feed_amount
Kruskal-Wallis chi-squared = 1.1152, df = 2, p-value = 0.5726
```

```
>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=TP)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("TP (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="TP",
+                              groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount        TP        sd
1         800 2.0148571 1.4887428
2        1000 1.7088235 1.0942456
3        1200 0.9883333 0.5207463
>

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=TP)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("TP (mg/L)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="TP",
+                             groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount       TP        sd
1         800 30.62718 11.574890
2        1000 32.95000 17.150848
3        1200 26.73571  6.222119
```

12. Total Coliform

```
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> Tcoli_effluent_data <- data_summary(data1, varname="Tcoli",
+                                           groupnames=c("Feed_amount"))
> Tcoli_effluent_data
```

```
  Feed_amount    Tcoli        sd
1         800 6.257871 0.9745939
2        1000 5.817774 0.1507249
3        1200 6.087790 0.4117186
> #write.csv(Tcoli_effluent, "Tcoli_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> Tcoli_feed_data <- data_summary(data2, varname="Tcoli",
+                                         groupnames=c("Feed_amount"))
> Tcoli_feed_data
  Feed_amount    Tcoli        sd
1         800 7.492791 0.2593434
2        1000 7.690945 0.1528799
3        1200 7.957077 0.2986723
> #write.csv(Tcoli_feed, "Tcoli_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$Tcoli)

        Shapiro-Wilk normality test

data:  data1$Tcoli
W = 0.98603, p-value = 0.9533

> # Variance
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> var.test(data11$Tcoli, data12$Tcoli)

        F test to compare two variances

data:  data11$Tcoli and data12$Tcoli
F = 41.81, num df = 17, denom df = 7, p-value = 4.231e-05
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
   9.248661 131.933864
sample estimates:
ratio of variances
          41.80974

> var.test(data11$Tcoli, data13$Tcoli)

        F test to compare two variances

data:  data11$Tcoli and data13$Tcoli
F = 5.6033, num df = 17, denom df = 3, p-value = 0.181
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.3942471 22.4758856
sample estimates:
ratio of variances
          5.603334

> var.test(data12$Tcoli, data13$Tcoli)

        F test to compare two variances

data:  data12$Tcoli and data13$Tcoli
F = 0.13402, num df = 7, denom df = 3, p-value = 0.02772
```

```
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.009164127 0.789352452
sample estimates:
ratio of variances
         0.1340198

>
> # Significance
> kruskal.test(Tcoli ~ Feed_amount, data = data1)

        Kruskal-Wallis rank sum test

data:  Tcoli by Feed_amount
Kruskal-Wallis chi-squared = 2.0048, df = 2, p-value = 0.367

>
> ## Normality check on data2
> shapiro.test(data2$Tcoli)

        Shapiro-Wilk normality test

data:  data2$Tcoli
W = 0.96111, p-value = 0.3703

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$Tcoli, data22$Tcoli)

        F test to compare two variances

data:  data21$Tcoli and data22$Tcoli
F = 2.8777, num df = 15, denom df = 7, p-value = 0.1627
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6300043 9.4773862
sample estimates:
ratio of variances
          2.877726

> var.test(data21$Tcoli, data23$Tcoli)

        F test to compare two variances

data:  data21$Tcoli and data23$Tcoli
F = 0.75398, num df = 15, denom df = 3, p-value = 0.6058
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05290088 3.13113541
sample estimates:
ratio of variances
          0.753981

> var.test(data22$Tcoli, data23$Tcoli)

        F test to compare two variances

data:  data22$Tcoli and data23$Tcoli
F = 0.26201, num df = 7, denom df = 3, p-value = 0.1314
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01791567 1.54316718
```

```
sample estimates:
ratio of variances
         0.2620059

>
> # Significance
> fit2 <- aov(Tcoli~Feed_amount, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2 0.7532  0.3766   6.537 0.0052 **
Residuals   25 1.4401  0.0576
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Tcoli ~ Feed_amount, data = data2)

$Feed_amount
               diff         lwr       upr     p adj
1000-800  0.1981543 -0.06070990 0.4570185 0.1576858
1200-800  0.4642863  0.13009405 0.7984785 0.0053432
1200-1000 0.2661320 -0.09995726 0.6322212 0.1866661

>
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=Tcoli)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Total coliform (Log)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="Tcoli",
+                              groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount    Tcoli        sd
1         800 6.257871 0.9745939
2        1000 5.817774 0.1507249
3        1200 6.087790 0.4117186
>

> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=Tcoli)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("Total coliform (Log)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
```

```
+          axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="Tcoli",
+                             groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount    Tcoli        sd
1         800 7.492791 0.2593434
2        1000 7.690945 0.1528799
3        1200 7.957077 0.2986723
```

13. E. Coli

```
> # Define factors for metadata -----
> metadata$Feed_amount <- factor(metadata$Feed_amount)
> metadata$Water_type <- factor(metadata$Water_type)
>
> #Data summary
> data1<-metadata[which(metadata$Water_type=="Effluent"),]
>
> Ecoli_effluent_data <- data_summary(data1, varname="Ecoli",
+                                      groupnames=c("Feed_amount"))
> Ecoli_effluent_data
  Feed_amount    Ecoli        sd
1         800 5.350805 0.9869731
2        1000 5.008772 0.5052080
3        1200 4.980259 0.5625779
> #write.csv(Ecoli_effluent, "Ecoli_effluent.csv")
>
> data2<-metadata[which(metadata$Water_type=="Feed"),]
>
> Ecoli_feed_data <- data_summary(data2, varname="Ecoli",
+                                  groupnames=c("Feed_amount"))
> Ecoli_feed_data
  Feed_amount    Ecoli        sd
1         800 6.768116 0.2774839
2        1000 7.037896 0.2073451
3        1200 7.196399 0.3075493
> #write.csv(Ecoli_feed, "Ecoli_feed.csv")
>
> # Statistical analysis
>
> # Normality and equal variance
> ## Normality check on data1
> shapiro.test(data1$Ecoli)

        Shapiro-Wilk normality test

data:  data1$Ecoli
W = 0.97031, p-value = 0.6529

> # Variance
> data11<-data1[which(data1$Feed_amount=="800"),]
> data12<-data1[which(data1$Feed_amount=="1000"),]
> data13<-data1[which(data1$Feed_amount=="1200"),]
> var.test(data11$Ecoli, data12$Ecoli)

        F test to compare two variances

data:  data11$Ecoli and data12$Ecoli
F = 3.8165, num df = 12, denom df = 7, p-value = 0.08476
```

```
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.8179774 13.7644184
sample estimates:
ratio of variances
          3.816543

> var.test(data11$Ecoli, data13$Ecoli)

        F test to compare two variances

data:  data11$Ecoli and data13$Ecoli
F = 3.0778, num df = 12, denom df = 3, p-value = 0.3852
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.2146844 13.7707999
sample estimates:
ratio of variances
          3.077834

> var.test(data12$Ecoli, data13$Ecoli)

        F test to compare two variances

data:  data12$Ecoli and data13$Ecoli
F = 0.80645, num df = 7, denom df = 3, p-value = 0.7302
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05514386 4.74981882
sample estimates:
ratio of variances
         0.8064456

>
> # Significance
> fit1 <- aov(Ecoli~Feed_amount, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2  0.773  0.3866    0.59  0.563
Residuals   22 14.426  0.6557
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ecoli ~ Feed_amount, data = data1)

$Feed_amount
                 diff       lwr       upr     p adj
1000-800  -0.34203215 -1.256099 0.5720346 0.6214134
1200-800  -0.37054561 -1.533621 0.7925294 0.7067352
1200-1000 -0.02851345 -1.274176 1.2171495 0.9981789

>
> ## Normality check on data2
> shapiro.test(data2$Ecoli)

        Shapiro-Wilk normality test

data:  data2$Ecoli
W = 0.95841, p-value = 0.3838

> # Equal variance check for data2
> data21<-data2[which(data2$Feed_amount=="800"),]
```

```
> data22<-data2[which(data2$Feed_amount=="1000"),]
> data23<-data2[which(data2$Feed_amount=="1200"),]
> var.test(data21$Ecoli, data22$Ecoli)

        F test to compare two variances

data:  data21$Ecoli and data22$Ecoli
F = 1.791, num df = 12, denom df = 7, p-value = 0.4485
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.383848 6.459158
sample estimates:
ratio of variances
          1.790969

> var.test(data21$Ecoli, data23$Ecoli)

        F test to compare two variances

data:  data21$Ecoli and data23$Ecoli
F = 0.81404, num df = 12, denom df = 3, p-value = 0.6841
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.05678079 3.64216914
sample estimates:
ratio of variances
         0.8140408

> var.test(data22$Ecoli, data23$Ecoli)

        F test to compare two variances

data:  data22$Ecoli and data23$Ecoli
F = 0.45453, num df = 7, denom df = 3, p-value = 0.3518
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.03107993 2.67707148
sample estimates:
ratio of variances
         0.4545252

>
> # Significance
> fit2 <- aov(Ecoli~Feed_amount, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Feed_amount  2 0.7165  0.3582   5.224 0.0139 *
Residuals   22 1.5087  0.0686
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ecoli ~ Feed_amount, data = data2)

$Feed_amount
               diff         lwr       upr     p adj
1000-800  0.2697799 -0.02582362 0.5653833 0.0779633
1200-800  0.4282825  0.05215131 0.8044137 0.0237357
1200-1000 0.1585026 -0.24433700 0.5613423 0.5916983

>
```

```
> # Plot
> ## Data 1 Effluent
> box_1 <- ggplot(data1, aes(x=Feed_amount, y=Ecoli)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("E. coli (Log)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 1 effluent
> box_1_data <- data_summary(data1, varname="Ecoli",
+                            groupnames=c("Feed_amount"))
> box_1_data
  Feed_amount    Ecoli        sd
1         800 5.350805 0.9869731
2        1000 5.008772 0.5052080
3        1200 4.980259 0.5625779
>


> ## Data 2 Feed
> box_2 <- ggplot(data2, aes(x=Feed_amount, y=Ecoli)) +
+   geom_violin(trim=TRUE, fill="gray") +
+   xlab("Feed Amount (Gallon/day)")+
+   ylab("E. coli (Log)") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"), legen
d.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>
> ## Mean and standard deviation for Data 2 feed
> box_2_data <- data_summary(data2, varname="Ecoli",
+                           groupnames=c("Feed_amount"))
> box_2_data
  Feed_amount    Ecoli        sd
1         800 6.768116 0.2774839
2        1000 7.037896 0.2073451
3        1200 7.196399 0.3075493
```

Microbial community analysis for the blackwater treatment

```
> ##Choose Blackwater_Frequency_Percentage_table.txt
> con <- file.choose(new = FALSE)
> ##choose Blackwater_Frequency_Table_Taxanomy.txt
> con1 <-file.choose(new = FALSE)
> OTU_Table <- read.table(con, header = T, row.names = 1)
> OTU_Table_taxonomy <- read.delim(con1, header = T, row.names = 1)
```

```
> # Dendogram ------------------------------------------------------------
>
> t.OTU.table <- t(OTU_Table)  # Conversion the data transposically
> class(t.OTU.table) # Check the class of the table
[1] "matrix"
> View(t.OTU.table)
>
> distance <-vegdist(t.OTU.table, method="euclidean") ## Production of Distan
ce Matrix
> cluster <- hclust(distance, method="complete", members = NULL) ## Productio
n of Hierarchical Cluster Production
> tree_m <- plot(cluster, xlab = "Samples", sub = NULL, main ="Dendogram")
>
> range(distance)
[1]  5.012491 76.377907
> rect.hclust(cluster, k = 3, border = "red")
> grp <- cutree(cluster, k = 3)

> ## Abundances ---------------------------------------------------
>
> #Phyloseq
> Full_OTU <- cbind.data.frame(OTU_Table, OTU_Table_taxonomy)
> View(OTU_Table_taxonomy) #Taxonomy table
> OTU <- otu_table(OTU_Table,taxa_are_rows = TRUE) # OTU Table production for
phyloseq
> TAX <- tax_table(as.matrix(OTU_Table_taxonomy)) ## Taxanomy production for
phyloseq
> # SAM <- sample_data(metadata)
> physeq <- phyloseq(OTU, TAX) ##physeq document production
> physeq0 <- tax_glom(physeq, taxrank=rank_names(physeq)[6], NArm=TRUE, bad_e
mpty=c(NA, "", " ", "\t"))
> tax_table(physeq0)
Taxonomy Table:     [48 taxa by 6 taxonomic ranks]:
            Domain      Phylum                  Class
Frequency1  "Unassigned" "Unassigned"           "Unassigned"
Frequency2  "Bacteria"   "Bacteria_unclassified" "Bacteria_unclassified"
Frequency3  "Bacteria"   "Actinobacteria"       "Actinobacteria_unclassified
"
Frequency4  "Bacteria"   "Actinobacteria"       "Actinobacteria"
Frequency5  "Bacteria"   "Actinobacteria"       "Actinobacteria"
Frequency6  "Bacteria"   "Actinobacteria"       "Actinobacteria"
Frequency7  "Bacteria"   "Bacteroidetes"        "Bacteroidetes_unclassified"
Frequency8  "Bacteria"   "Bacteroidetes"        "Cytophagia"
Frequency9  "Bacteria"   "Bacteroidetes"        "Cytophagia"
Frequency10 "Bacteria"   "Bacteroidetes"        "Flavobacteriia"
Frequency11 "Bacteria"   "Bacteroidetes"        "Flavobacteriia"
Frequency12 "Bacteria"   "Bacteroidetes"        "Sphingobacteriia"
Frequency13 "Bacteria"   "Bacteroidetes"        "[Saprospirae]"
Frequency14 "Bacteria"   "Bacteroidetes"        "[Saprospirae]"
Frequency15 "Bacteria"   "Bacteroidetes"        "[Saprospirae]"
Frequency16 "Bacteria"   "Cyanobacteria"        "Cyanobacteria_unclassified"
Frequency17 "Bacteria"   "Firmicutes"           "Bacilli"
Frequency18 "Bacteria"   "Firmicutes"           "Bacilli"
Frequency19 "Bacteria"   "Firmicutes"           "Clostridia"
Frequency20 "Bacteria"   "Firmicutes"           "Clostridia"
Frequency21 "Bacteria"   "Firmicutes"           "Clostridia"
Frequency23 "Bacteria"   "Planctomycetes"       "Planctomycetia"
Frequency24 "Bacteria"   "Proteobacteria"       "Proteobacteria_unclassified
"
Frequency25 "Bacteria"   "Proteobacteria"       "Alphaproteobacteria"
Frequency26 "Bacteria"   "Proteobacteria"       "Alphaproteobacteria"
Frequency27 "Bacteria"   "Proteobacteria"       "Alphaproteobacteria"
Frequency28 "Bacteria"   "Proteobacteria"       "Alphaproteobacteria"
Frequency29 "Bacteria"   "Proteobacteria"       "Alphaproteobacteria"
```

```
Frequency30 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency31 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency32 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency33 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency34 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency35 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency36 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency37 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency38 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency39 "Bacteria"   "Proteobacteria"        "Alphaproteobacteria"
Frequency40 "Bacteria"   "Proteobacteria"        "Betaproteobacteria"
Frequency41 "Bacteria"   "Proteobacteria"        "Betaproteobacteria"
Frequency42 "Bacteria"   "Proteobacteria"        "Betaproteobacteria"
Frequency43 "Bacteria"   "Proteobacteria"        "Epsilonproteobacteria"
Frequency44 "Bacteria"   "Proteobacteria"        "Gammaproteobacteria"
Frequency45 "Bacteria"   "Proteobacteria"        "Gammaproteobacteria"
Frequency46 "Bacteria"   "Verrucomicrobia"       "Verrucomicrobia_unclassifie
d"
Frequency47 "Bacteria"   "Verrucomicrobia"       "Verrucomicrobiae"
Frequency48 "Bacteria"   "Verrucomicrobia"       "Verrucomicrobiae"
Frequency49 "Bacteria"   "Verrucomicrobia"       "Verrucomicrobiae"
            Order                               Family
Frequency1  "Unassigned"                        "Unassigned"
Frequency2  "Bacteria_unclassified"             "Bacteria_unclassified"
Frequency3  "Actinobacteria_unclassified"       "Actinobacteria_unclassified"
Frequency4  "Actinomycetales"                   "Actinomycetales_unclassified"
Frequency5  "Actinomycetales"                   "Micrococcaceae"
Frequency6  "Actinomycetales"                   "Streptomycetaceae"
Frequency7  "Bacteroidetes_unclassified"        "Bacteroidetes_unclassified"
Frequency8  "Cytophagales"                      "Cytophagales_unclassified"
Frequency9  "Cytophagales"                      "Cyclobacteriaceae"
Frequency10 "Flavobacteriales"                  "Flavobacteriales_unclassified
"
Frequency11 "Flavobacteriales"                  "Flavobacteriaceae"
Frequency12 "Sphingobacteriales"                "Sphingobacteriaceae"
Frequency13 "[Saprospirae]_unclassified"        "[Saprospirae]_unclassified"
Frequency14 "[Saprospirales]"                   "[Saprospirales]_unclassified"
Frequency15 "[Saprospirales]"                   "Chitinophagaceae"
Frequency16 "Cyanobacteria_unclassified"        "Cyanobacteria_unclassified"
Frequency17 "Bacilli_unclassified"              "Bacilli_unclassified"
Frequency18 "Bacillales"                        "Bacillales_unclassified"
Frequency19 "Clostridiales"                     "Clostridiales_unclassified"
Frequency20 "Clostridiales"                     "Lachnospiraceae"
Frequency21 "Clostridiales"                     "Peptostreptococcaceae"
Frequency23 "Pirellulales"                      "Pirellulaceae"
Frequency24 "Proteobacteria_unclassified"       "Proteobacteria_unclassified"
Frequency25 "Alphaproteobacteria_unclassified"  "Alphaproteobacteria_unclassif
ied"
Frequency26 "Caulobacterales"                   "Caulobacteraceae"
Frequency27 "Caulobacterales"                   "Caulobacteraceae"
Frequency28 "Caulobacterales"                   "Caulobacteraceae"
Frequency29 "Rhizobiales"                       "Rhizobiales_unclassified"
Frequency30 "Rhizobiales"                       "Bradyrhizobiaceae"
Frequency31 "Rhizobiales"                       "Hyphomicrobiaceae"
Frequency32 "Rhizobiales"                       "Methylobacteriaceae"
Frequency33 "Rhizobiales"                       "Phyllobacteriaceae"
Frequency34 "Rhodospirillales"                  "Rhodobacteraceae"
Frequency35 "Rhodospirillales"                  "Rhodospirillales_unclassified
"
Frequency36 "Rhodospirillales"                  "Acetobacteraceae"
Frequency37 "Rhodospirillales"                  "Rhodospirillaceae"
Frequency38 "Sphingomonadales"                  "Sphingomonadales_unclassified
"
Frequency39 "Sphingomonadales"                  "Sphingomonadaceae"
```

```
Frequency40 "Betaproteobacteria_unclassified"   "Betaproteobacteria_unclassifi
ed"
Frequency41 "Burkholderiales"                   "Burkholderiales_unclassified"
Frequency42 "Neisseriales"                      "Neisseriaceae"
Frequency43 "Campylobacterales"                 "Helicobacteraceae"
Frequency44 "Gammaproteobacteria_unclassified" "Gammaproteobacteria_unclassif
ied"
Frequency45 "Xanthomonadales"                   "Xanthomonadaceae"
Frequency46 "Verrucomicrobia_unclassified"      "Verrucomicrobia_unclassified"
Frequency47 "Verrucomicrobiales"                "Verrucomicrobiaceae"
Frequency48 "Verrucomicrobiales"                "Verrucomicrobiaceae"
Frequency49 "Verrucomicrobiales"                "Verrucomicrobiaceae"
            Genus
Frequency1  "Unassigned"
Frequency2  "Bacteria_unclassified"
Frequency3  "Actinobacteria_unclassified"
Frequency4  "Actinomycetales_unclassified"
Frequency5  "Arthrobacter"
Frequency6  "Streptomycetaceae_unclassified"
Frequency7  "Bacteroidetes_unclassified"
Frequency8  "Cytophagales_unclassified"
Frequency9  "Cyclobacteriaceae_unclassified"
Frequency10 "Flavobacteriales_unclassified"
Frequency11 "Flavobacteriaceae_unclassified"
Frequency12 "Sphingobacteriaceae_unclassified"
Frequency13 "[Saprospirae]_unclassified"
Frequency14 "[Saprospirales]_unclassified"
Frequency15 "Chitinophagaceae_unclassified"
Frequency16 "Cyanobacteria_unclassified"
Frequency17 "Bacilli_unclassified"
Frequency18 "Bacillales_unclassified"
Frequency19 "Clostridiales_unclassified"
Frequency20 "Lachnospiraceae_unclassified"
Frequency21 "Clostridium"
Frequency23 "Pirellulaceae_unclassified"
Frequency24 "Proteobacteria_unclassified"
Frequency25 "Alphaproteobacteria_unclassified"
Frequency26 "Caulobacteraceae_unclassified"
Frequency27 "Brevundimonas"
Frequency28 "Nitrobacteria"
Frequency29 "Rhizobiales_unclassified"
Frequency30 "Bradyrhizobiaceae_unclassified"
Frequency31 "Hyphomicrobiaceae_unclassified"
Frequency32 "Methylobacteriaceae_unclassified"
Frequency33 "Phyllobacteriaceae_unclassified"
Frequency34 "Rhodobacteraceae_unclassified"
Frequency35 "Rhodospirillales_unclassified"
Frequency36 "Roseomonas"
Frequency37 "Rhodospirillaceae_unclassified"
Frequency38 "Sphingomonadales_unclassified"
Frequency39 "Sphingomonadaceae_unclassified"
Frequency40 "Betaproteobacteria_unclassified"
Frequency41 "Burkholderiales_unclassified"
Frequency42 "Neisseriaceae_unclassified"
Frequency43 "Helicobacter"
Frequency44 "Gammaproteobacteria_unclassified"
Frequency45 "Xanthomonadaceae_unclassified"
Frequency46 "Verrucomicrobia_unclassified"
Frequency47 "Verrucomicrobiaceae_unclassified"
Frequency48 "Haloferula"
Frequency49 "Verrucomicrobium"
> p = plot_bar(physeq0, fill = "Family", facet_grid=Domain~Phylum)
> p
```

```
> p + geom_bar(aes(color=Phylum, fill=Phylum), stat = "identity", position =
"stack")

> # Abundance Plotbar Domain
> physeqa <-tax_glom(physeq, taxrank=rank_names(physeq)[1], NArm=TRUE, bad_em
pty=c(NA, "", " ", "\t"))
> tablea <- otu_table(physeqa)
> write.csv(tablea, "domain.csv")
>
> a = plot_bar(physeqa, fill = "Domain") +
+   geom_bar(aes(color=Domain, fill=Domain), stat = "identity", position = "s
tack") +
+   xlab("") + ylab("Relative Abundance (%)") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> a
> a+scale_x_discrete(limits=c("S1", "S2",
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
+                              "S9", "S10", "S11",
+                              "S12", "S13", "S14",
+                              "S15"),
+                     labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                              "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                              "S5"="800 GPD AT day 27", "S6"="800 GPD at day
30",
+                              "S7"="800 GPD at day 31", "S8"="800 GPD at day
34",
+                              "S9"="800 GPD at day 38", "S10"="800 GPD at day
44",
+                              "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                              "S13"="1000 GPD at day 58", "S14"="1200 GPD at
day 73",
+                              "S15"="1200 GPD at day 74"))

#Abundance Plotbar Phylum
> physeqa1 <-tax_glom(physeq, taxrank=rank_names(physeq)[2], NArm=TRUE, bad_e
mpty=c(NA, "", " ", "\t"))
> tablea1 <- otu_table(physeqa1)
> write.csv(tablea1, "Phylum.csv")
>
> a1 = plot_bar(physeqa1, fill = "Phylum") +
+   geom_bar(aes(color=Phylum, fill=Phylum), stat = "identity", position = "s
tack") +
+   xlab("") + ylab("Relative Abundance (%)") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> a1
> a1+scale_x_discrete(limits=c("S1", "S2",
```

```
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
+                              "S9", "S10", "S11",
+                              "S12", "S13", "S14",
+                              "S15"),
+                     labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                              "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                              "S5"="800 GPD AT day 27", "S6"="800 GPD at day 
30",
+                              "S7"="800 GPD at day 31", "S8"="800 GPD at day 
34",
+                              "S9"="800 GPD at day 38", "S10"="800 GPD at day 
44",
+                              "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                              "S13"="1000 GPD at day 58", "S14"="1200 GPD at 
day 73",
+                              "S15"="1200 GPD at day 74"))


> ## Abundance Plotbar Bacteria at family level------------
> 
> #Abundance Plotbar Bacteroidetes (Family)
> physeq3 <-subset_taxa(physeq, Phylum == "Bacteroidetes")
> physeq3_1 <-tax_glom(physeq3, taxrank=rank_names(physeq3)[5], NArm=TRUE, ba
d_empty=c(NA, "", " ", "\t"))
> table3_1 <- otu_table(physeq3_1)
> write.csv(table3_1, "BacteroidetesFamily.csv")
> 
> d = plot_bar(physeq3_1, fill = "Family")+ geom_bar(aes(color=Family, fill=F
amily), stat = "identity",position = "stack") +
+   xlab("") + ylab("Bacteroidetes Abundance (%)") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> d
> d+scale_x_discrete(limits=c("S1", "S2",
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
+                              "S9", "S10", "S11",
+                              "S12", "S13", "S14",
+                              "S15"),
+                     labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                              "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                              "S5"="800 GPD AT day 27", "S6"="800 GPD at day 
30",
+                              "S7"="800 GPD at day 31", "S8"="800 GPD at day 
34",
+                              "S9"="800 GPD at day 38", "S10"="800 GPD at day 
44",
+                              "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                              "S13"="1000 GPD at day 58", "S14"="1200 GPD at 
day 73",
+                              "S15"="1200 GPD at day 74"))
```

```
> #Abundance Plotbar Firmicutes (Family)
> physeq4 <-subset_taxa(physeq, Phylum == "Firmicutes")
> physeq4_1 <-tax_glom(physeq4, taxrank=rank_names(physeq4)[5], NArm=TRUE, ba
d_empty=c(NA, "", " ", "\t"))
> table4_1 <- otu_table(physeq4_1)
> write.csv(table4_1, "FirmicutesFamily.csv")
>
> e = plot_bar(physeq4_1, fill = "Family")+ geom_bar(aes(color=Family, fill=F
amily), stat = "identity",position = "stack") +
+   ylab("Firmicutes Abundance (%)") + xlab("Samples") + labs(title = "") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> e
> e+scale_x_discrete(limits=c("S1", "S2",
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
+                              "S9", "S10", "S11",
+                              "S12", "S13", "S14",
+                              "S15"),
+                     labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                              "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                              "S5"="800 GPD AT day 27", "S6"="800 GPD at day
30",
+                              "S7"="800 GPD at day 31", "S8"="800 GPD at day
34",
+                              "S9"="800 GPD at day 38", "S10"="800 GPD at day
44",
+                              "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                              "S13"="1000 GPD at day 58", "S14"="1200 GPD at
day 73",
+                              "S15"="1200 GPD at day 74"))


> #Abundance Plotbar Actinobacteria (Family)
> physeq5 <-subset_taxa(physeq, Phylum == "Actinobacteria")
> physeq5_1 <-tax_glom(physeq5, taxrank=rank_names(physeq5)[5], NArm=TRUE, ba
d_empty=c(NA, "", " ", "\t"))
> table5_1 <- otu_table(physeq5_1)
> write.csv(table5_1, "ActinobacteriaFamily.csv")
>
> f = plot_bar(physeq5_1, fill = "Family")+ geom_bar(aes(color=Family, fill=F
amily), stat = "identity",position = "stack") +
+   ylab("Actinobacteria Abundance (%)") + xlab("") + labs(title = "") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> f
> f+scale_x_discrete(limits=c("S1", "S2",
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
```

```
+                               "S9", "S10", "S11",
+                               "S12", "S13", "S14",
+                               "S15"),
+                      labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                               "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                               "S5"="800 GPD AT day 27", "S6"="800 GPD at day
30",
+                               "S7"="800 GPD at day 31", "S8"="800 GPD at day
34",
+                               "S9"="800 GPD at day 38", "S10"="800 GPD at day
44",
+                               "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                               "S13"="1000 GPD at day 58", "S14"="1200 GPD at
day 73",
+                               "S15"="1200 GPD at day 74"))

> #Abundance Plotbar Proteobacteria (Family)
> physeq6 <-subset_taxa(physeq, Phylum == "Proteobacteria")
> physeq6_1 <-tax_glom(physeq6, taxrank=rank_names(physeq6)[5], NArm=TRUE, ba
d_empty=c(NA, "", " ", "\t"))
> table6_1 <- otu_table(physeq6_1)
> write.csv(table6_1, "ProteobacteriaFamily.csv")
>
> g = plot_bar(physeq6_1, fill = "Family")+ geom_bar(aes(color=Family, fill=F
amily), stat = "identity",position = "stack") +
+   ylab("Proteobacteria Abundance (%)") + xlab("") + labs(title = "") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> g
> g+scale_x_discrete(limits=c("S1", "S2",
+                               "S3", "S4", "S5",
+                               "S6", "S7", "S8",
+                               "S9", "S10", "S11",
+                               "S12", "S13", "S14",
+                               "S15"),
+                      labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                               "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                               "S5"="800 GPD AT day 27", "S6"="800 GPD at day
30",
+                               "S7"="800 GPD at day 31", "S8"="800 GPD at day
34",
+                               "S9"="800 GPD at day 38", "S10"="800 GPD at day
44",
+                               "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                               "S13"="1000 GPD at day 58", "S14"="1200 GPD at
day 73",
+                               "S15"="1200 GPD at day 74"))

> #Abundance Plotbar Verrucomicrobia (Family)
> physeq7 <-subset_taxa(physeq, Phylum == "Verrucomicrobia")
> physeq7_1 <-tax_glom(physeq7, taxrank=rank_names(physeq7)[5], NArm=TRUE, ba
d_empty=c(NA, "", " ", "\t"))
> table7_1 <- otu_table(physeq7_1)
```

```
> write.csv(table7_1, "VerrucomicrobiaFamily.csv")
>
> h = plot_bar(physeq7_1, fill = "Family")+ geom_bar(aes(color=Family, fill=F
amily), stat = "identity",position = "stack") +
+   ylab("Verrucomicrobia Abundance (%)") + xlab("") + labs(title = "") +
+   theme(legend.position="right",
+         axis.text.x = element_text(size = 18, family="Times New Roman", ang
le = 90, hjust = 1),
+         axis.text.y = element_text(size = 18, family="Times New Roman"),
+         axis.title.x = element_text(size = 18, family="Times New Roman"),
+         axis.title.y = element_text(size = 18, family="Times New Roman"),
+         legend.text = element_text(size = 18, family="Times New Roman"),
+         legend.title= element_text(size = 18, family="Times New Roman"))
> h
> h+scale_x_discrete(limits=c("S1", "S2",
+                              "S3", "S4", "S5",
+                              "S6", "S7", "S8",
+                              "S9", "S10", "S11",
+                              "S12", "S13", "S14",
+                              "S15"),
+                     labels=c("S1"="Blackwater", "S2"="800 GPD AT day 5",
+                              "S3"="800 GPD AT day 8", "S4"="800 GPD AT day 2
0",
+                              "S5"="800 GPD AT day 27", "S6"="800 GPD at day
30",
+                              "S7"="800 GPD at day 31", "S8"="800 GPD at day
34",
+                              "S9"="800 GPD at day 38", "S10"="800 GPD at day
44",
+                              "S11"="900 GPD at day 50", "S12"="900 GPD at da
y 51",
+                              "S13"="1000 GPD at day 58", "S14"="1200 GPD at
day 73",
+                              "S15"="1200 GPD at day 74"))


R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> ## NMDS analysis for the blackwater treatment
> ## Wei Liao, April 30, 2020
>
> # Loading Library and Tables ----------------
>
> # Load "vegan" and "MASS" libraries in R
>
> library(vegan)
Loading required package: permute
Loading required package: lattice
This is vegan 2.5-6
> library(MASS)
```

```
> species <- read.csv(file.choose(), head = TRUE, row.names = 1)
> env <- read.csv(file.choose(), head = TRUE, row.names = 1)
> performance <- read.csv(file.choose(), head= TRUE, row.names = 1)
> rarecurve(species, step=20, min(rowSums(species)), label=TRUE)

> # Statistical analysis ----------------------
> # When this step is done, type "species.mds" or "ef.sp" to obtain the stati
stical results
>
> species.mds <- metaMDS(species, trace=FALSE)
> ef.sp <- envfit(species.mds, env, permu=999)
> perf.sp <- envfit(species.mds, performance, permu=999)
> species.mds

Call:
metaMDS(comm = species, trace = FALSE)

global Multidimensional Scaling using monoMDS

Data:     wisconsin(sqrt(species))
Distance: bray

Dimensions: 2
Stress:     0.1422723
Stress type 1, weak ties
Two convergent solutions found after 20 tries
Scaling: centring, PC rotation, halfchange scaling
Species: expanded scores based on ‘wisconsin(sqrt(species))’

> ef.sp

***VECTORS

               NMDS1    NMDS2     r2 Pr(>r)
Feed_amount -0.51469  0.85738 0.6547  0.004 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999


> perf.sp

***VECTORS

                                            NMDS1    NMDS2     r2 Pr(>r)
Turbidity                                -0.26301  0.96479 0.1830  0.314
TS                                        0.33708 -0.94148 0.6794  0.006 **
TSS                                      -0.26966  0.96296 0.0598  0.688
COD                                      -0.93623 -0.35139 0.2738  0.181
NH3                                       0.93288 -0.36020 0.1089  0.540
NO2                                      -0.45895 -0.88846 0.3174  0.121
NO3                                       0.49927 -0.86645 0.2933  0.165
TKN                                       1.00000  0.00270 0.5512  0.018 *
TP                                        0.39066 -0.92053 0.2055  0.288
TOC                                       0.98687 -0.16152 0.0454  0.782
Proteobacteria_phylum                    -0.89944  0.43705 0.8023  0.001 ***
Proteobacteria_unclassified_family       -0.78295 -0.62209 0.5424  0.008 **
Alphaproteobacteria_unclassified_family -0.39404 -0.91909 0.4341  0.034 *
Caulobacteraceae_family                   0.99978  0.02089 0.5659  0.015 *
Rhizobiales_unclassified_family          -0.03813  0.99927 0.5721  0.017 *
Rhodobacteraceae_family                  -0.28556 -0.95836 0.7847  0.001 ***
Sphingomonadales_unclassified_family     -0.23373  0.97230 0.4537  0.054 .
Betaproteobacteria_unclassified_family  -0.70536 -0.70885 0.3985  0.043 *
```

```
Burkholderiales_unclassified_family      -0.53845  0.84265 0.4226  0.052 .
Gammaproteobacteria_unclassified_family -0.57398 -0.81887 0.6578  0.004 **
Xanthomonadaceae_family                  -0.12782 -0.99180 0.6262  0.006 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999


> # Name the performance parameters
> names(performance)[names(performance)=="TS"]<- "TS (p=0.003)"
> names(performance)[names(performance)=="TKN"]<- "TKN (p=0.021)"
> names(performance)[names(performance)=="Proteobacteria_phylum"]<- "Phylum P
roteobacteria (p=0.001)"
> names(performance)[names(performance)=="Proteobacteria_unclassified_family"
]<- "Unclassified Proteobacteria family (p=0.004)"
> names(performance)[names(performance)=="Alphaproteobacteria_unclassified_fa
mily"]<- "Unclassified Alphaproteobacteria family (p=0.045)"
> names(performance)[names(performance)=="Caulobacteraceae_family"]<- "Caulob
acteraceae (p=0.012)"
> names(performance)[names(performance)=="Rhizobiales_unclassified_family"]<-
"Unclassifed Rhizobiales family (p=0.017)"
> names(performance)[names(performance)=="Rhodobacteraceae_family"]<- "Rhodob
acteraceae (p=0.001)"
> names(performance)[names(performance)=="Sphingomonadales_unclassified_famil
y"]<- "Unclassified Sphingomonadales family (p=0.036)"
> names(performance)[names(performance)=="Betaproteobacteria_unclassified_fam
ily"]<- "Unclassified Betaproteobacteria family (p=0.042)"
> names(performance)[names(performance)=="Gammaproteobacteria_unclassified_fa
mily"]<- "Unclassified Gammaproteobacteria family (p=0.004)"
> names(performance)[names(performance)=="Xanthomonadaceae_family"]<- "Xantho
monadaceae (p=0.006)"
>
> # Plotting NMDS chart -------------------
> plot(species.mds, display="sites", type="points")
> #with(env, ordiellipse(species.mds, Feed_amount, kind= "se", draw="polygon"
, col="green", alpha=50, label=TRUE,border=NA, conf=0.95))
> with(env, ordisurf(species.mds, Feed_amount, main="", labcex = 0.6, add=TRU
E, col="dark green", alpha=50, label=TRUE, border=NA, conf=0.95))

Family: gaussian
Link function: identity

Formula:
y ~ s(x1, x2, k = 10, bs = "tp", fx = FALSE)

Estimated degrees of freedom:
3.69  total = 4.69

REML score: 81.29275
>
> #Plot the performance parameters
> ef.perf <- envfit(species.mds, performance[, c(2, 6, 7, 8)], permu=999)
> plot(ef.perf, col="red", cex=1.0)
>
> #Plot the significant bacterial families
> ef.perf <- envfit(species.mds, performance[, c(11,13,14,15,16,17,18,20,21)]
, permu=999)
> plot(ef.perf, col="blue", cex=0.7)
```

CHAPTER 3

WATER QUALITY

UV254

```
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater UV254
7  Effluent    1      PPG          S 0.037
9  Effluent    2      PPG          S 0.027
10 Effluent    2      PPG          S 0.015
11 Effluent    2      PPG          S 0.014
12 Effluent    3      PPG          S 0.025
13 Effluent    3      PPG          S 0.017
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)


> # Select treated sample data for shower wastewater----


> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
      Sample Cell Membrane Wastewater UV254
7   Effluent    1      PPG          S 0.037
9   Effluent    2      PPG          S 0.027
10  Effluent    2      PPG          S 0.015
11  Effluent    2      PPG          S 0.014
12  Effluent    3      PPG          S 0.025
13  Effluent    3      PPG          S 0.017
14  Effluent    3      PPG          S 0.016
96  Effluent    1     PVDF          S 0.033
97  Effluent    1     PVDF          S 0.036
98  Effluent    1     PVDF          S 0.037
99  Effluent    1     PVDF          S 0.031
100 Effluent    2     PVDF          S 0.031
101 Effluent    2     PVDF          S 0.040
102 Effluent    2     PVDF          S 0.039
103 Effluent    2     PVDF          S 0.031
104 Effluent    3     PVDF          S 0.030
105 Effluent    3     PVDF          S 0.040
106 Effluent    3     PVDF          S 0.039
107 Effluent    3     PVDF          S 0.030
138 Effluent    1      PES          S 0.024
139 Effluent    1      PES          S 0.022
142 Effluent    2      PES          S 0.026
143 Effluent    2      PES          S 0.026
146 Effluent    3      PES          S 0.024
147 Effluent    3      PES          S 0.021

> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
> # Statistical analysis on data1
>
> fit1 <- aov(UV254~Membrane, data1)
> summary(fit1)
            Df    Sum Sq   Mean Sq F value   Pr(>F)
Membrane     2 0.0009354 0.0004677   16.06 5.02e-05 ***
Residuals   22 0.0006408 0.0000291
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = UV254 ~ Membrane, data = data1)

$Membrane
                 diff          lwr         upr     p adj
PPG-PES  -0.002261905 -0.009804602 0.005280793 0.7348260
PVDF-PES  0.010916667  0.004137916 0.017695417 0.0015051
PVDF-PPG  0.013178571  0.006730693 0.019626449 0.0001090


> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=UV254)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("UV254") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="UV254",
+                            groupnames=c("Membrane"))
> box_1_data
  Membrane      UV254          sd
1      PES 0.02383333 0.002041241
2      PPG 0.02157143 0.008482475
3     PVDF 0.03475000 0.004136863


> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater UV254
29  Effluent    1      PPG          L 0.031
30  Effluent    1      PPG          L 0.052
31  Effluent    1      PPG          L 0.028
32  Effluent    1      PPG          L 0.027
33  Effluent    1      PPG          L 0.024
34  Effluent    1      PPG          L 0.024
35  Effluent    2      PPG          L 0.033
36  Effluent    2      PPG          L 0.033
37  Effluent    2      PPG          L 0.028
38  Effluent    2      PPG          L 0.025
39  Effluent    2      PPG          L 0.023
40  Effluent    2      PPG          L 0.024
41  Effluent    2      PPG          L 0.023
42  Effluent    3      PPG          L 0.049
43  Effluent    3      PPG          L 0.034
44  Effluent    3      PPG          L 0.027
```

```
45  Effluent    3      PPG          L 0.025
46  Effluent    3      PPG          L 0.025
47  Effluent    3      PPG          L 0.025
57  Effluent    1     PVDF          L 0.074
58  Effluent    1     PVDF          L 0.089
59  Effluent    1     PVDF          L 0.098
60  Effluent    1     PVDF          L 0.069
61  Effluent    2     PVDF          L 0.074
62  Effluent    2     PVDF          L 0.080
63  Effluent    2     PVDF          L 0.087
64  Effluent    2     PVDF          L 0.067
65  Effluent    3     PVDF          L 0.081
66  Effluent    3     PVDF          L 0.089
67  Effluent    3     PVDF          L 0.088
68  Effluent    3     PVDF          L 0.072
158 Effluent    1      PES          L 0.023
159 Effluent    1      PES          L 0.072
160 Effluent    1      PES          L 0.062
162 Effluent    2      PES          L 0.032
163 Effluent    2      PES          L 0.069
164 Effluent    2      PES          L 0.061
166 Effluent    3      PES          L 0.036
167 Effluent    3      PES          L 0.082
168 Effluent    3      PES          L 0.066
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(UV254~Membrane, data2)
> summary(fit2)
            Df   Sum Sq  Mean Sq F value   Pr(>F)
Membrane     2 0.019579 0.009789   65.21 7.43e-13 ***
Residuals   37 0.005554 0.000150
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = UV254 ~ Membrane, data = data2)

$Membrane
                diff         lwr         upr     p adj
PPG-PES  -0.02641520 -0.03851975 -0.01431066 0.0000150
PVDF-PES  0.02477778  0.01158715  0.03796840 0.0001459
PVDF-PPG  0.05119298  0.04016284  0.06222312 0.0000000


> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=UV254)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("UV254") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
```

```
> box_2 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="UV254",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane      UV254          sd
1      PES 0.05588889 0.020392673
2      PPG 0.02947368 0.008187853
3     PVDF 0.08066667 0.009632647


> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
      Sample Cell Membrane Wastewater UV254
77  Effluent    1      PPG         SL 0.085
78  Effluent    1      PPG         SL 0.067
79  Effluent    1      PPG         SL 0.063
80  Effluent    2      PPG         SL 0.085
81  Effluent    2      PPG         SL 0.070
82  Effluent    2      PPG         SL 0.066
83  Effluent    2      PPG         SL 0.112
84  Effluent    3      PPG         SL 0.080
118 Effluent    1     PVDF         SL 0.042
119 Effluent    1     PVDF         SL 0.081
120 Effluent    1     PVDF         SL 0.076
121 Effluent    1     PVDF         SL 0.061
122 Effluent    2     PVDF         SL 0.042
123 Effluent    2     PVDF         SL 0.075
124 Effluent    2     PVDF         SL 0.074
125 Effluent    2     PVDF         SL 0.056
126 Effluent    3     PVDF         SL 0.042
127 Effluent    3     PVDF         SL 0.079
128 Effluent    3     PVDF         SL 0.073
129 Effluent    3     PVDF         SL 0.053
179 Effluent    1      PES         SL 0.019
180 Effluent    1      PES         SL 0.048
181 Effluent    1      PES         SL 0.062
182 Effluent    1      PES         SL 0.044
183 Effluent    2      PES         SL 0.025
184 Effluent    2      PES         SL 0.044
185 Effluent    2      PES         SL 0.048
186 Effluent    2      PES         SL 0.049
187 Effluent    3      PES         SL 0.016
188 Effluent    3      PES         SL 0.035
189 Effluent    3      PES         SL 0.039
190 Effluent    3      PES         SL 0.044
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(UV254~Membrane, data3)
> summary(fit3)
            Df   Sum Sq  Mean Sq F value   Pr(>F)    
Membrane     2 0.007786 0.003893   17.63 9.77e-06 ***
Residuals   29 0.006405 0.000221                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = UV254 ~ Membrane, data = data3)

$Membrane
                diff          lwr         upr     p adj
PPG-PES   0.03908333  0.022331560 0.055835106 0.0000090
PVDF-PES  0.02341667  0.008433425 0.038399908 0.0016458
PVDF-PPG -0.01566667 -0.032418440 0.001085106 0.0703058

> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=UV254)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("UV254") + labs(title = "", subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile

> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="UV254",
+                             groupnames=c("Membrane"))
> box_3_data
  Membrane      UV254         sd
1      PES 0.03941667 0.01350056
2      PPG 0.07850000 0.01608016
3     PVDF 0.06283333 0.01534354
```

COD

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - COD
> ## Wei Liao, September 27, 2023
>
> # Load libraries -----
>
>   library (MASS)
>   library(ggplot2)
>   library(grid)
>   library(gridExtra)
>   library(ggpubr)
>   library(plyr)
>   library(inferr)
>   library(extrafont)
>   loadfonts(device="win", quiet=TRUE)
> # Plot bar chart with standard deviation -----
>   #data : a data frame
>   #varname : the name of a column containing the variable to be summarized
>   #groupnames : vector of column names to be used as
>   #grouping variables
>   data_summary <- function(data, varname, groupnames){
+     require(plyr)
+     summary_func <- function(x, col){
+       c(mean = mean(x[[col]], na.rm=TRUE),
```

```
+         sd = sd(x[[col]], na.rm=TRUE))
+     }
+     data_sum<-ddply(data, groupnames, .fun=summary_func,
+                     varname)
+     data_sum <- rename(data_sum, c("mean" = varname))
+     return(data_sum)
+   }
> # Choose data file COD.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
      Sample Cell Membrane Wastewater  COD
7   Effluent    1      PPG          S 80.0
8   Effluent    1      PPG          S 83.7
9   Effluent    2      PPG          S 92.7
10  Effluent    2      PPG          S 79.9
12  Effluent    3      PPG          S 84.1
13  Effluent    3      PPG          S 84.2
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)
> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
       Sample Cell Membrane Wastewater   COD
7    Effluent    1      PPG          S  80.0
8    Effluent    1      PPG          S  83.7
9    Effluent    2      PPG          S  92.7
10   Effluent    2      PPG          S  79.9
12   Effluent    3      PPG          S  84.1
13   Effluent    3      PPG          S  84.2
96   Effluent    1     PVDF          S 202.0
97   Effluent    1     PVDF          S 209.0
98   Effluent    1     PVDF          S 195.0
99   Effluent    1     PVDF          S 198.0
100  Effluent    2     PVDF          S 203.0
101  Effluent    2     PVDF          S 206.0
102  Effluent    2     PVDF          S 197.0
103  Effluent    2     PVDF          S 205.0
104  Effluent    3     PVDF          S 182.0
105  Effluent    3     PVDF          S 187.0
106  Effluent    3     PVDF          S 201.0
107  Effluent    3     PVDF          S 198.0
140  Effluent    1      PES          S  81.2
141  Effluent    1      PES          S  82.2
144  Effluent    2      PES          S  84.9
145  Effluent    2      PES          S  77.3
148  Effluent    3      PES          S  77.9
149  Effluent    3      PES          S  77.1
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
> # Statistical analysis on data1
>
> fit1 <- aov(COD~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2  81458   40729    1036 <2e-16 ***
Residuals   21    826      39
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
```

```
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = COD ~ Membrane, data = data1)

$Membrane
              diff        lwr      upr     p adj
PPG-PES    4.0000  -5.125898  13.1259 0.5217299
PVDF-PES 118.4833 110.580074 126.3866 0.0000000
PVDF-PPG 114.4833 106.580074 122.3866 0.0000000

> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=COD)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("COD (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 300)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="COD",
+                             groupnames=c("Membrane"))
> box_1_data
  Membrane      COD       sd
1      PES  80.1000 3.173011
2      PPG  84.1000 4.660043
3     PVDF 198.5833 7.786449

> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater COD
29  Effluent    1      PPG          L 272
30  Effluent    1      PPG          L 271
31  Effluent    1      PPG          L 188
32  Effluent    1      PPG          L 189
35  Effluent    2      PPG          L 264
36  Effluent    2      PPG          L 268
37  Effluent    2      PPG          L 273
38  Effluent    2      PPG          L 271
42  Effluent    3      PPG          L 266
43  Effluent    3      PPG          L 266
44  Effluent    3      PPG          L 285
45  Effluent    3      PPG          L 297
57  Effluent    1     PVDF          L 178
58  Effluent    1     PVDF          L 174
59  Effluent    1     PVDF          L 171
60  Effluent    1     PVDF          L 169
65  Effluent    3     PVDF          L 204
66  Effluent    3     PVDF          L 195
67  Effluent    3     PVDF          L 157
68  Effluent    3     PVDF          L 159
158 Effluent    1      PES          L 126
159 Effluent    1      PES          L 110
```

```
160 Effluent    1       PES          L 109
161 Effluent    1       PES          L 113
162 Effluent    2       PES          L 130
163 Effluent    2       PES          L 136
164 Effluent    2       PES          L 112
165 Effluent    2       PES          L 111
166 Effluent    3       PES          L 135
167 Effluent    3       PES          L 141
168 Effluent    3       PES          L 122
169 Effluent    3       PES          L 113
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(COD~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2 114967   57483   102.3 7.27e-14 ***
Residuals   29  16294     562
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = COD ~ Membrane, data = data2)

$Membrane
              diff         lwr       upr    p adj
PPG-PES  137.66667  113.76831 161.56503 0.00e+00
PVDF-PES  54.37500   27.65582  81.09418 6.84e-05
PVDF-PPG -83.29167 -110.01084 -56.57249 1.00e-07

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=COD)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("COD (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 300)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile



> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="COD",
+                              groupnames=c("Membrane"))
> box_2_data
  Membrane      COD       sd
1      PES 121.5000 11.67359
2      PPG 259.1667 34.26855
3     PVDF 175.8750 16.37452
```

```
> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
      Sample Cell Membrane Wastewater    COD
77  Effluent    1      PPG         SL 195.00
78  Effluent    1      PPG         SL 200.00
80  Effluent    2      PPG         SL 202.00
81  Effluent    2      PPG         SL 214.00
82  Effluent    2      PPG         SL 134.00
83  Effluent    2      PPG         SL 136.00
84  Effluent    3      PPG         SL 209.00
85  Effluent    3      PPG         SL 206.00
118 Effluent    1     PVDF         SL 179.00
119 Effluent    1     PVDF         SL 175.00
120 Effluent    1     PVDF         SL 136.00
121 Effluent    1     PVDF         SL 154.00
122 Effluent    2     PVDF         SL 191.00
123 Effluent    2     PVDF         SL 195.00
124 Effluent    2     PVDF         SL 153.00
125 Effluent    2     PVDF         SL 153.00
126 Effluent    3     PVDF         SL 172.00
127 Effluent    3     PVDF         SL 172.00
128 Effluent    3     PVDF         SL 156.00
129 Effluent    3     PVDF         SL 156.00
179 Effluent    1      PES         SL  92.35
180 Effluent    1      PES         SL 132.50
181 Effluent    1      PES         SL 117.50
183 Effluent    2      PES         SL 108.50
184 Effluent    2      PES         SL 138.50
185 Effluent    2      PES         SL 120.00
187 Effluent    3      PES         SL 235.00
188 Effluent    3      PES         SL 119.50
189 Effluent    3      PES         SL 114.00
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(COD~Membrane, data3)
> summary(fit3)
            Df Sum Sq Mean Sq F value  Pr(>F)   
Membrane     2  13877    6938    7.39 0.00288 **
Residuals   26  24411     939                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = COD ~ Membrane, data = data3)

$Membrane
              diff        lwr      upr     p adj
PPG-PES   56.12778  19.130353 93.12520 0.0023691
PVDF-PES  35.12778   1.553162 68.70239 0.0389868
PVDF-PPG -21.00000 -55.753029 13.75303 0.3066726

>
```

```
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=COD)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("COD (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 300)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="COD",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane      COD       sd
1      PES 130.8722 41.22943
2      PPG 187.0000 32.60587
3     PVDF 166.0000 17.50325
```

TURBIDITY

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - Turbidity
> ## Wei Liao, September 27, 2023
>
> # Load libraries -----

>   loadfonts(device="win", quiet=TRUE)
> # Plot bar chart with standard deviation -----
>   #data : a data frame
>   #varname : the name of a column containing the variable to be summarized
>   #groupnames : vector of column names to be used as
>   #grouping variables
>   data_summary <- function(data, varname, groupnames){
+     require(plyr)
+     summary_func <- function(x, col){
+       c(mean = mean(x[[col]], na.rm=TRUE),
+         sd = sd(x[[col]], na.rm=TRUE))
+     }
+     data_sum<-ddply(data, groupnames, .fun=summary_func,
+                     varname)
+     data_sum <- rename(data_sum, c("mean" = varname))
+     return(data_sum)
+   }
> # Choose data file Turbidity.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater Turbidity
7  Effluent    1      PPG          S      0.11
9  Effluent    2      PPG          S      0.14
12 Effluent    3      PPG          S      0.16
29 Effluent    1      PPG          L      0.11
30 Effluent    1      PPG          L      0.57
31 Effluent    1      PPG          L      0.36
> # Define factors for metadata -----
```

```
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)
> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
      Sample Cell Membrane Wastewater Turbidity
7   Effluent    1      PPG          S      0.11
9   Effluent    2      PPG          S      0.14
12  Effluent    3      PPG          S      0.16
96  Effluent    1     PVDF          S      0.16
97  Effluent    1     PVDF          S      0.50
98  Effluent    1     PVDF          S      0.45
99  Effluent    1     PVDF          S      0.40
100 Effluent    2     PVDF          S      0.25
101 Effluent    2     PVDF          S      0.53
102 Effluent    2     PVDF          S      0.44
103 Effluent    2     PVDF          S      0.48
104 Effluent    3     PVDF          S      0.12
105 Effluent    3     PVDF          S      0.89
106 Effluent    3     PVDF          S      0.79
107 Effluent    3     PVDF          S      0.63
138 Effluent    1      PES          S      0.35
139 Effluent    1      PES          S      0.64
142 Effluent    2      PES          S      0.32
143 Effluent    2      PES          S      0.59
146 Effluent    3      PES          S      0.33
147 Effluent    3      PES          S      0.52
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
> # Statistical analysis on data1
>
> fit1 <- aov(Turbidity~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.2796 0.13982   3.674 0.0459 *
Residuals   18 0.6850 0.03805
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Turbidity ~ Membrane, data = data1)

$Membrane
                diff         lwr        upr     p adj
PPG-PES  -0.32166667 -0.67370229 0.03036896 0.0767223
PVDF-PES  0.01166667 -0.23726011 0.26059345 0.9921465
PVDF-PPG  0.33333333  0.01197024 0.65469642 0.0413495


> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=Turbidity)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("Turbidity (NTU)") + labs(title = "", subtitle=NULL) + ylim(0, 3)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
```

```
+          axis.text.y=element_text(size=20, family="Times New Roman"),
+          axis.title.y = element_text(size = 20, family="Times New Roman"),
+          axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="Turbidity",
+                              groupnames=c("Membrane"))
> box_1_data
  Membrane Turbidity         sd
1      PES 0.4583333 0.14246637
2      PPG 0.1366667 0.02516611
3     PVDF 0.4700000 0.23005928


>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater Turbidity
29  Effluent    1      PPG          L      0.11
30  Effluent    1      PPG          L      0.57
31  Effluent    1      PPG          L      0.36
32  Effluent    1      PPG          L      0.36
33  Effluent    1      PPG          L      0.15
34  Effluent    1      PPG          L      0.17
35  Effluent    2      PPG          L      0.10
36  Effluent    2      PPG          L      0.14
37  Effluent    2      PPG          L      0.27
38  Effluent    2      PPG          L      0.24
39  Effluent    2      PPG          L      0.13
40  Effluent    2      PPG          L      0.14
41  Effluent    2      PPG          L      0.10
42  Effluent    3      PPG          L      0.10
43  Effluent    3      PPG          L      0.09
44  Effluent    3      PPG          L      0.25
45  Effluent    3      PPG          L      0.19
46  Effluent    3      PPG          L      0.13
47  Effluent    3      PPG          L      0.13
57  Effluent    1     PVDF          L      0.17
58  Effluent    1     PVDF          L      1.54
59  Effluent    1     PVDF          L      2.23
60  Effluent    1     PVDF          L      0.41
61  Effluent    2     PVDF          L      0.20
62  Effluent    2     PVDF          L      0.98
63  Effluent    2     PVDF          L      1.54
64  Effluent    2     PVDF          L      0.35
65  Effluent    3     PVDF          L      0.18
66  Effluent    3     PVDF          L      1.14
67  Effluent    3     PVDF          L      1.44
68  Effluent    3     PVDF          L      0.39
158 Effluent    1      PES          L      0.15
159 Effluent    1      PES          L      1.06
160 Effluent    1      PES          L      0.47
162 Effluent    2      PES          L      0.28
163 Effluent    2      PES          L      0.96
164 Effluent    2      PES          L      0.46
166 Effluent    3      PES          L      0.25
167 Effluent    3      PES          L      1.27
168 Effluent    3      PES          L      0.76
>
```

```
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(Turbidity~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2  3.642  1.8211   9.868 0.000367 ***
Residuals   37  6.828  0.1845
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Turbidity ~ Membrane, data = data2)

$Membrane
               diff        lwr          upr     p adj
PPG-PES  -0.4325731 -0.8569730 -0.008173173 0.0449566
PVDF-PES  0.2519444 -0.2105347  0.714423543 0.3878697
PVDF-PPG  0.6845175  0.2977876  1.071247455 0.0003233

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=Turbidity)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("Turbidity (NTU)") + labs(title = "", subtitle=NULL) + ylim(0,3)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>



> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="Turbidity",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane Turbidity        sd
1      PES 0.6288889 0.3987620
2      PPG 0.1963158 0.1230271
3     PVDF 0.8808333 0.6930362


> # Select treated sample data for shower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
      Sample Cell Membrane Wastewater Turbidity
77  Effluent    1      PPG         SL      0.48
78  Effluent    1      PPG         SL      1.39
79  Effluent    1      PPG         SL      1.27
```

```
80  Effluent    2      PPG        SL      0.67
81  Effluent    2      PPG        SL      1.64
82  Effluent    2      PPG        SL      1.41
83  Effluent    2      PPG        SL      3.40
84  Effluent    3      PPG        SL      0.55
118 Effluent    1     PVDF        SL      0.16
119 Effluent    1     PVDF        SL      2.74
120 Effluent    1     PVDF        SL      2.77
121 Effluent    1     PVDF        SL      1.71
122 Effluent    2     PVDF        SL      0.15
123 Effluent    2     PVDF        SL      2.23
124 Effluent    2     PVDF        SL      1.95
125 Effluent    2     PVDF        SL      1.29
126 Effluent    3     PVDF        SL      0.24
127 Effluent    3     PVDF        SL      2.66
128 Effluent    3     PVDF        SL      2.52
129 Effluent    3     PVDF        SL      1.20
179 Effluent    1      PES        SL      0.09
180 Effluent    1      PES        SL      0.60
181 Effluent    1      PES        SL      1.30
182 Effluent    1      PES        SL      0.44
183 Effluent    2      PES        SL      0.08
184 Effluent    2      PES        SL      0.14
185 Effluent    2      PES        SL      0.25
186 Effluent    2      PES        SL      0.50
187 Effluent    3      PES        SL      0.13
188 Effluent    3      PES        SL      0.20
189 Effluent    3      PES        SL      0.33
190 Effluent    3      PES        SL      0.54
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(Turbidity~Membrane, data3)
> summary(fit3)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2  10.10   5.051   7.759  0.002 **
Residuals   29  18.88   0.651
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Turbidity ~ Membrane, data = data3)

$Membrane
              diff         lwr      upr     p adj
PPG-PES  0.9679167  0.05841571 1.877418 0.0351838
PVDF-PES 1.2516667  0.43818429 2.065149 0.0019284
PVDF-PPG 0.2837500 -0.62575095 1.193251 0.7237969

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=Turbidity)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("Turbidity (NTU)") + labs(title = "", subtitle=NULL) + ylim(0, 4)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
```

```
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile
>
```

```
> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="Turbidity",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane Turbidity        sd
1      PES 0.3833333 0.3416892
2      PPG 1.3512500 0.9378918
3     PVDF 1.6350000 1.0196746
```

TP

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - TP
> ## Wei Liao, September 27, 2023
> # Plot bar chart with standard deviation -----
>   #data : a data frame
>   #varname : the name of a column containing the variable to be summarized
>   #groupnames : vector of column names to be used as
>   #grouping variables
>   data_summary <- function(data, varname, groupnames){
+     require(plyr)
+     summary_func <- function(x, col){
+       c(mean = mean(x[[col]], na.rm=TRUE),
+         sd = sd(x[[col]], na.rm=TRUE))
+     }
+     data_sum<-ddply(data, groupnames, .fun=summary_func,
+                     varname)
+     data_sum <- rename(data_sum, c("mean" = varname))
+     return(data_sum)
+   }
> # Choose data file TP.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater    TP
7  Effluent    1      PPG          S 1.120
8  Effluent    1      PPG          S 0.802
9  Effluent    2      PPG          S 1.120
10 Effluent    2      PPG          S 0.679
12 Effluent    3      PPG          S 1.340
13 Effluent    3      PPG          S 0.720
>
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)
> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
     Sample Cell Membrane Wastewater    TP
7  Effluent    1      PPG          S 1.120
```

```
8   Effluent    1      PPG          S 0.802
9   Effluent    2      PPG          S 1.120
10  Effluent    2      PPG          S 0.679
12  Effluent    3      PPG          S 1.340
13  Effluent    3      PPG          S 0.720
96  Effluent    1     PVDF          S 0.621
97  Effluent    1     PVDF          S 0.566
98  Effluent    1     PVDF          S 0.233
99  Effluent    1     PVDF          S 0.229
100 Effluent    2     PVDF          S 0.761
101 Effluent    2     PVDF          S 0.598
102 Effluent    2     PVDF          S 0.298
103 Effluent    2     PVDF          S 0.240
104 Effluent    3     PVDF          S 0.503
105 Effluent    3     PVDF          S 0.766
106 Effluent    3     PVDF          S 0.285
107 Effluent    3     PVDF          S 0.230
138 Effluent    1      PES          S 0.277
139 Effluent    1      PES          S 0.357
140 Effluent    1      PES          S 0.167
141 Effluent    1      PES          S 0.182
143 Effluent    2      PES          S 0.366
144 Effluent    2      PES          S 0.197
145 Effluent    2      PES          S 0.183
146 Effluent    3      PES          S 0.261
147 Effluent    3      PES          S 0.253
148 Effluent    3      PES          S 0.179
149 Effluent    3      PES          S 0.196
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)

> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
      Sample Cell Membrane Wastewater    TP
7   Effluent    1      PPG          S 1.120
8   Effluent    1      PPG          S 0.802
9   Effluent    2      PPG          S 1.120
10  Effluent    2      PPG          S 0.679
12  Effluent    3      PPG          S 1.340
13  Effluent    3      PPG          S 0.720
96  Effluent    1     PVDF          S 0.621
97  Effluent    1     PVDF          S 0.566
98  Effluent    1     PVDF          S 0.233
99  Effluent    1     PVDF          S 0.229
100 Effluent    2     PVDF          S 0.761
101 Effluent    2     PVDF          S 0.598
102 Effluent    2     PVDF          S 0.298
103 Effluent    2     PVDF          S 0.240
104 Effluent    3     PVDF          S 0.503
105 Effluent    3     PVDF          S 0.766
106 Effluent    3     PVDF          S 0.285
107 Effluent    3     PVDF          S 0.230
138 Effluent    1      PES          S 0.277
139 Effluent    1      PES          S 0.357
140 Effluent    1      PES          S 0.167
141 Effluent    1      PES          S 0.182
143 Effluent    2      PES          S 0.366
144 Effluent    2      PES          S 0.197
145 Effluent    2      PES          S 0.183
146 Effluent    3      PES          S 0.261
```

```
147 Effluent    3      PES           S 0.253
148 Effluent    3      PES           S 0.179
149 Effluent    3      PES           S 0.196
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
> # Statistical analysis on data1
>
> fit1 <- aov(TP~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value  Pr(>F)
Membrane     2 2.0610   1.030   29.43 2.1e-07 ***
Residuals   26 0.9103   0.035
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TP ~ Membrane, data = data1)

$Membrane
               diff         lwr        upr     p adj
PPG-PES   0.7255000  0.48952022  0.9614798 0.0000001
PVDF-PES  0.2061667  0.01207841  0.4002549 0.0357270
PVDF-PPG -0.5193333 -0.75181692 -0.2868497 0.0000230

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=TP)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TP (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 25)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="TP",
+                            groupnames=c("Membrane"))
> box_1_data
  Membrane        TP         sd
1      PES 0.2380000 0.07137787
2      PPG 0.9635000 0.26722706
3     PVDF 0.4441667 0.21369853

> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater    TP
29  Effluent    1      PPG           L 21.80
30  Effluent    1      PPG           L 21.50
31  Effluent    1      PPG           L 16.50
```

```
32  Effluent    1       PPG          L 17.00
35  Effluent    2       PPG          L 20.80
36  Effluent    2       PPG          L 20.60
37  Effluent    2       PPG          L 20.60
38  Effluent    2       PPG          L 20.90
42  Effluent    3       PPG          L 21.40
43  Effluent    3       PPG          L 21.40
44  Effluent    3       PPG          L 21.10
45  Effluent    3       PPG          L 20.10
57  Effluent    1      PVDF          L 11.90
58  Effluent    1      PVDF          L 11.40
59  Effluent    1      PVDF          L 10.90
60  Effluent    1      PVDF          L 11.60
65  Effluent    3      PVDF          L 11.50
66  Effluent    3      PVDF          L 11.20
67  Effluent    3      PVDF          L 10.10
68  Effluent    3      PVDF          L  9.89
158 Effluent    1       PES          L  7.67
159 Effluent    1       PES          L  7.87
160 Effluent    1       PES          L 10.40
161 Effluent    1       PES          L 10.10
162 Effluent    2       PES          L  8.30
163 Effluent    2       PES          L  8.14
164 Effluent    2       PES          L 10.40
165 Effluent    2       PES          L 10.30
166 Effluent    3       PES          L  8.95
167 Effluent    3       PES          L  8.90
168 Effluent    3       PES          L 11.40
169 Effluent    3       PES          L 11.60
> 
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
> 
> # Statistical analysis on data2
> 
> fit2 <- aov(TP~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)    
Membrane     2  789.3   394.6   200.3 <2e-16 ***
Residuals   29   57.1     2.0                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TP ~ Membrane, data = data2)

$Membrane
              diff          lwr       upr     p adj
PPG-PES  10.805833   9.39047973 12.221187 0.0000000
PVDF-PES  1.558750  -0.02366343  3.141163 0.0541612
PVDF-PPG -9.247083 -10.82949677 -7.664670 0.0000000


> 
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=TP)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TP (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0,25)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
```

```
+          axis.text.x = element_text(size=20, family="Times New Roman"),
+          axis.text.y=element_text(size=20, family="Times New Roman"),
+          axis.title.y = element_text(size = 20, family="Times New Roman"),
+          axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="TP",
+                             groupnames=c("Membrane"))
> box_2_data
  Membrane       TP        sd
1      PES  9.50250 1.368404
2      PPG 20.30833 1.729665
3     PVDF 11.06125 0.721317


> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
      Sample Cell Membrane Wastewater     TP
77  Effluent    1      PPG         SL 11.100
78  Effluent    1      PPG         SL 10.700
80  Effluent    2      PPG         SL 11.300
81  Effluent    2      PPG         SL 11.000
82  Effluent    2      PPG         SL 10.900
83  Effluent    2      PPG         SL 11.000
84  Effluent    3      PPG         SL 10.700
85  Effluent    3      PPG         SL 11.100
118 Effluent    1     PVDF         SL  6.560
119 Effluent    1     PVDF         SL  6.260
120 Effluent    1     PVDF         SL  6.810
121 Effluent    1     PVDF         SL  6.670
122 Effluent    2     PVDF         SL  6.540
123 Effluent    2     PVDF         SL  6.440
124 Effluent    2     PVDF         SL  6.650
125 Effluent    2     PVDF         SL  6.570
126 Effluent    3     PVDF         SL  6.740
127 Effluent    3     PVDF         SL  6.890
128 Effluent    3     PVDF         SL  7.130
129 Effluent    3     PVDF         SL  6.600
179 Effluent    1      PES         SL  5.650
180 Effluent    1      PES         SL  5.440
181 Effluent    1      PES         SL  6.515
183 Effluent    2      PES         SL  6.580
184 Effluent    2      PES         SL  6.450
185 Effluent    2      PES         SL  6.515
187 Effluent    3      PES         SL 12.900
188 Effluent    3      PES         SL  6.280
189 Effluent    3      PES         SL  6.140
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(TP~Membrane, data3)
> summary(fit3)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
Membrane     2 102.48   51.24   31.66 1.08e-07 ***
Residuals   26  42.08    1.62                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TP ~ Membrane, data = data3)

$Membrane
               diff       lwr       upr     p adj
PPG-PES   4.0338889  2.497718  5.570060 0.0000019
PVDF-PES -0.2861111 -1.680164  1.107942 0.8671630
PVDF-PPG -4.3200000 -5.762982 -2.877018 0.0000002

> 
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=TP)) + 
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TP (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 25)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile
> 

> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="TP", 
+                              groupnames=c("Membrane"))
> box_3_data
  Membrane        TP        sd
1      PES  6.941111 2.2705241
2      PPG 10.975000 0.2052873
3     PVDF  6.655000 0.2230165
```

TN

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - TN
> ## Wei Liao, September 27, 2023

> # Choose data file TN.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater   TN
7  Effluent    1      PPG          S 4.01
9  Effluent    2      PPG          S 3.40
12 Effluent    3      PPG          S 2.12
13 Effluent    3      PPG          S 2.06
```

```
14 Effluent    3      PPG          S 2.81
29 Effluent    1      PPG          L 4.20

> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)
>
> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
      Sample Cell Membrane Wastewater    TN
7   Effluent    1      PPG          S 4.010
9   Effluent    2      PPG          S 3.400
12  Effluent    3      PPG          S 2.120
13  Effluent    3      PPG          S 2.060
14  Effluent    3      PPG          S 2.810
96  Effluent    1     PVDF          S 2.510
97  Effluent    1     PVDF          S 2.570
98  Effluent    1     PVDF          S 1.280
99  Effluent    1     PVDF          S 1.420
100 Effluent    2     PVDF          S 2.320
101 Effluent    2     PVDF          S 2.370
102 Effluent    2     PVDF          S 1.190
103 Effluent    2     PVDF          S 1.360
104 Effluent    3     PVDF          S 2.020
105 Effluent    3     PVDF          S 2.340
106 Effluent    3     PVDF          S 1.260
107 Effluent    3     PVDF          S 1.220
138 Effluent    1      PES          S 2.100
139 Effluent    1      PES          S 2.440
140 Effluent    1      PES          S 0.121
141 Effluent    1      PES          S 0.210
142 Effluent    2      PES          S 2.140
143 Effluent    2      PES          S 1.650
144 Effluent    2      PES          S 0.199
145 Effluent    2      PES          S 0.464
146 Effluent    3      PES          S 1.610
147 Effluent    3      PES          S 1.760
148 Effluent    3      PES          S 0.517
149 Effluent    3      PES          S 0.262
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)

> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
>
> # Statistical analysis on data1
>
> fit1 <- aov(TN~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2  11.13   5.566   9.449 0.000824 ***
Residuals   26  15.32   0.589
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

```
Fit: aov(formula = TN ~ Membrane, data = data1)

$Membrane
               diff         lwr        upr     p adj
PPG-PES   1.7572500  0.74208656  2.7724134 0.0006027
PVDF-PES  0.6989167 -0.07967815  1.4775115 0.0846698
PVDF-PPG -1.0583333 -2.07349677 -0.0431699 0.0397869

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=TN)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TN (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 5)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="TN",
+                              groupnames=c("Membrane"))
> box_1_data
  Membrane       TN        sd
1      PES 1.122750 0.8983905
2      PPG 2.880000 0.8369886
3     PVDF 1.821667 0.5748649


> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater    TN
29  Effluent    1      PPG          L 4.200
30  Effluent    1      PPG          L 4.840
31  Effluent    1      PPG          L 1.060
32  Effluent    1      PPG          L 1.140
35  Effluent    2      PPG          L 2.890
36  Effluent    2      PPG          L 3.090
37  Effluent    2      PPG          L 3.470
38  Effluent    2      PPG          L 1.640
42  Effluent    3      PPG          L 3.210
43  Effluent    3      PPG          L 3.570
44  Effluent    3      PPG          L 3.330
45  Effluent    3      PPG          L 3.370
57  Effluent    1     PVDF          L 2.490
58  Effluent    1     PVDF          L 2.570
59  Effluent    1     PVDF          L 0.964
60  Effluent    1     PVDF          L 1.010
65  Effluent    3     PVDF          L 2.830
66  Effluent    3     PVDF          L 2.750
67  Effluent    3     PVDF          L 0.860
68  Effluent    3     PVDF          L 0.917
158 Effluent    1      PES          L 4.250
```

```
159 Effluent    1       PES           L 1.970
160 Effluent    1       PES           L 0.554
161 Effluent    1       PES           L 0.432
162 Effluent    2       PES           L 2.050
163 Effluent    2       PES           L 2.110
164 Effluent    2       PES           L 0.590
165 Effluent    2       PES           L 0.555
166 Effluent    3       PES           L 2.330
167 Effluent    3       PES           L 2.100
168 Effluent    3       PES           L 0.596
169 Effluent    3       PES           L 1.850
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(TN~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2  12.74   6.372   5.308 0.0109 *
Residuals   29  34.81   1.200
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TN ~ Membrane, data = data2)

$Membrane
               diff        lwr       upr     p adj
PPG-PES   1.3685833  0.2639182 2.4732485 0.0127579
PVDF-PES  0.1832917 -1.0517615 1.4183448 0.9288099
PVDF-PPG -1.1852917 -2.4203448 0.0497615 0.0619115


>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=TN)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TN (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0,5)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="TN",
+                             groupnames=c("Membrane"))
> box_2_data
  Membrane       TN        sd
1      PES 1.615583 1.1298820
2      PPG 2.984167 1.1580270
3     PVDF 1.798875 0.9272598
```

```
> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
       Sample Cell Membrane Wastewater    TN
77   Effluent    1      PPG         SL 4.030
78   Effluent    1      PPG         SL 4.150
80   Effluent    2      PPG         SL 4.390
81   Effluent    2      PPG         SL 4.390
82   Effluent    2      PPG         SL 1.230
83   Effluent    2      PPG         SL 1.300
84   Effluent    3      PPG         SL 4.580
85   Effluent    3      PPG         SL 3.420
118 Effluent    1     PVDF         SL 2.380
119 Effluent    1     PVDF         SL 2.110
120 Effluent    1     PVDF         SL 1.030
121 Effluent    1     PVDF         SL 1.170
122 Effluent    2     PVDF         SL 2.260
123 Effluent    2     PVDF         SL 2.050
124 Effluent    2     PVDF         SL 1.250
125 Effluent    2     PVDF         SL 1.140
126 Effluent    3     PVDF         SL 2.430
127 Effluent    3     PVDF         SL 1.920
128 Effluent    3     PVDF         SL 1.260
129 Effluent    3     PVDF         SL 1.660
179 Effluent    1      PES         SL 3.080
180 Effluent    1      PES         SL 2.005
181 Effluent    1      PES         SL 2.025
183 Effluent    2      PES         SL 3.645
184 Effluent    2      PES         SL 1.710
185 Effluent    2      PES         SL 2.025
187 Effluent    3      PES         SL 4.630
188 Effluent    3      PES         SL 2.430
189 Effluent    3      PES         SL 2.650
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(TN~Membrane, data3)
> summary(fit3)
            Df Sum Sq Mean Sq F value  Pr(>F)   
Membrane     2  14.60   7.301   8.005 0.00196 **
Residuals   26  23.71   0.912                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TN ~ Membrane, data = data3)

$Membrane
               diff       lwr         upr     p adj
PPG-PES   0.7473611 -0.405806  1.90052822 0.2592510
PVDF-PES -0.9672222 -2.013704  0.07925988 0.0740837
PVDF-PPG -1.7145833 -2.797795 -0.63137145 0.0015608
```

```
>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=TN)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("TN (mg/L)") + labs(title = "", subtitle=NULL) + ylim(0, 5)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile
>
```

```
> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="TN",
+                             groupnames=c("Membrane"))
> box_3_data
  Membrane       TN        sd
1      PES 2.688889 0.9494070
2      PPG 3.436250 1.3846499
3     PVDF 1.721667 0.5294394
```

Conductivity

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - Conductivity
> ## Wei Liao, September 27, 2023
> # Plot bar chart with standard deviation -----
>   #data : a data frame
>   #varname : the name of a column containing the variable to be summarized
>   #groupnames : vector of column names to be used as
>   #grouping variables
>   data_summary <- function(data, varname, groupnames){
+     require(plyr)
+     summary_func <- function(x, col){
+       c(mean = mean(x[[col]], na.rm=TRUE),
+         sd = sd(x[[col]], na.rm=TRUE))
+     }
+     data_sum<-ddply(data, groupnames, .fun=summary_func,
+                     varname)
+     data_sum <- rename(data_sum, c("mean" = varname))
+     return(data_sum)
+   }
> # Choose data file Conductivity.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater Conductivity
7  Effluent    1      PPG          S          330
9  Effluent    2      PPG          S          329
10 Effluent    2      PPG          S          316
11 Effluent    2      PPG          S          321
12 Effluent    3      PPG          S          336
13 Effluent    3      PPG          S          320
>
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
```

```
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)


> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
       Sample Cell Membrane Wastewater Conductivity
7   Effluent    1      PPG          S          330
9   Effluent    2      PPG          S          329
10  Effluent    2      PPG          S          316
11  Effluent    2      PPG          S          321
12  Effluent    3      PPG          S          336
13  Effluent    3      PPG          S          320
14  Effluent    3      PPG          S          326
96  Effluent    1     PVDF          S          310
97  Effluent    1     PVDF          S          327
98  Effluent    1     PVDF          S          327
99  Effluent    1     PVDF          S          329
100 Effluent    2     PVDF          S          310
101 Effluent    2     PVDF          S          325
102 Effluent    2     PVDF          S          326
103 Effluent    2     PVDF          S          325
104 Effluent    3     PVDF          S          278
105 Effluent    3     PVDF          S          312
106 Effluent    3     PVDF          S          312
107 Effluent    3     PVDF          S          311
138 Effluent    1      PES          S          270
139 Effluent    1      PES          S          255
142 Effluent    2      PES          S          279
143 Effluent    2      PES          S          268
146 Effluent    3      PES          S          276
147 Effluent    3      PES          S          262
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
>
> # Statistical analysis on data1
>
> fit1 <- aov(Conductivity~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2  12319    6159   46.36 1.29e-08 ***
Residuals   22   2923     133
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Conductivity ~ Membrane, data = data1)

$Membrane
              diff       lwr       upr     p adj
PPG-PES  57.095238  40.98566 73.204817 0.0000000
PVDF-PES 47.666667  33.18871 62.144620 0.0000001
PVDF-PPG -9.428571 -23.19985  4.342709 0.2203367


>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=Conductivity)) +
```

```
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("Conductivity") + labs(title = "", subtitle=NULL) + ylim(0, 500)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="Conductivity",
+                             groupnames=c("Membrane"))
> box_1_data
  Membrane Conductivity        sd
1      PES     268.3333  8.869423
2      PPG     325.4286  6.876461
3     PVDF     316.0000 14.289220

> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater Conductivity
29  Effluent    1      PPG          L        371.0
30  Effluent    1      PPG          L        366.0
31  Effluent    1      PPG          L        346.0
32  Effluent    1      PPG          L        346.0
33  Effluent    1      PPG          L        354.0
34  Effluent    1      PPG          L        355.0
35  Effluent    2      PPG          L        377.0
36  Effluent    2      PPG          L        362.0
37  Effluent    2      PPG          L        342.0
38  Effluent    2      PPG          L        344.0
39  Effluent    2      PPG          L        351.0
40  Effluent    2      PPG          L        354.0
41  Effluent    2      PPG          L        359.0
42  Effluent    3      PPG          L        365.0
43  Effluent    3      PPG          L        377.0
44  Effluent    3      PPG          L        346.0
45  Effluent    3      PPG          L        346.0
46  Effluent    3      PPG          L        357.0
47  Effluent    3      PPG          L        358.0
57  Effluent    1     PVDF          L        457.0
58  Effluent    1     PVDF          L        449.0
59  Effluent    1     PVDF          L        448.0
60  Effluent    1     PVDF          L        461.0
61  Effluent    2     PVDF          L        457.0
62  Effluent    2     PVDF          L        452.0
63  Effluent    2     PVDF          L        452.0
64  Effluent    2     PVDF          L        463.0
65  Effluent    3     PVDF          L        454.0
66  Effluent    3     PVDF          L        436.0
67  Effluent    3     PVDF          L        436.0
68  Effluent    3     PVDF          L        453.0
158 Effluent    1      PES          L        209.8
159 Effluent    1      PES          L        266.0
160 Effluent    1      PES          L        279.0
162 Effluent    2      PES          L        219.3
```

```
163 Effluent    2       PES          L        266.0
164 Effluent    2       PES          L        279.0
166 Effluent    3       PES          L        223.0
167 Effluent    3       PES          L        285.0
168 Effluent    3       PES          L        304.0
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(Conductivity~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 192080   96040   301.5 <2e-16 ***
Residuals   37  11787     319
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Conductivity ~ Membrane, data = data2)

$Membrane
              diff       lwr      upr p adj
PPG-PES   97.62047  79.98697 115.2540     0
PVDF-PES 192.48889 173.27322 211.7046     0
PVDF-PPG  94.86842  78.80008 110.9368     0


>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=Conductivity)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("Conductivity") + labs(title = "", subtitle=NULL) + ylim(0,500)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile
>

> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="Conductivity",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane Conductivity        sd
1      PES     259.0111 33.338658
2      PPG     356.6316 10.812598
3     PVDF     451.5000  8.479923

> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
      Sample Cell Membrane Wastewater Conductivity
```

```
77  Effluent    1      PPG         SL          414
78  Effluent    1      PPG         SL          413
79  Effluent    1      PPG         SL          414
80  Effluent    2      PPG         SL          417
81  Effluent    2      PPG         SL          412
82  Effluent    2      PPG         SL          415
83  Effluent    2      PPG         SL          456
84  Effluent    3      PPG         SL          407
118 Effluent    1     PVDF         SL          263
119 Effluent    1     PVDF         SL          280
120 Effluent    1     PVDF         SL          280
121 Effluent    1     PVDF         SL          294
122 Effluent    2     PVDF         SL          266
123 Effluent    2     PVDF         SL          279
124 Effluent    2     PVDF         SL          281
125 Effluent    2     PVDF         SL          298
126 Effluent    3     PVDF         SL          251
127 Effluent    3     PVDF         SL          277
128 Effluent    3     PVDF         SL          280
129 Effluent    3     PVDF         SL          286
179 Effluent    1      PES         SL          247
180 Effluent    1      PES         SL          261
181 Effluent    1      PES         SL          264
182 Effluent    1      PES         SL          274
183 Effluent    2      PES         SL          277
184 Effluent    2      PES         SL          283
185 Effluent    2      PES         SL          292
186 Effluent    2      PES         SL          303
187 Effluent    3      PES         SL          235
188 Effluent    3      PES         SL          272
189 Effluent    3      PES         SL          278
190 Effluent    3      PES         SL          289
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(Conductivity~Membrane, data3)
> summary(fit3)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 122987   61494   237.4 <2e-16 ***
Residuals   29   7512     259
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Conductivity ~ Membrane, data = data3)

$Membrane
              diff        lwr        upr     p adj
PPG-PES   145.5833  127.44120  163.72546 0.0000000
PVDF-PES    5.0000  -11.22682   21.22682 0.7294925
PVDF-PPG -140.5833 -158.72546 -122.44120 0.0000000



>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=Conductivity)) +
+   geom_violin(trim=TRUE, fill="green") +
```

```
+   xlab("Membrane")+
+   ylab("Conductivity") + labs(title = "", subtitle=NULL) + ylim(0, 500)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile
>

> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="Conductivity",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane Conductivity       sd
1      PES     272.9167 19.08097
2      PPG     418.5000 15.42725
3     PVDF     277.9167 12.93662
```

pH

```
> ## Statistical analysis
> ## Flat cell analysis
> ## Water quality data - pH
> ## Wei Liao, September 27, 2023
>
> # Load libraries -----

>   loadfonts(device="win", quiet=TRUE)
> # Plot bar chart with standard deviation -----
>   #data : a data frame
>   #varname : the name of a column containing the variable to be summarized
>   #groupnames : vector of column names to be used as
>   #grouping variables
>   data_summary <- function(data, varname, groupnames){
+     require(plyr)
+     summary_func <- function(x, col){
+       c(mean = mean(x[[col]], na.rm=TRUE),
+         sd = sd(x[[col]], na.rm=TRUE))
+     }
+     data_sum<-ddply(data, groupnames, .fun=summary_func,
+                     varname)
+     data_sum <- rename(data_sum, c("mean" = varname))
+     return(data_sum)
+   }
> # Choose data file pH.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
     Sample Cell Membrane Wastewater   pH
7  Effluent    1      PPG          S 6.68
9  Effluent    2      PPG          S 6.44
10 Effluent    2      PPG          S 7.04
11 Effluent    2      PPG          S 7.27
12 Effluent    3      PPG          S 6.50
13 Effluent    3      PPG          S 7.10
>
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Cell <- factor(metadata$Cell)
> metadata$Wastewater <- factor(metadata$Wastewater)
```

```
> # Select treated sample data for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
      Sample Cell Membrane Wastewater   pH
7   Effluent    1      PPG          S 6.68
9   Effluent    2      PPG          S 6.44
10  Effluent    2      PPG          S 7.04
11  Effluent    2      PPG          S 7.27
12  Effluent    3      PPG          S 6.50
13  Effluent    3      PPG          S 7.10
14  Effluent    3      PPG          S 7.31
96  Effluent    1     PVDF          S 7.68
97  Effluent    1     PVDF          S 6.72
98  Effluent    1     PVDF          S 6.74
99  Effluent    1     PVDF          S 6.66
100 Effluent    2     PVDF          S 7.62
101 Effluent    2     PVDF          S 6.78
102 Effluent    2     PVDF          S 6.78
103 Effluent    2     PVDF          S 6.71
104 Effluent    3     PVDF          S 7.64
105 Effluent    3     PVDF          S 6.89
106 Effluent    3     PVDF          S 6.84
107 Effluent    3     PVDF          S 6.81
138 Effluent    1      PES          S 6.64
139 Effluent    1      PES          S 6.92
142 Effluent    2      PES          S 6.74
143 Effluent    2      PES          S 7.06
146 Effluent    3      PES          S 6.78
147 Effluent    3      PES          S 7.08
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Cell <- factor(data1$Cell)
>
> # Statistical analysis on data1
>
> fit1 <- aov(pH~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.0664 0.03322   0.269  0.767
Residuals   22 2.7163 0.12347
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pH ~ Membrane, data = data1)

$Membrane
               diff        lwr       upr     p adj
PPG-PES  0.03571429 -0.4553656 0.5267942 0.9817833
PVDF-PES 0.11916667 -0.3221752 0.5605085 0.7784019
PVDF-PPG 0.08345238 -0.3363475 0.5032522 0.8723548


> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=pH)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("pH") + labs(title = "", subtitle=NULL) + ylim(6, 8)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
```

```
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
> box_1 + geom_boxplot(width=0.1) # Add median and quartile

> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="pH",
+                              groupnames=c("Membrane"))
> box_1_data
  Membrane       pH        sd
1      PES 6.870000 0.1792205
2      PPG 6.905714 0.3615641
3     PVDF 6.989167 0.4012811


> # Select treated sample data for laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
      Sample Cell Membrane Wastewater   pH
29  Effluent    1      PPG          L 6.88
30  Effluent    1      PPG          L 6.82
31  Effluent    1      PPG          L 6.87
32  Effluent    1      PPG          L 7.35
33  Effluent    1      PPG          L 6.89
34  Effluent    1      PPG          L 7.31
35  Effluent    2      PPG          L 7.05
36  Effluent    2      PPG          L 6.90
37  Effluent    2      PPG          L 6.95
38  Effluent    2      PPG          L 7.37
39  Effluent    2      PPG          L 7.10
40  Effluent    2      PPG          L 7.31
41  Effluent    2      PPG          L 7.15
42  Effluent    3      PPG          L 7.27
43  Effluent    3      PPG          L 6.99
44  Effluent    3      PPG          L 7.05
45  Effluent    3      PPG          L 7.37
46  Effluent    3      PPG          L 7.21
47  Effluent    3      PPG          L 7.32
57  Effluent    1     PVDF          L 7.56
59  Effluent    1     PVDF          L 7.26
60  Effluent    1     PVDF          L 7.09
61  Effluent    2     PVDF          L 7.66
62  Effluent    2     PVDF          L 7.06
63  Effluent    2     PVDF          L 7.25
64  Effluent    2     PVDF          L 7.17
65  Effluent    3     PVDF          L 7.73
66  Effluent    3     PVDF          L 7.13
67  Effluent    3     PVDF          L 7.28
68  Effluent    3     PVDF          L 7.24
158 Effluent    1      PES          L 7.20
159 Effluent    1      PES          L 7.36
160 Effluent    1      PES          L 6.83
162 Effluent    2      PES          L 7.22
163 Effluent    2      PES          L 6.54
164 Effluent    2      PES          L 6.91
166 Effluent    3      PES          L 7.27
167 Effluent    3      PES          L 6.58
168 Effluent    3      PES          L 6.98
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
```

```
> data2$Cell <- factor(data2$Cell)
>
> # Statistical analysis on data2
>
> fit2 <- aov(pH~Membrane, data2)
> summary(fit2)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.5464 0.27321   5.072 0.0115 *
Residuals   36 1.9394 0.05387
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pH ~ Membrane, data = data2)

$Membrane
              diff         lwr       upr     p adj
PPG-PES  0.1259064 -0.10366247 0.3554753 0.3825054
PVDF-PES 0.3240404  0.06904667 0.5790341 0.0100610
PVDF-PPG 0.1981340 -0.01680719 0.4130751 0.0758305

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=pH)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("pH") + labs(title = "", subtitle=NULL) + ylim(6, 8)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
> box_2 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="pH",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane       pH        sd
1      PES 6.987778 0.2989472
2      PPG 7.113684 0.1955394
3     PVDF 7.311818 0.2315521


> # Select treated sample data for slower and laundry combined wastewater---
-
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
     Sample Cell Membrane Wastewater   pH
77  Effluent    1      PPG         SL 7.32
78  Effluent    1      PPG         SL 7.03
79  Effluent    1      PPG         SL 6.66
80  Effluent    2      PPG         SL 7.34
81  Effluent    2      PPG         SL 7.07
82  Effluent    2      PPG         SL 6.78
83  Effluent    2      PPG         SL 6.84
```

```
84  Effluent    3      PPG         SL 7.32
118 Effluent    1     PVDF         SL 6.90
119 Effluent    1     PVDF         SL 6.63
120 Effluent    1     PVDF         SL 6.76
121 Effluent    1     PVDF         SL 6.77
122 Effluent    2     PVDF         SL 6.96
123 Effluent    2     PVDF         SL 6.64
124 Effluent    2     PVDF         SL 6.83
125 Effluent    2     PVDF         SL 6.53
126 Effluent    3     PVDF         SL 7.07
127 Effluent    3     PVDF         SL 6.69
128 Effluent    3     PVDF         SL 6.84
129 Effluent    3     PVDF         SL 6.52
179 Effluent    1      PES         SL 6.96
180 Effluent    1      PES         SL 7.57
181 Effluent    1      PES         SL 7.22
182 Effluent    1      PES         SL 7.36
183 Effluent    2      PES         SL 7.08
184 Effluent    2      PES         SL 7.06
185 Effluent    2      PES         SL 7.20
186 Effluent    2      PES         SL 7.40
187 Effluent    3      PES         SL 7.18
188 Effluent    3      PES         SL 6.86
189 Effluent    3      PES         SL 7.25
190 Effluent    3      PES         SL 7.39
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Cell <- factor(data3$Cell)
>
> # Statistical analysis on data2
>
> fit3 <- aov(pH~Membrane, data3)
> summary(fit3)
            Df Sum Sq Mean Sq F value   Pr(>F)    
Membrane     2  1.231  0.6156   14.16 5.13e-05 ***
Residuals   29  1.261  0.0435                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pH ~ Membrane, data = data3)

$Membrane
               diff        lwr         upr     p adj
PPG-PES  -0.1658333 -0.4008956  0.06922894 0.2070057
PVDF-PES -0.4491667 -0.6594128 -0.23892058 0.0000343
PVDF-PPG -0.2833333 -0.5183956 -0.04827106 0.0155957

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=pH)) +
+   geom_violin(trim=TRUE, fill="green") +
+   xlab("Membrane")+
+   ylab("pH") + labs(title = "", subtitle=NULL) + ylim(6, 8)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
```

```
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
> box_3 + geom_boxplot(width=0.1) # Add median and quartile


> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="pH",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane       pH        sd
1      PES 7.210833 0.2017405
2      PPG 7.045000 0.2671543
3     PVDF 6.761667 0.1688912
```

FLUX

```
> # Choose data file Metadata-Flux(r2).txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
  Replicate Membrane Wastewater Startup_time       Flux Water_treated
1         1      PES          S          183 0.16159797      431.7898
2         2      PES          S          167 0.13918776      374.1367
3         3      PES          S          192 0.09343019      248.8046
4         1      PES          L          220 0.22104614      613.4030
5         2      PES          L          212 0.23009171      640.3452
6         3      PES          L          214 0.18323401      509.5738
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Wastewater <- factor(metadata$Wastewater)
> ## Individual wastewater --------------------------------------------------
---
>
> # Shower water -----
>
> # Select shower water data
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1$Membrane<-factor(data1$Membrane)
>
> # Startup time
>
> fit3 <- aov(Startup_time~Membrane, data1)
> summary(fit3)
            Df Sum Sq Mean Sq F value   Pr(>F)    
Membrane     2   9094    4547    66.6 0.000249 ***
Residuals    5    341      68                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Startup_time ~ Membrane, data = data1)

$Membrane
             diff       lwr      upr     p adj
PPG-PES  44.33333 19.790798 68.87587 0.0046975
PVDF-PES 77.66667 55.715155 99.61818 0.0002042
PVDF-PPG 33.33333  8.790798 57.87587 0.0157307

> # Flux
>
```

```
> fit31 <- aov(Flux~Membrane, data1)
> summary(fit31)
            Df  Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.11535 0.05767   6.673 0.0388 *
Residuals    5 0.04321 0.00864
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey31 <- TukeyHSD(fit31, conf.level=0.95) #Tukey multiple comparison
> Tukey31 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Flux ~ Membrane, data = data1)

$Membrane
               diff         lwr       upr     p adj
PPG-PES   0.2962222  0.02008550 0.5723589 0.0389546
PVDF-PES  0.1916956 -0.05528861 0.4386797 0.1125146
PVDF-PPG -0.1045266 -0.38066334 0.1716101 0.4870118

> box_6 <- ggplot(data1, aes(x=Membrane, y=Flux)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Flow rate (m3 wastewater/m2 membrane/min)") + labs(title = "",
subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_6
>

> ## Mean and standard deviation
> box_6_data <- data_summary(data1, varname="Flux",
+                            groupnames=c("Membrane"))
> box_6_data
  Membrane      Flux          sd
1      PES 0.1314053 0.034743868
2      PPG 0.4276275 0.006050448
3     PVDF 0.3231009 0.142757662


> # Laundry wastewater -----
>
> # Select data
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2$Membrane<-factor(data2$Membrane)
> # Flux
>
> fit41 <- aov(Flux~Membrane, data2)
> summary(fit41)
            Df  Sum Sq Mean Sq F value   Pr(>F)
Membrane     2 0.07133 0.03567   35.28 0.000481 ***
Residuals    6 0.00607 0.00101
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey41 <- TukeyHSD(fit41, conf.level=0.95) #Tukey multiple comparison
> Tukey41 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

```
Fit: aov(formula = Flux ~ Membrane, data = data2)

$Membrane
               diff         lwr          upr     p adj
PPG-PES   0.2162245  0.13656577  0.295883323 0.0003990
PVDF-PES  0.1326548  0.05299607  0.212313622 0.0052817
PVDF-PPG -0.0835697 -0.16322848 -0.003910925 0.0415873

> # Plot - Flux
>
> box_8 <- ggplot(data2, aes(x=Membrane, y=Flux)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Flux (m3 wastewater/m2 membrane/min)") + labs(title = "",
subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_8
>

> ## Mean and standard deviation
> box_8_data <- data_summary(data2, varname="Flux",
+                            groupnames=c("Membrane"))
> box_8_data
  Membrane      Flux          sd
1      PES 0.2114573 0.024857006
2      PPG 0.4276818 0.009166393
3     PVDF 0.3441121 0.048282905


# Select data
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3$Membrane<-factor(data3$Membrane)
> # Flux
>
> fit51 <- aov(Flux~Membrane, data3)
> summary(fit51)
            Df  Sum Sq  Mean Sq F value Pr(>F)
Membrane     2 0.02386 0.011928   5.452 0.0554 .
Residuals    5 0.01094 0.002188
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey51 <- TukeyHSD(fit51, conf.level=0.95) #Tukey multiple comparison
> Tukey51 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Flux ~ Membrane, data = data3)

$Membrane
                diff          lwr        upr     p adj
PPG-PES   0.13693316 -0.002009642 0.27587596 0.0525908
PVDF-PES  0.08165465 -0.042619572 0.20592887 0.1764276
PVDF-PPG -0.05527851 -0.194221313 0.08366429 0.4566203

> # Plot - Flux
>
> box_10 <- ggplot(data3, aes(x=Membrane, y=Flux)) +
+   geom_boxplot(fill="green") +
```

```
+   xlab("Membrane")+
+   ylab("Flux (m3 wastewater/m2 membrane/min)") + labs(title = "",
subtitle=NULL) +
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_10
>


> ## Mean and standard deviation
> box_10_data <- data_summary(data2, varname="Flux",
+                              groupnames=c("Membrane"))
> box_10_data
  Membrane      Flux          sd
1      PES 0.2114573 0.024857006
2      PPG 0.4276818 0.009166393
3     PVDF 0.3441121 0.048282905
```

SEM

```
C

> # Carbon ------------------------------
>
> # Select treated sample data with controls for shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Stub Wastewater         C         O          N           P
2       PES   Al          S 1.8305411 0.9194594 0.09185618 0.017952348
5       PES   Si          S 1.8769180 0.8611143 0.11848550 0.017054731
6       PPG   Al          S 0.1601523 0.1130765 0.01826353 0.004282484
9       PPG   Si          S 0.1653794 0.1058340 0.01716142 0.003999084
12     PVDF   Al          S 0.7704856 0.4361732 0.05031181 0.006323343
16     PVDF   Si          S 0.7707605 0.4233890 0.05113660 0.010172334
            Ca           S
2  0.021542818 0.025731699
5  0.018849965 0.022440435
6  0.006108837 0.001416998
9  0.005605015 0.001354021
12 0.005361095 0.024193659
16 0.005910951 0.024056195
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
>
> fit1 <- aov(C~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2 2.9346  1.4673    4042 7.15e-06 ***
```

```
Residuals    3 0.0011  0.0004
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = C ~ Membrane, data = data1)

$Membrane
               diff        lwr        upr    p adj
PPG-PES  -1.6909637 -1.7705837 -1.6113436 0.00e+00
PVDF-PES -1.0831065 -1.1627266 -1.0034865 0.00e+00
PVDF-PPG  0.6078571  0.5282371  0.6874772 5.94e-05

> box_1 <- ggplot(data1, aes(x=Membrane, y=C)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Carbon (g/m2 membrane/100 m3 treated water)") + labs(title = "",
subtitle=NULL) + ylim(0, 2)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
>

> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="C",
+                            groupnames=c("Membrane"))
> box_1_data
  Membrane         C           sd
1      PES 1.8537295 0.0327934198
2      PPG 0.1627659 0.0036961525
3     PVDF 0.7706230 0.0001944035


> # Laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Stub
Wastewater         C         O          N           P          Ca
1       PES   Al          L 0.5616050 0.3080315 0.06421472 0.023296502
0.019712425
4       PES   Si          L 0.5734524 0.2953877 0.06839614 0.022002252
0.018019944
7       PPG   Al          L 0.3273256 0.2006818 0.03916217 0.008442923
0.008767651
11     PVDF   Al          L 0.8543386 0.3780259 0.07601758 0.015808357
0.023643803
15     PVDF   Si          L 0.8581876 0.3667539 0.08935158 0.017732852
0.022269163
             S
1  0.010055328
4  0.009358424
7  0.003052441
11 0.016083285
15 0.015533429
>
```

```
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data1
>
> fit2 <- aov(C~Membrane, data2)
> summary(fit2)
            Df  Sum Sq Mean Sq F value   Pr(>F)    
Membrane     2 0.20168 0.10084    2599 0.000385 ***
Residuals    2 0.00008 0.00004                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = C ~ Membrane, data = data2)

$Membrane
               diff        lwr        upr    p adj
PPG-PES  -0.2402030 -0.2851393 -0.1952668 7.68e-04
PVDF-PES  0.2887344  0.2520441  0.3254247 2.88e-05
PVDF-PPG  0.5289374  0.4840012  0.5738737 0.00e+00

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=C)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Carbon (g/m2 membrane/100 m3 treated water)") + labs(title = "",
subtitle=NULL) + ylim(0,2)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
>


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="C",
+                              groupnames=c("Membrane"))
> box_2_data
  Membrane         C          sd
1      PES 0.5675287 0.008377353
2      PPG 0.3273256          NA
3     PVDF 0.8562631 0.002721648


#  SL wastewater----
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
   Membrane Stub
Wastewater         C         O          N          P         Ca
3       PES   AL          SL 0.3782840 0.1837664 0.00000000 0.00000000
0.02363830
8       PPG   Al          SL 0.5836795 0.4122643 0.06376837 0.01844340
0.01735850
```

```
10      PPG   Si          SL 0.5994709 0.4059960 0.07220653 0.01976940
0.01687632
13     PVDF   Al          SL 0.9135652 0.5688396 0.09469366 0.03150809
0.01880321
14     PVDF   Si          SL 0.9582863 0.5325883 0.07046970 0.02964471
0.01846442
             S
3  0.023851900
8  0.008920339
10 0.006509436
13 0.015754045
14 0.015754045
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
> # Statistical analysis on data1
>
> fit3 <- aov(C~Membrane, data3)
> summary(fit3)
            Df  Sum Sq Mean Sq F value  Pr(>F)
Membrane     2 0.23744 0.11872   211.1 0.00471 **
Residuals    2 0.00112 0.00056
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = C ~ Membrane, data = data3)

$Membrane
              diff        lwr       upr     p adj
PPG-PES  0.2132912 0.04220495 0.3843774 0.0327580
PVDF-PES 0.5576418 0.38655554 0.7287280 0.0047397
PVDF-PPG 0.3443506 0.20465928 0.4840419 0.0085782

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=C)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Carbon (g/m2 membrane/100 m3 treated water)") + labs(title = "",
subtitle=NULL) + ylim(0,2)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
>


> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="C",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane         C         sd
1      PES 0.3782840         NA
2      PPG 0.5915752 0.01116621
```

```
3      PVDF 0.9359258 0.03162263
```

O

```
> # Oxygen -----------------------------
>
> # Shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Stub
Wastewater         C          O          N           P          Ca
2        PES    Al         S 1.8305411 0.9194594 0.09185618 0.017952348
0.021542818
5        PES    Si         S 1.8769180 0.8611143 0.11848550 0.017054731
0.018849965
6        PPG    Al         S 0.1601523 0.1130765 0.01826353 0.004282484
0.006108837
9        PPG    Si         S 0.1653794 0.1058340 0.01716142 0.003999084
0.005605015
12      PVDF    Al         S 0.7704856 0.4361732 0.05031181 0.006323343
0.005361095
16      PVDF    Si         S 0.7707605 0.4233890 0.05113660 0.010172334
0.005910951
             S
2  0.025731699
5  0.022440435
6  0.001416998
9  0.001354021
12 0.024193659
16 0.024056195
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
>
> fit1 <- aov(O~Membrane, data1)
> summary(fit1)
            Df Sum Sq Mean Sq F value   Pr(>F)
Membrane     2 0.6162  0.3081   510.7 0.000158 ***
Residuals    3 0.0018  0.0006
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = O ~ Membrane, data = data1)

$Membrane
               diff        lwr        upr     p adj
PPG-PES  -0.7808316 -0.8834745 -0.6781887 0.0000616
PVDF-PES -0.4605058 -0.5631487 -0.3578629 0.0007358
PVDF-PPG  0.3203259  0.2176830  0.4229688 0.0020243

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=O)) +
+   geom_boxplot(fill="green") +
```

```
+   xlab("Membrane")+
+   ylab("Oxygen (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 1)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="O",
+                            groupnames=c("Membrane"))
> box_1_data
  Membrane         O          sd
1      PES 0.8902869 0.041256238
2      PPG 0.1094552 0.005121176
3     PVDF 0.4297811 0.009039758


> # Laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Stub
Wastewater         C         O          N           P          Ca
1       PES   Al           L 0.5616050 0.3080315 0.06421472 0.023296502
0.019712425
4       PES   Si           L 0.5734524 0.2953877 0.06839614 0.022002252
0.018019944
7       PPG   Al           L 0.3273256 0.2006818 0.03916217 0.008442923
0.008767651
11     PVDF   Al           L 0.8543386 0.3780259 0.07601758 0.015808357
0.023643803
15     PVDF   Si           L 0.8581876 0.3667539 0.08935158 0.017732852
0.022269163
             S
1  0.010055328
4  0.009358424
7  0.003052441
11 0.016083285
15 0.015533429
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data1
>
> fit2 <- aov(O~Membrane, data2)
> summary(fit2)
            Df   Sum Sq  Mean Sq F value  Pr(>F)
Membrane     2 0.019873 0.009936   138.5 0.00717 **
Residuals    2 0.000143 0.000072
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

```
Fit: aov(formula = O ~ Membrane, data = data2)

$Membrane
                diff         lwr          upr     p adj
PPG-PES  -0.10102783 -0.16213201 -0.03992365 0.0188936
PVDF-PES  0.07068029  0.02078893  0.12057164 0.0255581
PVDF-PPG  0.17170811  0.11060393  0.23281230 0.0065700

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=O)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Oxygen (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0,1)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
>



> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="O",
+                              groupnames=c("Membrane"))
> box_2_data
  Membrane         O          sd
1      PES 0.3017096 0.008940537
2      PPG 0.2006818          NA
3     PVDF 0.3723899 0.007970540

> # SL wastewater----
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
   Membrane Stub
Wastewater         C         O         N          P          Ca
3       PES   AL          SL 0.3782840 0.1837664 0.00000000 0.00000000
0.02363830
8       PPG   Al          SL 0.5836795 0.4122643 0.06376837 0.01844340
0.01735850
10      PPG   Si          SL 0.5994709 0.4059960 0.07220653 0.01976940
0.01687632
13     PVDF   Al          SL 0.9135652 0.5688396 0.09469366 0.03150809
0.01880321
14     PVDF   Si          SL 0.9582863 0.5325883 0.07046970 0.02964471
0.01846442
             S
3  0.023851900
8  0.008920339
10 0.006509436
13 0.015754045
14 0.015754045
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
```

```
> # Statistical analysis on data1
>
> fit3 <- aov(O~Membrane, data3)
> summary(fit3)
            Df  Sum Sq Mean Sq F value  Pr(>F)
Membrane     2 0.09021 0.04511   133.3 0.00745 **
Residuals    2 0.00068 0.00034
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = O ~ Membrane, data = data3)

$Membrane
              diff        lwr       upr     p adj
PPG-PES  0.2253637 0.09265278 0.3580746 0.0179279
PVDF-PES 0.3669475 0.23423661 0.4996585 0.0067945
PVDF-PPG 0.1415838 0.03322581 0.2499418 0.0299098

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=O)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Oxygen (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0,1)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
>


> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="O",
+                             groupnames=c("Membrane"))
> box_3_data
  Membrane         O         sd
1      PES 0.1837664         NA
2      PPG 0.4091301 0.00443239
3     PVDF 0.5507140 0.02563350



N


> # shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Stub
Wastewater         C         O          N            P           Ca
2        PES   Al          S 1.8305411 0.9194594 0.09185618 0.017952348
0.021542818
5        PES   Si          S 1.8769180 0.8611143 0.11848550 0.017054731
0.018849965
```

```
6       PPG   Al          S 0.1601523 0.1130765 0.01826353 0.004282484
0.006108837
9       PPG   Si          S 0.1653794 0.1058340 0.01716142 0.003999084
0.005605015
12      PVDF  Al          S 0.7704856 0.4361732 0.05031181 0.006323343
0.005361095
16      PVDF  Si          S 0.7707605 0.4233890 0.05113660 0.010172334
0.005910951
             S
2  0.025731699
5  0.022440435
6  0.001416998
9  0.001354021
12 0.024193659
16 0.024056195
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
>
> fit1 <- aov(N~Membrane, data1)
> summary(fit1)
            Df   Sum Sq  Mean Sq F value Pr(>F)
Membrane     2 0.007802 0.003901   32.92 0.0091 **
Residuals    3 0.000356 0.000119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = N ~ Membrane, data = data1)

$Membrane
                diff         lwr          upr     p adj
PPG-PES  -0.08745836 -0.13294791 -0.041968814 0.0082417
PVDF-PES -0.05444663 -0.09993618 -0.008957087 0.0310636
PVDF-PPG  0.03301173 -0.01247782  0.078501273 0.1099776

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=N)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Nitronge (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.15)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
>

> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="N",
+                              groupnames=c("Membrane"))
> box_1_data
  Membrane          N          sd
```

```
1      PES 0.10517084 0.0188297699
2      PPG 0.01771248 0.0007793087
3     PVDF 0.05072420 0.0005832104


> # Laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Stub
Wastewater         C          O          N          P          Ca
1        PES    Al          L 0.5616050 0.3080315 0.06421472 0.023296502
0.019712425
4        PES    Si          L 0.5734524 0.2953877 0.06839614 0.022002252
0.018019944
7        PPG    Al          L 0.3273256 0.2006818 0.03916217 0.008442923
0.008767651
11      PVDF    Al          L 0.8543386 0.3780259 0.07601758 0.015808357
0.023643803
15      PVDF    Si          L 0.8581876 0.3667539 0.08935158 0.017732852
0.022269163
             S
1  0.010055328
4  0.009358424
7  0.003052441
11 0.016083285
15 0.015533429
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data2
>
> fit2 <- aov(N~Membrane, data2)
> summary(fit2)
            Df    Sum Sq   Mean Sq F value Pr(>F)
Membrane     2 0.0012670 0.0006335   12.98 0.0715 .
Residuals    2 0.0000976 0.0000488
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = N ~ Membrane, data = data2)

$Membrane
                diff          lwr        upr     p adj
PPG-PES  -0.02714326 -0.077553097 0.02326659 0.1535480
PVDF-PES  0.01637915 -0.024780314 0.05753861 0.2491902
PVDF-PPG  0.04352241 -0.006887437 0.09393225 0.0659442

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=N)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Nitronge (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.15)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
```

```
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
>

> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="N",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane          N          sd
1      PES 0.06630543 0.002956713
2      PPG 0.03916217          NA
3     PVDF 0.08268458 0.009428566

> # SL wastewater----
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
   Membrane Stub
Wastewater         C          O          N          P         Ca
3       PES   AL         SL 0.3782840 0.1837664 0.00000000 0.00000000
0.02363830
8       PPG   Al         SL 0.5836795 0.4122643 0.06376837 0.01844340
0.01735850
10      PPG   Si         SL 0.5994709 0.4059960 0.07220653 0.01976940
0.01687632
13     PVDF   Al         SL 0.9135652 0.5688396 0.09469366 0.03150809
0.01880321
14     PVDF   Si         SL 0.9582863 0.5325883 0.07046970 0.02964471
0.01846442
             S
3  0.023851900
8  0.008920339
10 0.006509436
13 0.015754045
14 0.015754045
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
> # Statistical analysis on data1
>
> fit3 <- aov(N~Membrane, data3)
> summary(fit3)
            Df   Sum Sq   Mean Sq F value Pr(>F)
Membrane     2 0.004747 0.0023736   14.43 0.0648 .
Residuals    2 0.000329 0.0001645
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = N ~ Membrane, data = data3)

$Membrane
               diff          lwr        upr     p adj
PPG-PES  0.06798745 -0.024546334 0.16052123 0.0888138
PVDF-PES 0.08258168 -0.009952097 0.17511547 0.0619806
PVDF-PPG 0.01459424 -0.060959279 0.09014775 0.5842389
```

```
>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=N)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Nitrogen (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.15)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
>


> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="N",
+                             groupnames=c("Membrane"))
> box_3_data
  Membrane          N          sd
1      PES 0.00000000          NA
2      PPG 0.06798745 0.005966679
3     PVDF 0.08258168 0.017128927
```

P

```
> # Phosphorous ------------------------------
>
> # shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Stub
Wastewater         C         O          N          P          Ca
2       PES    Al          S 1.8305411 0.9194594 0.09185618 0.017952348
0.021542818
5       PES    Si          S 1.8769180 0.8611143 0.11848550 0.017054731
0.018849965
6       PPG    Al          S 0.1601523 0.1130765 0.01826353 0.004282484
0.006108837
9       PPG    Si          S 0.1653794 0.1058340 0.01716142 0.003999084
0.005605015
12     PVDF    Al          S 0.7704856 0.4361732 0.05031181 0.006323343
0.005361095
16     PVDF    Si          S 0.7707605 0.4233890 0.05113660 0.010172334
0.005910951
             S
2  0.025731699
5  0.022440435
6  0.001416998
9  0.001354021
12 0.024193659
16 0.024056195
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
```

```
>
> fit1 <- aov(P~Membrane, data1)
> summary(fit1)
            Df    Sum Sq  Mean Sq F value  Pr(>F)
Membrane     2 1.874e-04 9.37e-05   35.81 0.00806 **
Residuals    3 7.850e-06 2.62e-06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = P ~ Membrane, data = data1)

$Membrane
                 diff          lwr          upr     p adj
PPG-PES  -0.013362756 -0.020122535 -0.006602976 0.0076084
PVDF-PES -0.009255701 -0.016015480 -0.002495922 0.0215030
PVDF-PPG  0.004107055 -0.002652725  0.010866834 0.1627951

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=P)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Phosphorous(g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.04)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
>

> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="P",
+                             groupnames=c("Membrane"))
> box_1_data
  Membrane           P           sd
1      PES 0.017503539 0.0006347111
2      PPG 0.004140784 0.0002003941
3     PVDF 0.008247839 0.0027216476


> # Laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Stub
Wastewater         C         O          N          P          Ca
1        PES   Al          L 0.5616050 0.3080315 0.06421472 0.023296502
0.019712425
4        PES   Si          L 0.5734524 0.2953877 0.06839614 0.022002252
0.018019944
7        PPG   Al          L 0.3273256 0.2006818 0.03916217 0.008442923
0.008767651
11      PVDF   Al          L 0.8543386 0.3780259 0.07601758 0.015808357
0.023643803
15      PVDF   Si          L 0.8581876 0.3667539 0.08935158 0.017732852
0.022269163
```

```
                S
1  0.010055328
4  0.009358424
7  0.003052441
11 0.016083285
15 0.015533429
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data2
>
> fit2 <- aov(P~Membrane, data2)
> summary(fit2)
            Df    Sum Sq   Mean Sq F value Pr(>F)
Membrane     2 1.361e-04 6.806e-05   50.61 0.0194 *
Residuals    2 2.690e-06 1.340e-06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = P ~ Membrane, data = data2)

$Membrane
                 diff           lwr           upr     p adj
PPG-PES  -0.014206454 -2.257264e-02 -0.0058402648 0.0179295
PVDF-PES -0.005878773 -1.270974e-02  0.0009521924 0.0663402
PVDF-PPG  0.008327681 -3.850771e-05  0.0166938707 0.0504400

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=P)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Phosphorous (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.03)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
>


> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="P",
+                            groupnames=c("Membrane"))
> box_2_data
  Membrane           P          sd
1      PES 0.022649377 0.000915173
2      PPG 0.008442923          NA
3     PVDF 0.016770605 0.001360823


> # SL wastewater----
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
```

```
> data3
   Membrane Stub
Wastewater         C         O          N          P          Ca
3        PES    AL         SL 0.3782840 0.1837664 0.00000000 0.00000000
0.02363830
8        PPG    Al         SL 0.5836795 0.4122643 0.06376837 0.01844340
0.01735850
10       PPG    Si         SL 0.5994709 0.4059960 0.07220653 0.01976940
0.01687632
13      PVDF    Al         SL 0.9135652 0.5688396 0.09469366 0.03150809
0.01880321
14      PVDF    Si         SL 0.9582863 0.5325883 0.07046970 0.02964471
0.01846442
             S
3  0.023851900
8  0.008920339
10 0.006509436
13 0.015754045
14 0.015754045
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
> # Statistical analysis on data1
>
> fit3 <- aov(P~Membrane, data3)
> summary(fit3)
            Df    Sum Sq   Mean Sq F value   Pr(>F)
Membrane     2 0.0006252 3.126e-04   239.1 0.00417 **
Residuals    2 0.0000026 1.310e-06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = P ~ Membrane, data = data3)

$Membrane
              diff         lwr        upr     p adj
PPG-PES  0.0191064 0.010856355 0.02735645 0.0097158
PVDF-PES 0.0305764 0.022326351 0.03882644 0.0034035
PVDF-PPG 0.0114700 0.004733862 0.01820613 0.0178329

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=P)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Phosphorous (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.15)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
>
```

```
> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="P",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane         P           sd
1      PES 0.0000000           NA
2      PPG 0.0191064 0.0009376215
3     PVDF 0.0305764 0.0013176100
```

S

```
>
> # shower wastewater----
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Stub
Wastewater        C          O          N          P          Ca
2       PES   Al          S 1.8305411 0.9194594 0.09185618 0.017952348
0.021542818
5       PES   Si          S 1.8769180 0.8611143 0.11848550 0.017054731
0.018849965
6       PPG   Al          S 0.1601523 0.1130765 0.01826353 0.004282484
0.006108837
9       PPG   Si          S 0.1653794 0.1058340 0.01716142 0.003999084
0.005605015
12     PVDF   Al          S 0.7704856 0.4361732 0.05031181 0.006323343
0.005361095
16     PVDF   Si          S 0.7707605 0.4233890 0.05113660 0.010172334
0.005910951
             S
2  0.025731699
5  0.022440435
6  0.001416998
9  0.001354021
12 0.024193659
16 0.024056195
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
>
> fit1 <- aov(P~Membrane, data1)
> summary(fit1)
            Df    Sum Sq  Mean Sq F value  Pr(>F)
Membrane     2 1.874e-04 9.37e-05   35.81 0.00806 **
Residuals    3 7.850e-06 2.62e-06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey1 <- TukeyHSD(fit1, conf.level=0.95) #Tukey multiple comparison
> Tukey1 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = P ~ Membrane, data = data1)

$Membrane
                 diff          lwr          upr     p adj
PPG-PES  -0.013362756 -0.020122535 -0.006602976 0.0076084
```

```
PVDF-PES -0.009255701 -0.016015480 -0.002495922 0.0215030
PVDF-PPG  0.004107055 -0.002652725  0.010866834 0.1627951

>
> # Plot
> box_1 <- ggplot(data1, aes(x=Membrane, y=S)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Sulfur(g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.03)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_1
>


> ## Mean and standard deviation
> box_1_data <- data_summary(data1, varname="S",
+                            groupnames=c("Membrane"))
> box_1_data
  Membrane           S           sd
1      PES 0.024086067 2.327275e-03
2      PPG 0.001385509 4.453146e-05
3     PVDF 0.024124927 9.720173e-05


# Laundry wastewater----
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Stub
Wastewater          C         O          N           P          Ca
1       PES   Al          L 0.5616050 0.3080315 0.06421472 0.023296502
0.019712425
4       PES   Si          L 0.5734524 0.2953877 0.06839614 0.022002252
0.018019944
7       PPG   Al          L 0.3273256 0.2006818 0.03916217 0.008442923
0.008767651
11     PVDF   Al          L 0.8543386 0.3780259 0.07601758 0.015808357
0.023643803
15     PVDF   Si          L 0.8581876 0.3667539 0.08935158 0.017732852
0.022269163
             S
1  0.010055328
4  0.009358424
7  0.003052441
11 0.016083285
15 0.015533429
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data2
>
> fit2 <- aov(S~Membrane, data2)
> summary(fit2)
            Df    Sum Sq   Mean Sq F value   Pr(>F)
Membrane     2 1.126e-04 5.629e-05   285.7 0.00349 **
```

```
Residuals    2 3.900e-07 2.000e-07
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = S ~ Membrane, data = data2)

$Membrane
                 diff          lwr          upr     p adj
PPG-PES  -0.006654435 -0.009856677 -0.003452193 0.0120436
PVDF-PES  0.006101481  0.003486861  0.008716101 0.0095699
PVDF-PPG  0.012755916  0.009553674  0.015958158 0.0027369

>
> # Plot
> box_2 <- ggplot(data2, aes(x=Membrane, y=S)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Sulfur (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.03)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_2
>

> ## Mean and standard deviation
> box_2_data <- data_summary(data2, varname="S",
+                             groupnames=c("Membrane"))
> box_2_data
  Membrane           S           sd
1      PES 0.009706876 0.0004927855
2      PPG 0.003052441           NA
3     PVDF 0.015808357 0.0003888069


# SL wastewater----
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
   Membrane Stub
Wastewater         C         O          N          P          Ca
3       PES   AL          SL 0.3782840 0.1837664 0.00000000 0.00000000
0.02363830
8       PPG   Al          SL 0.5836795 0.4122643 0.06376837 0.01844340
0.01735850
10      PPG   Si          SL 0.5994709 0.4059960 0.07220653 0.01976940
0.01687632
13     PVDF   Al          SL 0.9135652 0.5688396 0.09469366 0.03150809
0.01880321
14     PVDF   Si          SL 0.9582863 0.5325883 0.07046970 0.02964471
0.01846442
             S
3  0.023851900
8  0.008920339
10 0.006509436
13 0.015754045
```

```
14 0.015754045
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
> # Statistical analysis on data1
>
> fit3 <- aov(S~Membrane, data3)
> summary(fit3)
            Df    Sum Sq   Mean Sq F value Pr(>F)
Membrane     2 1.821e-04 9.105e-05   62.66 0.0157 *
Residuals    2 2.910e-06 1.450e-06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = S ~ Membrane, data = data3)

$Membrane
                 diff           lwr          upr     p adj
PPG-PES  -0.016137012 -0.0248339465 -0.007440079 0.0150607
PVDF-PES -0.008097855 -0.0167947890  0.000599079 0.0572299
PVDF-PPG  0.008039158  0.0009381407  0.015140174 0.0394475

>
> # Plot
> box_3 <- ggplot(data3, aes(x=Membrane, y=S)) +
+   geom_boxplot(fill="green") +
+   xlab("Membrane")+
+   ylab("Sulfur (g/m2 membrane/100 m3 wastewater)") + labs(title = "",
subtitle=NULL) + ylim(0, 0.03)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_3
>

> ## Mean and standard deviation
> box_3_data <- data_summary(data3, varname="S",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane           S          sd
1      PES 0.023851900          NA
2      PPG 0.007714887 0.001704766
3     PVDF 0.015754045 0.000000000
```

Dry Matter Mass

```
> # Choose data file Metadata(r5)-DryMatter.txt -----
> con <-file.choose(new = FALSE)
> metadata <- read.table(con, header = T, row.names = 1, fill = TRUE)
> head(metadata)
```

```
  Membrane Wastewater DryMatter
1      PPG          S 0.4572842
2      PPG          S 0.2030975
3      PPG          S 0.2842838
4      PPG          L 0.5879786
5      PPG          L 0.6092973
6      PPG          L 0.7510909
> # Define factors for metadata -----
> metadata$Membrane <- factor(metadata$Membrane)
> metadata$Wastewater <- factor(metadata$Wastewater)
> ## Shower wastewater -----------------------------------------------------
>
>
> data1 <- metadata[which(metadata$Wastewater=="S"),]
> data1
   Membrane Wastewater DryMatter
1       PPG          S 0.4572842
2       PPG          S 0.2030975
3       PPG          S 0.2842838
10     PVDF          S 1.8491363
11     PVDF          S 1.4143656
12     PVDF          S 0.8604172
19      PES          S 2.9920583
>
> # Define factors for data1
> data1$Membrane <- factor(data1$Membrane)
> data1$Wastewater <- factor(data1$Wastewater)
>
> # Statistical analysis on data1
>
> fit2<- aov(DryMatter~Membrane, data1)
> summary(fit2)
            Df Sum Sq Mean Sq F value  Pr(>F)
Membrane     2  5.637  2.8184   21.48 0.00726 **
Residuals    4  0.525  0.1312
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey2 <- TukeyHSD(fit2, conf.level=0.95) #Tukey multiple comparison
> Tukey2 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = DryMatter ~ Membrane, data = data1)

$Membrane
              diff          lwr        upr     p adj
PPG-PES  -2.677170 -4.167895013 -1.1864447 0.0067381
PVDF-PES -1.617419 -3.108143816 -0.1266935 0.0386546
PVDF-PPG  1.059751  0.005649313  2.1138531 0.0491773

>
> # Plot
> box_3 <- ggplot(data1, aes(x=Membrane, y=DryMatter)) +
+   geom_boxplot(fill="green") +
+   xlab("")+
+   ylab("Dry matter (g/m2 membrane/m3 treated water)") + labs(title = "",
subtitle=NULL) + ylim(0, 3)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
```

```
> box_3
>


> ## Mean and standard deviation
> box_3_data <- data_summary(data1, varname="DryMatter",
+                            groupnames=c("Membrane"))
> box_3_data
  Membrane DryMatter        sd
1      PES 2.9920583        NA
2      PPG 0.3148885 0.1298276
3     PVDF 1.3746397 0.4955552


## Laundry wastewater -----------------------------------------------------
>
>
> data2 <- metadata[which(metadata$Wastewater=="L"),]
> data2
   Membrane Wastewater DryMatter
4       PPG          L 0.5879786
5       PPG          L 0.6092973
6       PPG          L 0.7510909
13     PVDF          L 0.7056501
14     PVDF          L 1.3627678
15     PVDF          L 1.0236492
22      PES          L 0.9210376
23      PES          L 0.9123201
24      PES          L 1.1533741
>
> # Define factors for data2
> data2$Membrane <- factor(data2$Membrane)
> data2$Wastewater <- factor(data2$Wastewater)
>
> # Statistical analysis on data1
>
> fit3<- aov(DryMatter~Membrane, data2)
> summary(fit3)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.2664 0.13319    2.97  0.127
Residuals    6 0.2691 0.04485
> Tukey3 <- TukeyHSD(fit3, conf.level=0.95) #Tukey multiple comparison
> Tukey3 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = DryMatter ~ Membrane, data = data2)

$Membrane
               diff        lwr       upr     p adj
PPG-PES  -0.3461217 -0.8766619 0.1844186 0.1925331
PVDF-PES  0.0351118 -0.4954284 0.5656520 0.9776092
PVDF-PPG  0.3812334 -0.1493068 0.9117737 0.1487777

>
> # Plot
> box_4 <- ggplot(data2, aes(x=Membrane, y=DryMatter)) +
+   geom_boxplot(fill="green") +
+   xlab("")+
+   ylab("Dry matter (g/m2 membrane/100 m3 treated water)") + labs(title =
"", subtitle=NULL) + ylim(0, 3)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
```

```
+           axis.text.y=element_text(size=20, family="Times New Roman"),
+           axis.title.y = element_text(size = 20, family="Times New Roman"),
+           axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_4
>


> ## Mean and standard deviation
> box_4_data <- data_summary(data2, varname="DryMatter",
+                            groupnames=c("Membrane"))
> box_4_data
  Membrane DryMatter         sd
1      PES 0.9955773 0.13672554
2      PPG 0.6494556 0.08866181
3     PVDF 1.0306891 0.32861543


## Laundry/shower wastewater ------------------------------------------------
-----
>
>
> data3 <- metadata[which(metadata$Wastewater=="SL"),]
> data3
   Membrane Wastewater DryMatter
7       PPG         SL 1.1158961
8       PPG         SL 0.9166341
9       PPG         SL 1.5838235
16     PVDF         SL 1.4023320
17     PVDF         SL 1.8095556
18     PVDF         SL 1.8700622
26      PES         SL 0.7119966
>
> # Define factors for data3
> data3$Membrane <- factor(data3$Membrane)
> data3$Wastewater <- factor(data3$Wastewater)
>
> # Statistical analysis on data1
>
> fit4<- aov(DryMatter~Membrane, data3)
> summary(fit4)
            Df Sum Sq Mean Sq F value Pr(>F)
Membrane     2 0.8245  0.4122    4.53 0.0938 .
Residuals    4 0.3640  0.0910
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Tukey4 <- TukeyHSD(fit4, conf.level=0.95) #Tukey multiple comparison
> Tukey4 #Output Tukey results
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = DryMatter ~ Membrane, data = data3)

$Membrane
              diff        lwr      upr     p adj
PPG-PES  0.4934546 -0.7480251 1.734934 0.4157041
PVDF-PES 0.9819867 -0.2594930 2.223466 0.0992387
PVDF-PPG 0.4885321 -0.3893267 1.366391 0.2315406

>
> # Plot
> box_5 <- ggplot(data3, aes(x=Membrane, y=DryMatter)) +
+   geom_boxplot(fill="green") +
+   xlab("")+
```

```
+   ylab("Dry matter (g/m2 membrane/100 m3 treated water)") + labs(title =
"", subtitle=NULL) + ylim(0, 3)+
+   theme_classic() +
+   theme(title=element_text(size=20, family="Times New Roman"),
+         axis.text.x = element_text(size=20, family="Times New Roman"),
+         axis.text.y=element_text(size=20, family="Times New Roman"),
+         axis.title.y = element_text(size = 20, family="Times New Roman"),
+         axis.title.x=element_text(size=20, family="Times New Roman"),
legend.position = "top")
> box_5
>
```

```
> ## Mean and standard deviation
> box_5_data <- data_summary(data3, varname="DryMatter",
+                            groupnames=c("Membrane"))
> box_5_data
  Membrane DryMatter        sd
1      PES 0.7119966        NA
2      PPG 1.2054512 0.3424916
3     PVDF 1.6939833 0.2543828
```

MOO

```
> # Loading the libraries
> library(rPref)
> library(dplyr)
> library(igraph)
> library(ggplot2)
> library(rmoo)
>
> ## the .txt file needs to be saved as the type of "Tab delimited".
>
> ## Choose "meta_data_MOO.txt", and the data table should be .txt
>
> con1 <- file.choose(new = FALSE)
> metadata <- read.table(con1, header=T)
> ## View the data structure
> View(metadata)
> # Calculate and plot Skyline for flux and powder mass
> sky1 <- psel(metadata,  high(Flux)*low(Powder_mass))
> ggplot(metadata, aes(x = Flux, y = Powder_mass)) + xlim(0.1, 0.5) + ylim(0,
4) +
+   geom_point(shape = 21) + geom_point(data = sky1, color="Blue",size = 3)
+
+   geom_text(aes(label=Name), hjust=-0.1, vjust=1)
> # Calculate and plot Skyline for flux and Turbidity reduction
> sky2 <- psel(metadata, high(Flux) * high(Turbidity_reduction))
> ggplot(metadata, aes(x = Flux, y = Turbidity_reduction)) + xlim(0.1, 0.5) +
ylim(85, 100) +
+   geom_point(shape = 21) + geom_point(data = sky2, color="Blue",size = 3)
+
+   geom_text(aes(label=Name), hjust=-0.1, vjust=1)
> # Calculate and plot Skyline for flux and COD reduction
> sky3 <- psel(metadata, high(Flux) * high(COD_reduction))
> ggplot(metadata, aes(x = Flux, y = COD_reduction)) + xlim(0.1, 0.5) +
ylim(0, 100) +
```

```
+   geom_point(shape = 21) + geom_point(data = sky3, color="Blue",size = 3)
+
+   geom_text(aes(label=Name), hjust=-0.1, vjust=1)
> # Calculate and plot Skyline for flux and TN reduction
> sky4 <- psel(metadata, high(Flux) * high(TN_reduction))
> ggplot(metadata, aes(x = Flux, y = TN_reduction)) + xlim(0.1, 0.5) +
ylim(0, 100) +
+   geom_point(shape = 21) + geom_point(data = sky4, color="Blue",size = 3)
+
+   geom_text(aes(label=metadata$Name), hjust=-0.1, vjust=1)
> # Calculate and plot Skyline for flux and TP reduction
> sky5 <- psel(metadata, high(Flux) * high(TP_reduction))
> ggplot(metadata, aes(x = Flux, y = TP_reduction)) + xlim(0.1, 0.5) +
ylim(0, 100) +
+   geom_point(shape = 21) + geom_point(data = sky5, color="Blue",size = 3)
+
+   geom_text(aes(label=metadata$Name), hjust=-0.1, vjust=1)
> # Calculate and plot Skyline for flux and UV254 reduction
> sky6 <- psel(metadata, high(Flux) * high(UV254_reduction))
> ggplot(metadata, aes(x = Flux, y = UV254_reduction)) + xlim(0, 0.5) +
ylim(0, 100) +
+   geom_point(shape = 21) + geom_point(data = sky6, color="Blue",size = 3)
+
+   geom_text(aes(label=metadata$Name), hjust=-0.1, vjust=1)
> # Consider the preference from above
> p1 <- high(Flux) * low(Powder_mass)
> # Calculate the level-value w.r.t. p by using top-all
> res1 <- psel(metadata, p1, top=nrow(metadata))
> # Visualize the level values by the color of the points
> gp1 <- ggplot(res1, aes(x = Flux, y=Powder_mass, color=factor(.level))) +
+   xlim(0.1, 0.5) + ylim(0,4)+
+   geom_point(size = 3) + geom_text(aes(label=res1$Name), size=5,
family="Times New Roman", hjust=-0.1, vjust=0)+
+   labs(x="Flux (m3/m2/min)", y="Powder mass (g/m2 membrane/100 m3 treated
water)", color="Level")+
+   theme(title=element_text(size=15, family ="Times New Roman"),
axis.title.x=element_text(size=15, family="Times New Roman"),
axis.title.y=element_text(size=15, family="Times New Roman"))+
+   theme(legend.position="right", legend.text=element_text(size=15,
family="Times New Roman"))
> gp1
> # gp1+geom_step(direction="vh")
> gp1+geom_line()
> # Consider the preference from above
> p2 <- high(Flux) * high(Turbidity_reduction)
> # Calculate the level-value w.r.t. p by using top-all
> res2 <- psel(metadata, p2, top=nrow(metadata))
> # Visualize the level values by the color of the points
> gp2 <- ggplot(res2, aes(x = Flux, y=Turbidity_reduction,
color=factor(.level))) +
+   xlim(0.1, 0.5) + ylim(85,100)+
+   geom_point(size = 3) + geom_text(aes(label=res2$Name),size=5,
family="Times New Roman", hjust=-0.1, vjust=0)+
+   labs(x="Flux (m3/m2/minute)", y="Turbidity reduction (%)",
color="Level")+
+   theme(title=element_text(size=15, family ="Times New Roman"),
axis.title.x=element_text(size=12, family ="Times New Roman"),
axis.title.y=element_text(size=15, family ="Times New Roman"))+
+   theme(legend.position="right", legend.text=element_text(size=12,
family="Times New Roman"))
> gp2
> # gp1+geom_step(direction="vh")
> gp2+geom_line()
> # Consider the preference from above
```

```
> p3 <- high(Flux) * high(COD_reduction)
> # Calculate the level-value w.r.t. p by using top-all
> res3 <- psel(metadata, p3, top=nrow(metadata))
> # Visualize the level values by the color of the points
> gp3 <- ggplot(res3, aes(x = Flux, y=COD_reduction, color=factor(.level)))
+
+ xlim(0.1, 0.5) + ylim(0,80)+
+   geom_point(size = 3) + geom_text(aes(label=res3$Name),size=5,
family="Times New Roman", hjust=-0.1, vjust=0)+
+   labs(x="Flux (m3/m2/minute)", y="COD reduction (%)", color="Level")+
+   theme(title=element_text(size=15, family ="Times New Roman"),
axis.title.x=element_text(size=12, family ="Times New Roman"),
axis.title.y=element_text(size=15, family ="Times New Roman"))+
+   theme(legend.position="right", legend.text=element_text(size=12,
family="Times New Roman"))
> gp3
> # gp1+geom_step(direction="vh")
> gp3+geom_line()
> # Consider the preference from above
> p4 <- high(Flux) * high(UV254_reduction)
> # Calculate the level-value w.r.t. p by using top-all
> res4 <- psel(metadata, p4, top=nrow(metadata))
> # Visualize the level values by the color of the points
> gp4 <- ggplot(res4, aes(x = Flux, y=UV254_reduction, color=factor(.level)))
+
+   xlim(0.1, 0.5) + ylim(50,90)+
+   geom_point(size = 3) + geom_text(aes(label=res4$Name),size=5,
family="Times New Roman", hjust=-0.1, vjust=0)+
+   labs(x="Flux (m3/m2/minute)", y="UV254 reduction (%)", color="Level")+
+   theme(title=element_text(size=15, family ="Times New Roman"),
axis.title.x=element_text(size=12, family ="Times New Roman"),
axis.title.y=element_text(size=15, family ="Times New Roman"))+
+   theme(legend.position="right", legend.text=element_text(size=12,
family="Times New Roman"))
> gp4

> # gp1+geom_step(direction="vh")
> gp4+geom_line()
```

APPENDIX B : SUPPLEMENTAL TABLES AND FIGURES

Table S1. Analytic methods of pharmaceuticals and personal care products (PPCPs).

Characteristic	Analysis Method
Acetone (ug/L)	EPA 8260B
Benzyl alcohol (ug/L)	EPA 8270C
Caffeine (ug/L)	L220
Chloroform (ug/L)	EPA 8260B
N, N-Diethyl-Meta-Toluamide (DEET) (ug/L)	L220
Di(2-ethylhexyl) phthalate (ug/L)	EPA 8270C
Ibuprofen (ug/L)	L221
Methylphenol (ug/L)	EPA 8270C
Nicotine (ug/L)	L220
Permethrin (ug/L)	EPA 8081B
Phenol (ug/L)	EPA 8270C
Salicylic acid (ug/L)	L221

Table S2. Life cycle inventory of different treatment combinations.

		Item	Value	Unit	Data source
Life cycle impacts	GWP	Methane conversion factor of the activated sludge treatment	0		(International, 2010)
		N2O emission factor	0.005	g N emitted as N_2O/g TN in the wastewater	(International, 2010)
		Molecular weight conversion of N2O per N2	1.5714		
		GWP factor of N2O emission	298	Kg CO_2-e/kg N_2O	(Bare, 2011)
		GWP factor of CH4 emission	25		(Bare, 2011)
		GWP factor of natural gas electricity	0.491	Kg CO_2-e/kWh	(Bare, 2011; EPA, 1995)
		GWP factor of diesel electricity	0.731	Kg CO_2-e/kWh	(Bare, 2011)

Table S2 (cont'd)

		GWP factor of diesel electricity	0.731	Kg CO_2-e/kWh	(Bare, 2011; EPA, 1995)
	WEP	WEP factor of TN	0.9864	Kg N-eq/kg TN	(Bare, 2011)
		WEP factor of TP	7.29	Kg N-eq/kg TP	(Bare, 2011)
		WEP factor of COD	0.05	Kg N-eq/kg COD	(Bare, 2011)
	Smog potential	Smog factor of CH_4	0.01438	Kg O_3-eq/kg CH_4	(Bare, 2011)
		Smog factor of N_2O	24.8	Kg O_3-eq/kg N_2O	(Bare, 2011)
		Smog factor of natural gas electricity	0.035	Kg O_3-eq/kWh	(Bare, 2011; EPA, 1995)
		Smog factor of diesel electricity	0.486	Kg O_3-eq/kWh	(Bare, 2011; EPA, 1995)
	Eco-toxicity on freshwater	Eco-Tox factor of caffeine	69878.8	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of methylphenol	0	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of permethrin	1176813.7	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of phenol	933.05	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of Di-2-ethylhexyl-phthalate	322.45	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of salicylic acid	160.91	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of nicotine	3950.20	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of DEET	224.43	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of benzyl alcohol	200.44	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of ibuprofen	208.93	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox of chloroform	41.18	CTUeco/kg substance	(Bare, 2011)

Table S2 (cont'd)

		Eco-Tox of acetone	1.21	CTUeco/kg substance	(Bare, 2011)
	Eco-toxicity on soil	Eco-Tox factor of caffeine	16573.55	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of methylphenol	0	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of permethrin	781.10	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of phenol	68.89	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of Di-2-ethylhexyl-phthalate	0.04	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of salicylic acid	28.24	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of nicotine	138.70	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of DEET	24.87	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of benzyl alcohol	39.30	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox factor of ibuprofen	3.67	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox of chloroform	2.09	CTUeco/kg substance	(Bare, 2011)
		Eco-Tox of acetone	0.26	CTUeco/kg substance	(Bare, 2011)
Treatment	A	Mixed wastewater to the activated sludge treatment	55.6	m^3/day	Data
		Total nitrogen of the mixed wastewater	49	mg/L	Data
		Daily methane production from anaerobic digestion treatment of activated sludge	2.3	m^3/day	Data
		Methane leaking factor of the methane combustion	2	%	(International, 2010)
		Energy demand of the activated sludge treatment	110.2	kWh-e/day	Data
		Energy demand of the UF/RO treatment	108.3	kWh-e/day	Data
		Energy demand of AD treatment	-11.55	kWh-e/day	Data
		Recycled water amount	39.5	m^3/day	Data
		TN concentration of the recycled water	0.16	mg/L	Data

Table S2 (cont'd)

		TP concentration of the recycled water	0.01	mg/L	Data
		COD concentration of the recycled water	0.42	mg/L	Data
		Caffeine in the recycled water	0	ug/L	Data
		Methylphenol in the recycled water	0	ug/L	Data
		Permethrin in the recycled water	0.09	ug/L	Data
		Phenol in the recycled water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the recycled water	0.08	ug/L	Data
		Salicylic acid in the recycled water	2.62	ug/L	Data
		Nicotine in the recycled water	0	ug/L	Data
		DEET in the recycled water	15.95	ug/L	Data
		Benzyl alcohol in the recycled water	0.12	ug/L	Data
		Ibuprofen in the recycled water	0	ug/L	Data
		Chloroform in the recycled water	0.12	ug/L	Data
		Acetone in the recycled water	0	ug/L	Data
		Digestion sludge amount	0.056	m^3/day	Data
		TN concentration of the digestion sludge	3810	mg/L	Data
		TP concentration of the digestion sludge	3165	mg/L	Data
		COD concentration of the digestion sludge	44242	mg/L	Data
		Caffeine in the digestion sludge	0	ug/L	Data
		Methylphenol in the digestion sludge	0	ug/L	Data
		Permethrin in the digestion sludge	0.07	ug/L	Data
		Phenol in the digestion sludge	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the digestion sludge	498	ug/L	Data
		Salicylic acid in the digestion sludge	3.71	ug/L	Data
		Nicotine in the digestion sludge	0	ug/L	Data
		DEET in the digestion sludge	169.5	ug/L	Data
		Benzyl alcohol in the digestion sludge	0	ug/L	Data
		Ibuprofen in the digestion sludge	0	ug/L	Data

Table S2 (cont'd)

		Chloroform in the digestion sludge	0.69	ug/L	Data
		Acetone in the digestion sludge	0	ug/L	Data
	B	Mixed water to the activated sludge treatment	24.7	m^3/day	Data
		TN of the mixed water	73.4	mg/L	Data
		COD of the mixed water	761.4	mg/L	Data
		Daily methane production from anaerobic digestion treatment of activated sludge	1.3	m^3/day	Data
		Methane leaking factor of the methane combustion	2	%	(International, 2010)
		Energy demand of the activated sludge treatment	69	kWh-e/day	Data
		Energy demand of the UF/RO treatment	108.4	kWh-e/day	Data
		Energy demand of AD treatment	-8.2	kWh-e/day	Data
		Recycled water amount	39.8	m^3/day	Data
		TN concentration of the recycled water	0.9	mg/L	Data
		TP concentration of the recycled water	0.04	mg/L	Data
		COD concentration of the recycled water	3.3	mg/L	Data
		Caffeine in the recycled water	0	ug/L	Data
		Methylphenol in the recycled water	0	ug/L	Data
		Permethrin in the recycled water	1.6	ug/L	Data
		Phenol in the recycled water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the recycled water	1.7	ug/L	Data
		Salicylic acid in the recycled water	3.9	ug/L	Data
		Nicotine in the recycled water	1.5	ug/L	Data
		DEET in the recycled water	86	ug/L	Data
		Benzyl alcohol in the recycled water	3.3	ug/L	Data
		Ibuprofen in the recycled water	4.9	ug/L	Data
		Chloroform in the recycled water	0.8	ug/L	Data
		Acetone in the recycled water	2.4	ug/L	Data
		Digestion sludge amount	0.025	m^3/day	Data

Table S2 (cont'd)

		TN concentration of the digestion sludge	5110	mg/L	Data
		TP concentration of the digestion sludge	4376	mg/L	Data
		COD concentration of the digestion sludge	59337	mg/L	Data
		Caffeine in the digestion sludge	0	ug/L	Data
		Methylphenol in the digestion sludge	0	ug/L	Data
		Permethrin in the digestion sludge	0.12	ug/L	Data
		Phenol in the digestion sludge	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the digestion sludge	754	ug/L	Data
		Salicylic acid in the digestion sludge	5.6	ug/L	Data
		Nicotine in the digestion sludge	0	ug/L	Data
		DEET in the digestion sludge	279	ug/L	Data
		Benzyl alcohol in the digestion sludge	0	ug/L	Data
		Ibuprofen in the digestion sludge	0	ug/L	Data
		Chloroform in the digestion sludge	1.1	ug/L	Data
		Acetone in the digestion sludge	0	ug/L	Data
	C	Mixed water to the activated sludge treatment	54.9	m^3/day	Data
		TN of the mixed water	53	mg/L	Data
		COD of the mixed water	539	mg/L	Data
		Energy demand of the activated sludge treatment	97.7	kWh-e/day	Data
		Energy demand of the UF/RO treatment	107	kWh-e/day	Data
		Recycled water amount	39	m^3/day	Data
		TN concentration of the recycled water	0.2	mg/L	Data
		TP concentration of the recycled water	0.01	mg/L	Data
		COD concentration of the recycled water	0.45	mg/L	Data
		Caffeine in the recycled water	0	ug/L	Data
		Methylphenol in the recycled water	0	ug/L	Data
		Permethrin in the recycled water	0.1	ug/L	Data

Table S2 (cont'd)

		Phenol in the recycled water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the recycled water	0.06	ug/L	Data
		Salicylic acid in the recycled water	3.2	ug/L	Data
		Nicotine in the recycled water	0	ug/L	Data
		DEET in the recycled water	18.5	ug/L	Data
		Benzyl alcohol in the recycled water	0.1	ug/L	Data
		Ibuprofen in the recycled water	0	ug/L	Data
		Chloroform in the recycled water	0.14	ug/L	Data
		Acetone in the recycled water	0	ug/L	Data
		Activated sludge amount	0.55	m^3/day	Data
		TN concentration of the digestion sludge	671	mg/L	Data
		TP concentration of the activated sludge	512	mg/L	Data
		COD concentration of the activated sludge	12983	mg/L	Data
		Caffeine in the digestion sludge	0	ug/L	Data
		Methylphenol in the activated sludge	0	ug/L	Data
		Permethrin in the activated sludge	0.8	ug/L	Data
		Phenol in the activated sludge	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the activated sludge	839	ug/L	Data
		Salicylic acid in the activated sludge	92	ug/L	Data
		Nicotine in the activated sludge	0	ug/L	Data
		DEET in the activated sludge	196	ug/L	Data
		Benzyl alcohol in the activated sludge	1.1	ug/L	Data
		Ibuprofen in the activated sludge	0	ug/L	Data
		Chloroform in the activated sludge	1.2	ug/L	Data
		Acetone in the activated sludge	0	ug/L	Data
	D	Mixed water to the activated sludge treatment	24	m^3/day	Data
		TN of the mixed water	70	mg/L	Data
		COD of the mixed water	815	mg/L	Data
		Greywater to the UF treatment	30.7	m^3/day	Data
		TN of the greywater	38	mg/L	Data

Table S2 (cont'd)

		COD of the greywater	386	mg/L	Data
		Energy demand of the activated sludge treatment	62.6	kWh-e/day	Data
		Energy demand of the UF/RO treatment	108	kWh-e/day	Data
		Recycled water amount	39.5	m^3/day	Data
		TN concentration of the recycled water	0.9	mg/L	Data
		TP concentration of the recycled water	0.04	mg/L	Data
		COD concentration of the recycled water	3.3	mg/L	Data
		Caffeine in the recycled water	0	ug/L	Data
		Methylphenol in the recycled water	0	ug/L	Data
		Permethrin in the recycled water	1.6	ug/L	Data
		Phenol in the recycled water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the recycled water	1.7	ug/L	Data
		Salicylic acid in the recycled water	4.3	ug/L	Data
		Nicotine in the recycled water	1.6	ug/L	Data
		DEET in the recycled water	88	ug/L	Data
		Benzyl alcohol in the recycled water	3.4	ug/L	Data
		Ibuprofen in the recycled water	5	ug/L	Data
		Chloroform in the recycled water	0.8	ug/L	Data
		Acetone in the recycled water	2.4	ug/L	Data
		Activated sludge amount	0.24	m^3/day	Data
		TN concentration of the digestion sludge	978	mg/L	Data
		TP concentration of the activated sludge	850	mg/L	Data
		COD concentration of the activated sludge	18937	mg/L	Data
		Caffeine in the digestion sludge	0	ug/L	Data
		Methylphenol in the activated sludge	0	ug/L	Data
		Permethrin in the activated sludge	1.4	ug/L	Data
		Phenol in the activated sludge	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the activated sludge	1368	ug/L	Data

Table S2 (cont'd)

		Salicylic acid in the activated sludge	129	ug/L	Data
		Nicotine in the activated sludge	0	ug/L	Data
		DEET in the activated sludge	312	ug/L	Data
		Benzyl alcohol in the activated sludge	1.8	ug/L	Data
		Ibuprofen in the activated sludge	0	ug/L	Data
		Chloroform in the activated sludge	1.8	ug/L	Data
		Acetone in the activated sludge	0	ug/L	Data
	E	Mixed water to the activated sludge treatment	17.7	m^3/day	Data
		TN of the mixed water	89	mg/L	Data
		COD of the mixed water	925	mg/L	Data
		Greywater to the UF treatment	30.7	m^3/day	Data
		TN of the greywater	38	mg/L	Data
		COD of the greywater	386	mg/L	Data
		Energy demand of the activated sludge treatment	53.4	kWh-e/day	Data
		Energy demand of the UF/RO treatment	60.3	kWh-e/day	Data
		Recycled water amount	22.2	m^3/day	Data
		TN concentration of the recycled water	1.4	mg/L	Data
		TP concentration of the recycled water	0.06	mg/L	Data
		COD concentration of the recycled water	5	mg/L	Data
		Caffeine in the recycled water	0	ug/L	Data
		Methylphenol in the recycled water	0	ug/L	Data
		Permethrin in the recycled water	2.4	ug/L	Data
		Phenol in the recycled water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the recycled water	2.5	ug/L	Data
		Salicylic acid in the recycled water	2.9	ug/L	Data
		Nicotine in the recycled water	2.4	ug/L	Data
		DEET in the recycled water	110.5	ug/L	Data
		Benzyl alcohol in the recycled water	4.9	ug/L	Data
		Ibuprofen in the recycled water	7.5	ug/L	Data

Table S2 (cont'd)

		Chloroform in the recycled water	1.1	ug/L	Data
		Acetone in the recycled water	3.7	ug/L	Data
		Activated sludge amount	0.18	m^3/day	Data
		TN concentration of the digestion sludge	1156	mg/L	Data
		TP concentration of the activated sludge	879	mg/L	Data
		COD concentration of the activated sludge	22367	mg/L	Data
		Caffeine in the digestion sludge	0	ug/L	Data
		Methylphenol in the activated sludge	0	ug/L	Data
		Permethrin in the activated sludge	1.3	ug/L	Data
		Phenol in the activated sludge	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the activated sludge	1337	ug/L	Data
		Salicylic acid in the activated sludge	89	ug/L	Data
		Nicotine in the activated sludge	0	ug/L	Data
		DEET in the activated sludge	271	ug/L	Data
		Benzyl alcohol in the activated sludge	1.9	ug/L	Data
		Ibuprofen in the activated sludge	0	ug/L	Data
		Chloroform in the activated sludge	1.7	ug/L	Data
		Acetone in the activated sludge	0	ug/L	Data
		Discharging water amount	17.5	m^3/day	Data
		TN concentration of the recycled water	8	mg/L	Data
		TP concentration of the recycled water	0.6	mg/L	Data
		COD concentration of the discharging water	56	mg/L	Data
		Caffeine in the discharging water	0	ug/L	Data
		Methylphenol in the discharging water	0	ug/L	Data
		Permethrin in the discharging water	1.3	ug/L	Data
		Phenol in the discharging water	0	ug/L	Data
		Di-2-ethylhexyl-phthalate in the discharging water	0.9	ug/L	Data

Table S2 (cont'd)

		Salicylic acid in the discharging water	89	ug/L	Data
		Nicotine in the discharging water	0	ug/L	Data
		DEET in the discharging water	271	ug/L	Data
		Benzyl alcohol in the discharging water	1.9	ug/L	Data
		Ibuprofen in the discharging water	0	ug/L	Data
		Chloroform in the discharging water	1.7	ug/L	Data
		Acetone in the discharging water	0	ug/L	Data

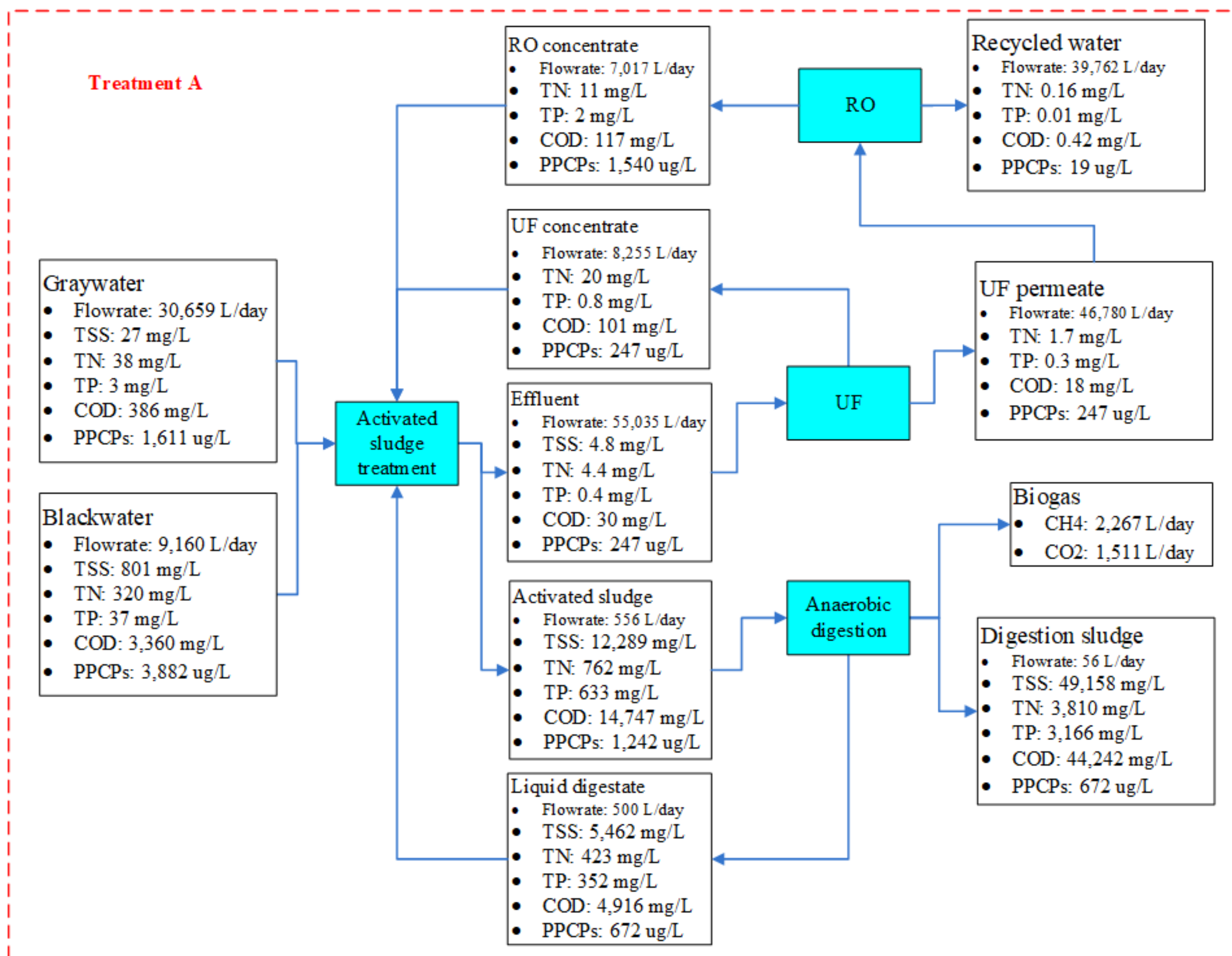


Figure S1. Detailed mass balance of different treatment combinations (Treatment A – E).

Figure S1 (cont'd)

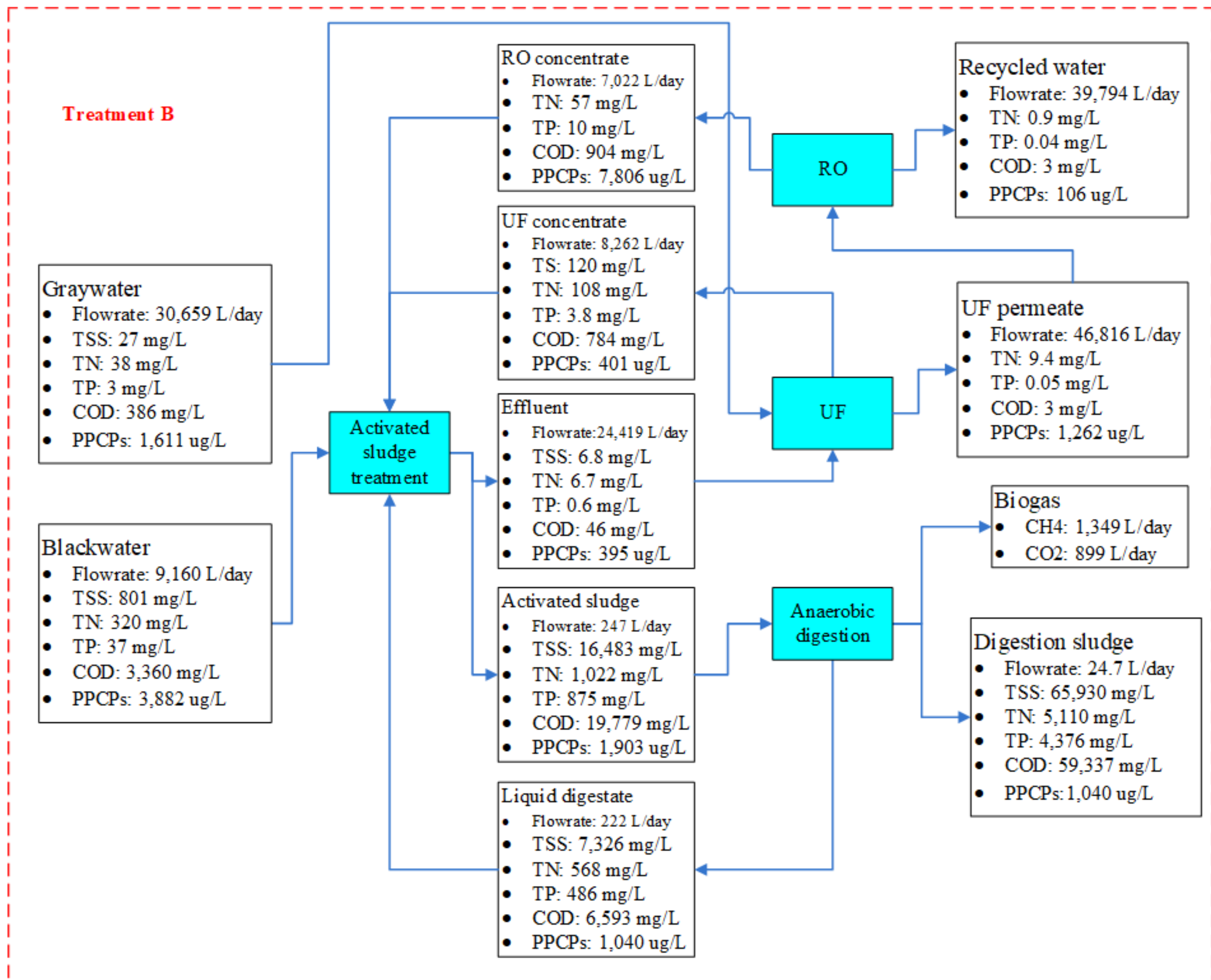


Figure S1 (cont'd)

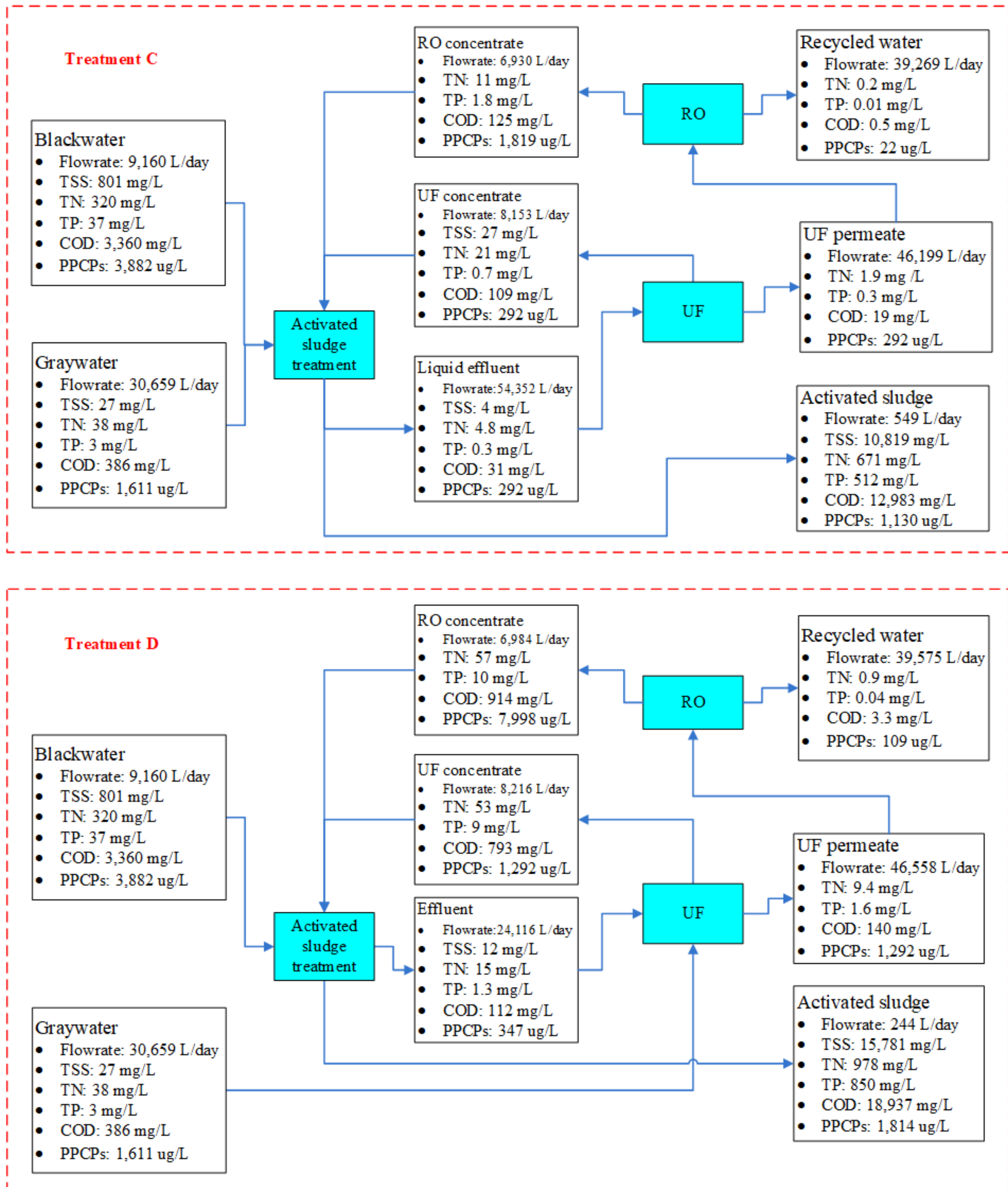


Figure S1 (cont'd)

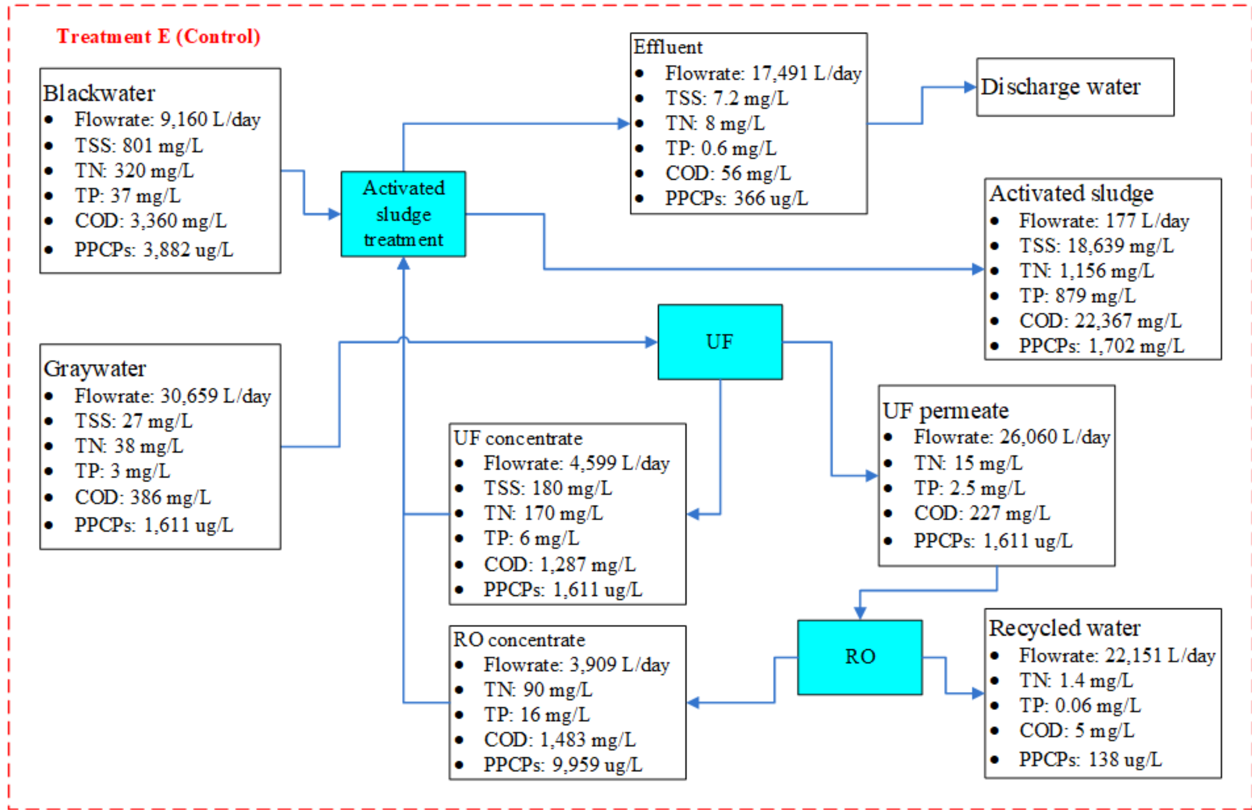


Table S3. Calculation of exergy rates of the treatment.*

Feed amount (L/day)	Stream	Component	Mass flow rate (kg/d)	Specific chemical exergy (kJ/kg)	Chemical exergy rate (kW)	Physical exergy rate (kW)	Exergy rate for each component (kW)
3000	Feed	Organic matter (COD)	8.50	13600	1.3374	-	1.3374
		TN	0.30	23007	0.0790	-	0.0790
		TP	0.09	432	0.0005	-	0.0005
		Electricity for the feeding pump				0.0124	0.0124
		Electricity for the treatment				0.7868	0.7868
	Treated water	Organic matter (COD)	0.40	13600	0.0625		0.0625
		TN	0.04	23007	0.0106		0.0106
		TP	0.01	432	0.00003		0.0000
	Sludge	Organic matter (COD)	4.70	13600	0.7391		0.7391
		TN	1.20	23007	0.3193		0.3192
		TP	0.05	432	0.00026		0.0003
3750	Feed	Organic matter (COD)	10.62	13600	1.6718		1.6718
		TN	0.37	23007	0.0988		0.0988
		TP	0.12	432	0.0006		0.0006
		Electricity for the feeding pump				0.0156	0.0156
		Electricity for the treatment				0.7876	0.7858
	Reclaimed water	Organic matter (COD)	0.52	13600	0.0822		0.0821
		TN	0.03	23007	0.0083		0.0083
		TP	0.01	432	0.00003		0.0000

Table S3 (cont'd)

	Sludge	Organic matter (COD)	6.56	13600	1.0319		1.0319
		TN	1.67	23007	0.4457		0.4457
		TP	0.07	432	0.0004		00004
4500	Feed	Organic matter (COD)	12.74	13600	2.0061		2.0061
		TN	0.45	23007	0.1185		0.1185
		TP	0.14	432	0.0007		0.0007
		Electricity for the feeding pump				0.0187	0.0187
		Electricity for the treatment				0.8073	0.8073
	Reclaimed water	Organic matter (COD)	0.71	13600	0.1184		0.1118
		TN	0.02	23007	0.0060		0.0060
		TP	0.00	432	0.00002		0.0000
	Sludge	Organic matter (COD)	7.34	13600	1.1551		1.1551
		TN	1.87	23007	0.4989		0.4989
		TP	0.08	432	0.0004		0.0004

Table S4. Characteristics of the treated wastewater.

Parameter	**Treated wastewater**		
	3000 LPD	**3750 LPD**	**4500 LPD**
Turbidity (NTU) [a]	23.25 ± 17.14	36.72 ± 21.38	19.59 ± 10.61
TS (mg/L) [b]	833.19 ± 138.98	754.17 ± 46.51	791.43 ± 184.25
TSS (mg/L) [c]	31.95 ± 22.54	54.23±35.30	36.50±34.03
COD (mg/L) [d]	139.61 ± 62.41	147.94±36.17	166.86±74.33
BOD (mg/L) [e]	132.41 ± 116.83	161.60±66.19	121.68±10.41
NH_3 (mg/L) [f]	6.74 ± 2.84	4.96±1.81	1.89±0.90
NO_2 (mg/L) [g]	0.093 ± 0.072	0.065±0.049	0.033±0.014
NO_3 (mg/L) [h]	0.52 ± 0.21	0..43±0.06	0.34±0.096
TOC (mg/L) [i]	65.40 ± 25.43	61.50±31.60	40.25±16.93
TN (mg/L) [j]	14.03 ± 13.13	8.79±2.72	5.26±1.52
TP (mg/L) [k]	2.01 ± 1.49	1.71±1.09	0.99±0.52
Total coliform (Log/ml) [l]	6.26 ± 0.97	5.82±0.15	6.09±0.41
E. coli (Log/ml) [m]	5.35 ± 0.99	5.01±0.51	4.98±0.56

n. Turbidity data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 24, 13, and 8 samples, respectively, with standard deviations.
o. TS data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 36, 12, and 7 samples, respectively, with standard deviations.
p. TSS data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 31, 14, and 8 samples, respectively, with standard deviations.
q. COD data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 31, 17, and 7 samples, respectively, with standard deviations.
r. BOD data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 11, 3, and 3 samples, respectively, with standard deviations.
s. NH_3 *data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 21, 17, and 8 samples, respectively, with standard deviations.*
t. NO_2 *data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 25, 17, and 5 samples, respectively, with standard deviations.*
u. NO_3 *data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 29, 13, and 7 samples, respectively, with standard deviations.*
v. TOC data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 14, 6, and 4 samples, respectively, with standard deviations.
w. TN data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 37, 16, and 7 samples, respectively, with standard deviations.
x. TP data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 35, 17, and 7 samples, respectively, with standard deviations.
y. Total coliform data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 18, 8, and 4 samples, respectively, with standard deviations.
z. E. coli data for the feed amounts of 3000, 3750, and 4500 LPD are averages of 13, 8, and 4 samples, respectively, with standard deviations.

Table S5. Microbial genus identified in all samples.

	Domain	Phylum	Class	Order	Family	Genus
1	*Unassigned*	*Unassigned*	*Unassigned*	*Unassigned*	*Unassigned*	*Unassigned*
2	*Bacteria*	*Bacteria unclassified*	*Bacteria unclassified*	*Bacteria unclassified*	*Bacteria unclassified*	*Bacteria_unclassified*
3	*Bacteria*	*Actinobacteria*	*Actinobacteria unclassified*	*Actinobacteria unclassified*	*Actinobacteria unclassified*	*Actinobacteria_unclassified*
4	*Bacteria*	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*Actinomycetales unclassified*	*Actinomycetales_unclassified*
5	*Bacteria*	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter*
6	*Bacteria*	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomycetaceae_unclassified*
7	*Bacteria*	*Bacteroidetes*	*Bacteroidetes unclassified*	*Bacteroidetes unclassified*	*Bacteroidetes unclassified*	*Bacteroidetes_unclassified*
8	*Bacteria*	*Bacteroidetes*	*Cytophagia*	*Cytophagales*	*Cytophagales unclassified*	*Cytophagales_unclassified*
9	*Bacteria*	*Bacteroidetes*	*Cytophagia*	*Cytophagales*	*Cyclobacteriaceae*	*Cyclobacteriaceae_unclassified*
10	*Bacteria*	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriales unclassified*	*Flavobacteriales_unclassified*
11	*Bacteria*	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacteriaceae_unclassified*
12	*Bacteria*	*Bacteroidetes*	*Sphingobacteriia*	*Sphingobacteriales*	*Sphingobacteriaceae*	*Sphingobacteriaceae_unclassified*
13	*Bacteria*	*Bacteroidetes*	*[Saprospirae]*	*[Saprospirae] unclassified*	*[Saprospirae] unclassified*	*[Saprospirae]_unclassified*
14	*Bacteria*	*Bacteroidetes*	*[Saprospirae]*	*[Saprospirales]*	*[Saprospirales] unclassified*	*[Saprospirales]_unclassified*
15	*Bacteria*	*Bacteroidetes*	*[Saprospirae]*	*[Saprospirales]*	*Chitinophagaceae*	*Chitinophagaceae_unclassified*
16	*Bacteria*	*Cyanobacteria*	*Cyanobacteria unclassified*	*Cyanobacteria unclassified*	*Cyanobacteria unclassified*	*Cyanobacteria_unclassified*
17	*Bacteria*	*Firmicutes*	*Bacilli*	*Bacilli unclassified*	*Bacilli unclassified*	*Bacilli_unclassified*
18	*Bacteria*	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillales unclassified*	*Bacillales_unclassified*
19	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Clostridiales unclassified*	*Clostridiales_unclassified*
20	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*Lachnospiraceae_unclassified*
21	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Clostridium*
22	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*	*Clostridium*
23	*Bacteria*	*Planctomycetes*	*Planctomycetia*	*Pirellulales*	*Pirellulaceae*	*Pirellulaceae_unclassified*

Table S5 (cont'd)

24	*Bacteria*	*Proteobacteria*	*Proteobacteria unclassified*	*Proteobacteria unclassified*	*Proteobacteria unclassified*	*Proteobacteria_unclassified*
25	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Alphaproteobacteria unclassified*	*Alphaproteobacteria unclassified*	*Alphaproteobacteria_unclassified*
26	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Caulobacteraceae_unclassified*
27	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Brevundimonas*
28	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Nitrobacteria*
29	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiales unclassified*	*Rhizobiales_unclassified*
30	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*	*Bradyrhizobiaceae_unclassified*
31	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*	*Hyphomicrobiaceae_unclassified*
32	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylobacteriaceae*	*Methylobacteriaceae_unclassified*
33	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Phyllobacteriaceae*	*Phyllobacteriaceae_unclassified*
34	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodobacteraceae*	*Rhodobacteraceae_unclassified*
35	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillales unclassified*	*Rhodospirillales_unclassified*
36	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Acetobacteraceae*	*Roseomonas*
37	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*	*Rhodospirillaceae_unclassified*
38	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadales unclassified*	*Sphingomonadales_unclassified*
39	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingomonadaceae_unclassified*
40	*Bacteria*	*Proteobacteria*	*Betaproteobacteria*	*Betaproteobacteria unclassified*	*Betaproteobacteria unclassified*	*Betaproteobacteria_unclassified*
41	*Bacteria*	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiales unclassified*	*Burkholderiales_unclassified*
42	*Bacteria*	*Proteobacteria*	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*	*Neisseriaceae_unclassified*
43	*Bacteria*	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacterales*	*Helicobacteraceae*	*Helicobacter*
44	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria unclassified*	*Gammaproteobacteria unclassified*	*Gammaproteobacteria_unclassified*
45	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	*Xanthomonadaceae_unclassified*
46	*Bacteria*	*Verrucomicrobia*	*Verrucomicrobia unclassified*	*Verrucomicrobia unclassified*	*Verrucomicrobia unclassified*	*Verrucomicrobia_unclassified*
47	*Bacteria*	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobiaceae_unclassified*
48	*Bacteria*	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Haloferula*

Table S5 (cont'd)

49	*Bacteria*	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*	*Verrucomicrobium*

*Table S6. Relative abundance of key microbial communities of the treatment at different feed amounts *.*

Microbial communities		Relative abundance (%)			
		Blackwater	Feed amount (L/day)		
			3000	3750	4500
Phylum	*Un-assigned bacteria*	0.19	0.34±0.11	0.51±0.30	3.21±0.74
	Unclassified bacteria	0.85	48.83±4.59	36.67±1.94	18.50±2.21
	Actinobacteria	0.00	1.30±0.34	2.05±1.43	1.81±0.19
	Bacteroidetes	43.75	18.27±6.51	20.77±1.76	15.76±0.26
	Cyanobacteria	0.57	0.54±0.21	0.48±0.04	0.22±0.09
	Firmicutes	0.44	0.65±0.11	0.81±0.09	2.28±0.01
	Planctomycetes	0.00	0.20±0.13	0.02±0.02	0.00±0.00
	proteobacteria	54.21	23.84±2.81	32.59±4.18	47.04±2.82
	Verrucomicrobia	0.00	6.03±0.99	6.10±1.78	11.17±0.03
Bacteroidetes family	*Unclassified bacteroidetes*	1.85	9.70±6.18	5.06±1.01	5.81±0.27
	Unclassified cytophagales	0.00	0.00±0.00	0.00±0.00	0.00±0.00
	Cyclobacteriaceae	0.00	0.09±0.08	0.05±0.04	0.00±0.00
	Unclassified flavobacteriales	20.27	0.02±0.05	0.04±0.07	0.30±0.12
	Flavobacteriales	20.30	1.10±0.46	8.39±0.96	4.62±0.27
	Sphingobacteriaceae	1.32	0.08±0.08	0.25±0.20	1.29±0.01
	Unclassified saprospirae	0.00	0.00±0.00	0.00±0.00	0.00±0.00
	Unclassified saprospirales	0.00	0.05±0.08	0.08±0.14	0.00±0.00
	Chitinophagaceae	0.00	7.23±1.41	6.90±0.94	3.73±0.38
Proteobacteria family	*Unclassified proteobacteria*	47.80	1.95±1.11	2.25±0.63	2.95±0.58
	Unclassified alphaproteobacteria	0.06	0.95±0.78	0.48±0.32	0.26±0.17
	Caulobacteraceae	0.00	1.14±0.34	0.94±0.48	0.07±0.10
	Unclassified rhizobiales	0.06	3.19±2.22	12.25±5.12	10.07±1.08
	Bradyrhizobiaceae	0.00	0.14±0.09	0.02±0.04	0.00±0.00
	Rhodobacteraceae	0.00	1.33±0.64	0.81±0.48	0.75±0.30
	Unclassified rhodospirillales	0.00	0.05±0.05	0.09±0.09	0.05±0.08
	Acetobacteraceae	0.00	0.03±0.05	0.06±0.05	0.00±0.00
	Rhodospirillaceae	0.00	0.02±0.04	0.18±0.14	0.03±0.04
	Unclassified sphingomonadales	0.00	4.02±0.74	5.86±1.35	11.39±0.12

Table S6 (cont'd)

	Unclassified burkholderiales	0.35	3.44±1.74	3.03±0.29	15.84±1.30
	Neisseriaceae	0.00	0.00±0.00	0.03±0.05	0.22±0.15
	Helicobacteraceae	0.79	0.00±0.00	0.00±0.00	0.00±0.00
	Unclassified gammaproteobacteria	1.60	1.48±1.23	0.97±0.23	1.43±0.26
	Xanthomonadaceae	0.25	2.69±0.96	1.52±0.71	1.35±0.85
Verrucomicrobia family	*Unclassified verrucomicrobia*	0.00	1.16±0.53	1.38±0.93	2.58±1.05
	Verrucomicrobiaceae	0.00	4.87±1.14	4.71±0.97	8.59±1.08

**: Data for three feed amounts are average and standard deviation of 2-6 replicates.*