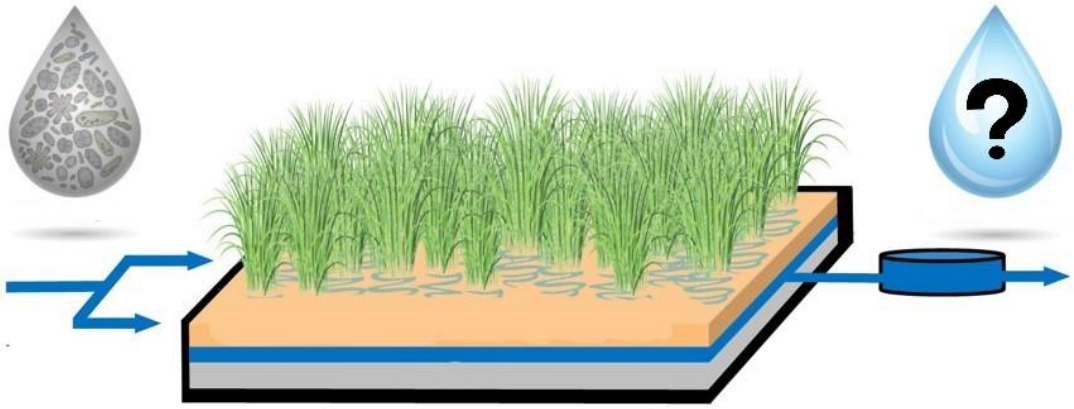
Master's Thesis Internship
Master Water Science and Management

Quantifying the log reduction of pathogenic microorganisms by constructed wetlands as a basis for QMRA of water reuse applications.



Utrecht University

KWR

Name: Sotirios Paraskevopoulos
Student number: 6312829
Email: s.paraskevopoulos@students.uu.nl
Supervisor UU: Steven de Jong
Supervisor KWR: Patrick Smeets
Supervisor KWR email: Patrick.Smeets@kwrwater.nl
Supervisor UU email: S.M.deJong@uu.nl
ECTS: 30
Duration: February-July 2019

Contents

List of abbreviations

BOD= Biological Oxygen Demand
C *in*= Concentration influent
C *out*= Concentration effluent
CFU= Colony-Forming Unit
cNES= combined Natural and Engineered Systems
COD= Chemical Oxygen Demand
CW= Constructed Wetland
E. *Coli* = Escherichia *coli*
FC= Fecal Coliforms
FS= Fecal Streptococcus
FWS= Free Water Surface
GUI= Graphical User Interface
HLR= Hydraulic Loading Rate
HRT= Hydraulic Retention Time
KWR= KWR Watercycle Research Institute
LRV= Log Removal Value
ml=mililiters
MPN= Most Probable Number
n= Porosity
QMRA= Quantitative Microbial Risk Assessment
RIVM= National Institute for Health and Environment
SA= Sensitivity Analysis
SSHFCW = Subsurface Horizontal Flow Constructed Wetland
T= Temperature
TC= Total Coliforms
TKN= Total Kjeldahl Nitrogen
TN= Total Nitrogen
TSS= Total Suspended Solids
VF=Vertical Flow
VFCW= Vertical Flow Constructed Wetland
WW=WasteWater
WWTP=WasteWater Treatment Plant

Preface

This Master's thesis has been written as the final part of the Water Science and Management Masters Programme of Utrecht University, in collaboration with KWR Watercycle Research Institute where I conducted my internship for the last five months. Moreover, the exact duration of the thesis/internship was 21.5 weeks worth of 30 EC and regular meetings were held with the two supervisors both at KWR and Utrecht University.

Abstract

Over the last 30 years, constructed wetlands (CWs) have been used as an alternative, cost-efficient way of treating wastewater, mainly in combination with conventional wastewater techniques. In the context of circular economy and water reuse applications, policymakers have to take precautionary measures regarding water safety, water supply and the acute effects of contamination from surface water. Consequently, water and wastewater companies in the Netherlands are obliged to perform Quantitative Microbial Risk Assessment (QMRA) every three years in order to monitor water quality, calculate the risk of infection, and improve water supply safety. Moreover, RIVM has developed a tool named QMRAspot to analyze and conduct QMRA for specific pathogens that can be found in water sources. KWR contributes to this effort by building a knowledge base regarding pathogen removal and providing RIVM with log removal values from various wastewater processes using another computational tool called Watershare Treatment Calculator. The process of constructed wetlands for wastewater treatment is currently missing from their knowledge base.

The objective of this paper is to quantify the log removal of pathogens from constructed wetlands as a basis for QMRA of water reuse applications. In that way, it will contribute to the efforts of KWR for the development of a knowledge base. To do that, a systematic literature review for the creation of a complete dataset with all the necessary information regarding pathogen removal was performed. Furthermore, three predictive models (one for every type of CW) were developed in order to be incorporated into the Watershare treatment calculator. The models gave log removal values of pathogens based on specific parameters such as Hydraulic Retention Time (HRT) and Hydraulic Loading Rate (HLR) of CWs. Additionally, a storyboard for the Graphical User Interface (GUI) in the platform of the treatment calculator was created.

Based on the systematic literature review, the average log removal for Free Water Surface, Subsurface Horizontal, and Vertical Flow CWs was 1.6, 1.7, and 2.12 log respectively suggesting that CWs can only be used for wastewater polishing when they are combined with conventional wastewater treatment systems. Moreover, all three types of CW showed great variability in log removal with ranges between 0.07-6.08 log, indicating that the complexity of these systems makes it difficult to draw robust conclusions on their performance. Contrary, the results of the predictive models were promising since the predictive models showed great residuals in terms of R^2, RMSE, MAE, and the correlation between observed and predicted values. That means that predicted values of pathogen removal with high precision can be extracted.

Overall, the created dataset, and predictive models can provide guidance to municipalities and water boards regarding the polishing ability of CWs and lay the foundations for a better understanding of the design and operational parameters of CWs

since the decision makers are able to know the required values of HRT and HLR, in order to achieve a certain degree of pathogen removal from their CW.

Acknowledgments

The past five months have been a very interesting and inspiring journey for me, therefore I would like to thank a group of people.

First of all, I would like to express my gratitude to my KWR supervisor Dr. Patrick Smeets. Thank you for helping me and always being there for me in times where I felt lost and desperate. Thank you for guiding and pushing me to be "a good researcher", as well as teaching me how to think critically. Your impact on my scientific career goes beyond this internship. Moreover, special thanks go to my UU supervisor Dr. Steven de Jong for his critical comments and his support since I always had a smile on my face after our meetings. Also, big thanks go to my colleagues from KWR for their contributions during our meetings whenever I needed advice.

In addition, I would like to thank my friends and colleagues from Utrecht University. Bart, Oscar, Jack and little Joel, you have no idea how much you helped me during our journey. Cheers to our endless study sessions at Oscar's casa.

Furthermore, I would like to thank my beloved family who has been by my side through my whole life. I couldn't have made it thus far without you and I am the man I am today because of you. I know you are proud of me and this is what motivates me every day to keep going and try to get better.

Last but not least, I would like to thank my partner in crime. Foteini, thank you so much for your constant support and love during this challenge! You have been my beacon in times I needed light. For this (and for so many more reasons) I will be forever grateful.

Sotiris Paraskevopoulos,
Utrecht, the Netherlands
July 2019

1. Introduction

1.1 Constructed wetlands for wastewater treatment

Constructed wetlands (CWs) have proven to be an alternative, easily operated and cost-efficient system that can be applied to wastewater (WW) purification, municipal sewage (Vymazal, 2005), or polishing of a wastewater treatment plant (WWTP) effluent (Toet et al., 2005). Moreover, the use of CWs provides an additional ecological and recreational value through the creation of habitats, preservation of wildlife and social acceptance, making them preferable compared to conventional technical systems in terms of ecosystem services (Rousseau et al., 2008). However, CWs can have disadvantages like low nutrient removal efficiency (Vymazal, 2005) as well as limitations regarding microbiological degradation processes with several factors affecting the process like temperature and seasonal variations (Akratos & Tsihrintzis, 2007). Furthermore, while conventional engineered wastewater treatment has already proven to be efficient in the removal of the majority of compounds, they come with the disadvantage of environmental degradation and high energy consumption amongst others (Weber & Legge, 2008). Therefore, a potential combination of these natural systems with engineered treatments like oxidation, MBR (membrane bio-reactor) or anaerobic reactors as a pre or post-treatment, can overcome these disadvantages and even enhance the efficiency of the combined Natural and Engineered Systems (cNES) as it combines the removal mechanisms of both system types (Liu et al., 2015).

When the end product of wastewater treatment is destined for water reuse applications like drinking water or irrigation, pathogenic bacteria that can be found in the effluents, are of great importance for effective removal. While natural systems can remove the remaining micropollutants of conventional WW treatments adequately when they are attached at the end of them (a schematic representation of pathogen removal and cNES can be found in Figures 1 and 2 respectively), recent studies have shown that when CWs are used as the main wastewater treatment method for agricultural reuse of effluents, they perform poorly on meeting the accepted limit of microbial contamination (Lavrnić & Mancini, 2016; Marecos do Monte & Albuquerque, 2010). Therefore, the need for a comprehensive exploration of the performance of CWs on pathogen removal as a standalone wastewater treatment method or as a combined treatment is imperative.

1.2 KWR and QMRA treatment calculator

KWR has a water-wise world research program which focuses on an optimal organization and management of the water cycle, having the concept of circular economy as a key driver. Part of the planning of KWR is to build a knowledge base for pathogen removal by various water and wastewater treatment processes, through collaborations with different projects such as AquaNES[1] and BTO[2] (Smeets, 2017). The process of natural systems for wastewater treatment or the combination between natural and engineered systems is currently missing from the knowledge base.

[1] AquaNES will catalyze innovations in water and wastewater treatment processes and management through improved combinations of natural and engineered components. Among the demonstrated solutions are natural treatment processes such as bank filtration (BF), managed aquifer recharge (MAR) and constructed wetlands (CW) plus engineered pre- and post-treatment options. More info: www.aquanes.eu

[2] Joint drinking water research program.

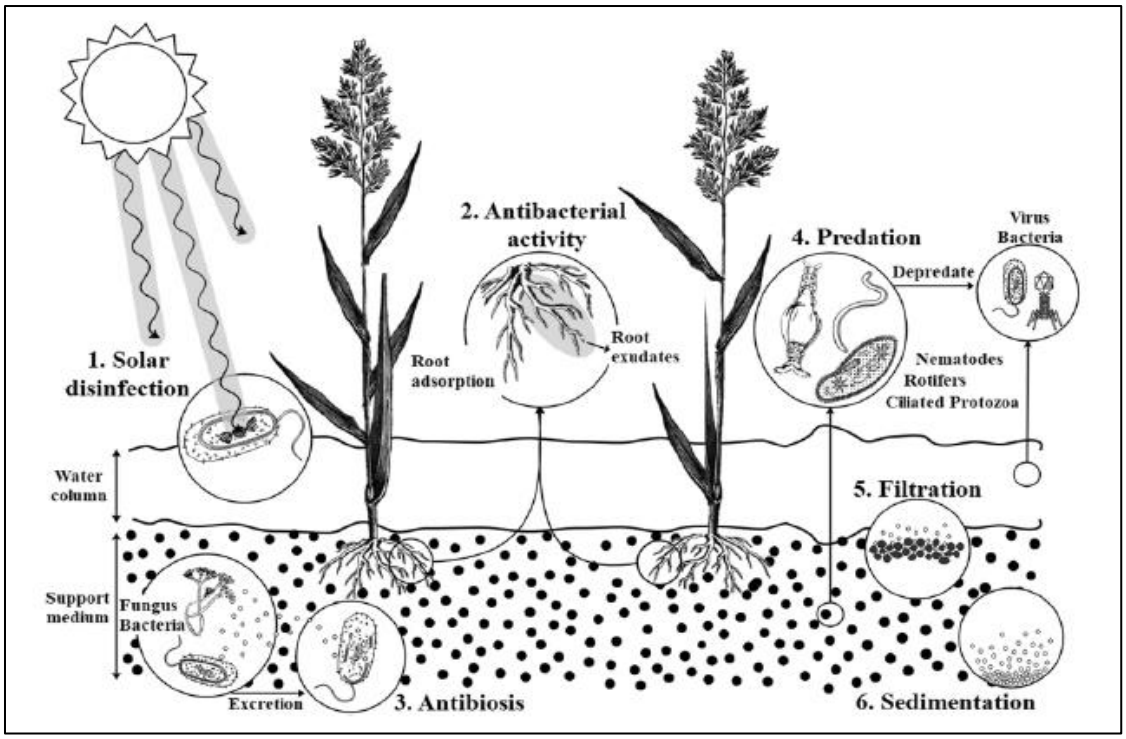


Figure 1: Schematic representation of the removal mechanisms of pathogens in constructed wetlands. Source: Lopez et al., (2019).

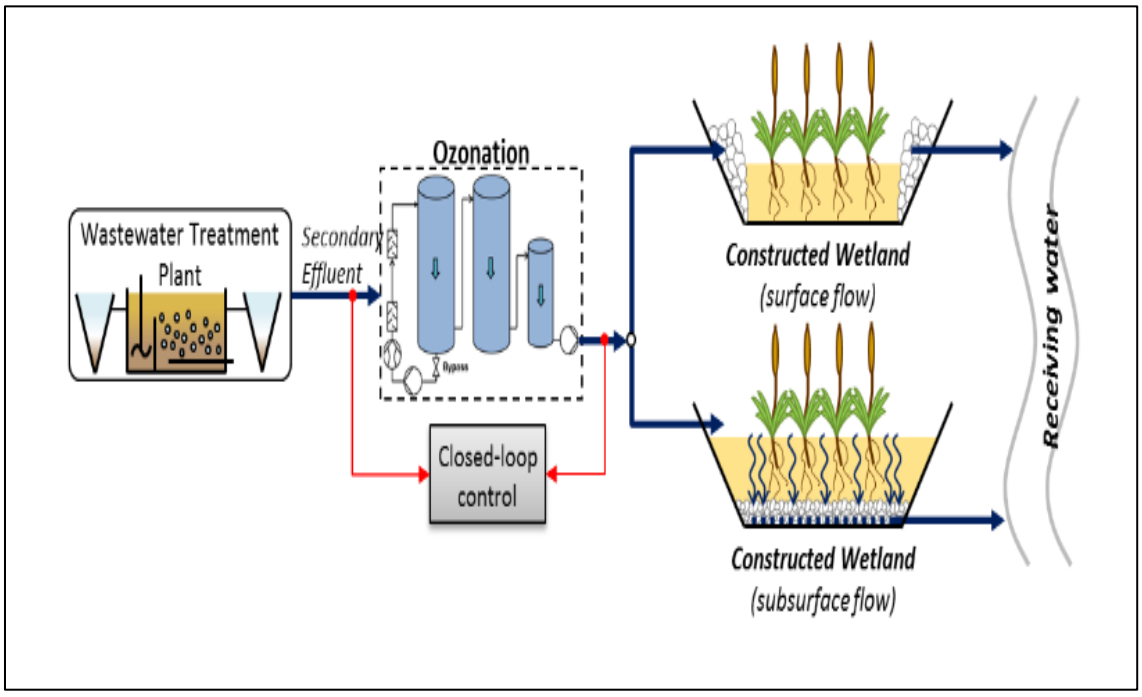


Figure 2: Combination of constructed wetlands and technical treatments. Source: AquaNES-WP3: Constructed wetlands and other natural systems for improved wastewater treatment.

Water can contain various contaminants which can be either microbial or chemical and can cause detrimental health effects if a specific dose is consumed. Specifically, microbial contamination can have harmful effects, even at low levels of exposure (Hrudey & Hrudey, 2004). Conventional wastewater treatment techniques for pathogen removal have proven to be both efficient and reliable. Techniques like activated sludge, or slow sand filtration are usually the preferred methods. In many cases, tertiary treatment is required when strict guidelines for drinking water are applied (Weber & Legge, 2008). If tertiary treatment via UV disinfection or chlorination is not performed, chances are that pathogens can be found in discharge locations from wastewater effluents. Also, livestock and wildlife can contribute to the recontamination of surface water (McAllister & Topp, 2012). Considering the acute effects of contamination and in the context of circular economy and water reuse applications that continuously come to light (Sgroi et al., 2018; Miller, 2006), precautionary measures must be taken by policymakers and water managers. Consequently, the microbial risk is the main concern when it comes to water supply systems. Therefore, water companies in the Netherlands that treat wastewater, are obliged to perform a quantitative microbial risk assessment (QMRA) every three years as described by Bichai & Smeets (2013) Figure 3 demonstrates the organizational composition and the relevant stakeholders of the Dutch water industry involved in the QMRA cycle. Several steps need to be taken from different departments in order to monitor water quality, calculate the risk of infection, evaluate the assessment and finally improve water supply safety.

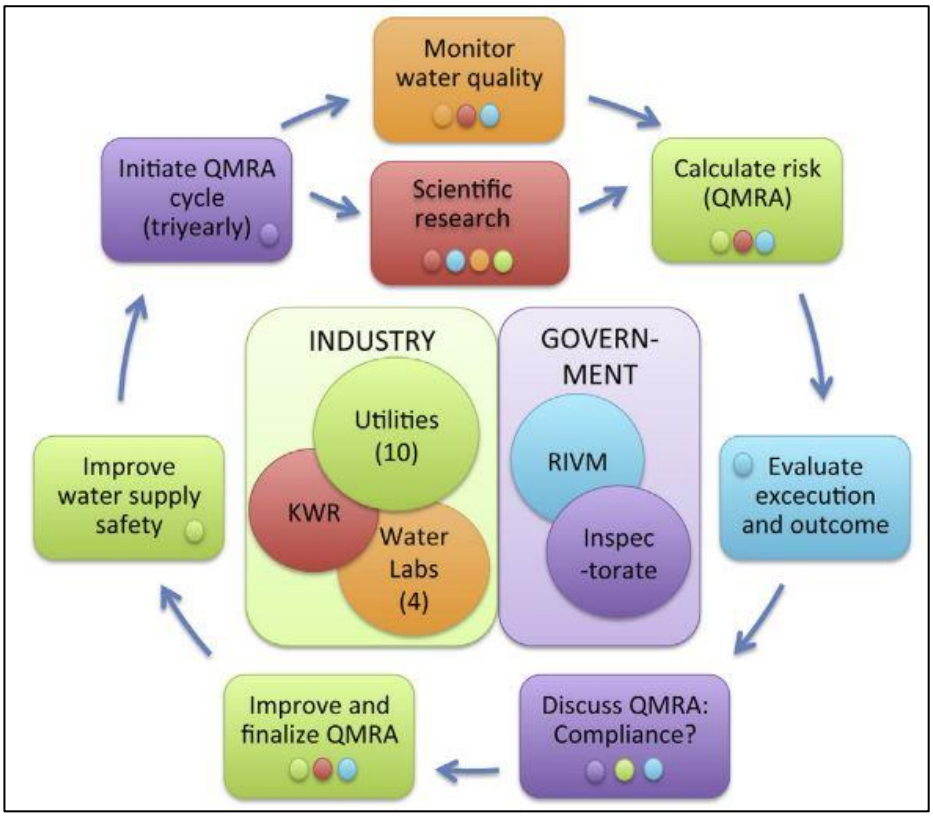


Figure 3: Steps and stakeholders involved in the Dutch legislative QMRA cycle (drinking water). Box color in the cycle corresponds to the main actor in that step. Dots colors correspond to other actors involved. Source: Bichai and Smeets, (2013).

1.2.1 QMRA

The adoption of alternative wastewater treatment methods requires the necessary attention and appropriate scientific tools that will enable any water manager to assess any risk associated with water reuse applications. Moreover, the design of a framework that has as a priority to define the safety of public health through targets, limits, and estimations of risks, is considered a reliable option. Furthermore, a quantitative risk estimate has to be implemented in order to interpret how risks are associated with the corresponding costs. Consequently, a need for a Quantitative Microbial Risk Assessment (QMRA) for the implementation of a multidisciplinary and flexible regulation that meets the challenges of water supply and water scarcity on a global scale is needed (Figure 4). Therefore, when it comes to complying with a health target through the implementation of QMRA the following steps have to be taken (Bichai & Smeets, 2013).

1. Characterization of the microbial contamination from the water source.
2. Assessment of the pathogen removal from various treatment processes based on literature review, modeling or removal of indicator organisms that are easier to monitor.
3. A calculation of risk of infection of a specific population based on the measurement of pathogen dose response and the courses of exposure.
4. Disability-Adjusted Life Years (DALYs[3]) are calculated from the risk of infection of a person, by estimating the possibility of sickness and the potential impact on the person's quality of life from its severity or even death.
5. Finally, the estimated risk is evaluated according to the selected health target that has been set by the regulator.

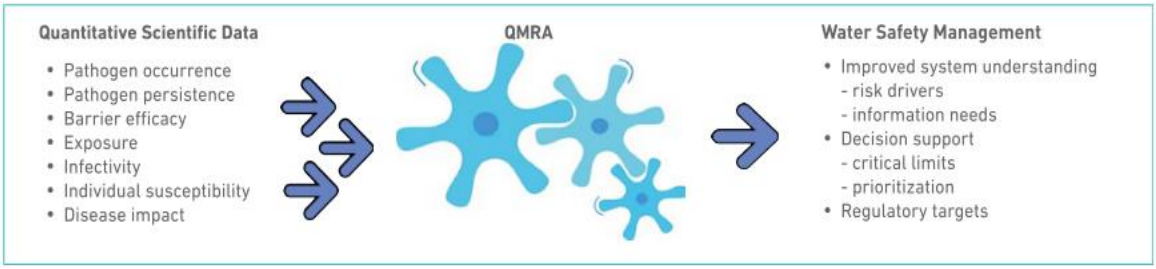


Figure 4: QMRA is a tool for combining quantitative scientific data related to water-related disease pathways to support water safety management. Source: WHO, 2016.

[3] The tolerable recommended value from WHO is 10^{-6} DALYs (WHO, 2006). Although for the Dutch legislative the theoretical risk of infection is 10^{-4} DALYs (Bichai & Smeets, 2013).

1.2.2 Treatment calculator

For the Dutch legislative QMRA of drinking water, RIVM[4] has developed a tool to analyze and conduct QMRA for specific index pathogens named QMRAspot using as input raw data in the proper format (Schijven et al., 2014). As already mentioned, KWR has the ambition of creating a knowledge base regarding pathogen removal from various WW treatment techniques (Smeets, 2017). If no raw data is available as an input in QMRAspot for the implementation of QMRA, KWR provides log removal values based on literature review in the form of another user-friendly computational tool called Watershare Treatment Calculator. This tool provides access to the most up to date information in treatment efficiency and log removal values of various pathogens, so that the end user like municipalities or water companies, does not have to perform literature reviews every time a QMRA is needed. This treatment calculator is based on a predictive model that uses as input a large quantitative data set of pathogen removal values from different case studies and different treatments of wastewater. By implementing a systematic literature review, the end user has access to transparent, evidence-based information, the format of which enables interactive meta-analysis. Therefore, it enables the user to have a complete overview of log removal values depending on the selected pathogen and their corresponding attributes, thus being able to calculate the risk of infection by pathogenic microorganisms in drinking water (Smeets, 2017).

Figure 5 is a screenshot of the computational tool, using UV as an option of treatment and it will serve as an example here since the CW option is missing. The user has access to all available data from the literature review, including the original references of the data. Moreover, there is an option to choose a variety of attributes like type of treatment, pathogen or type of water source. After selecting the desired attributes, the meta-analysis of the data determines the model constants, that lead to the creation of a predictive model. The final outcome provides a fitted line that shows the variation of log removal values regarding a specific UV dose, as well as the equation that is used to calculate the reduction of pathogens based on specific parameters, generated from the model.

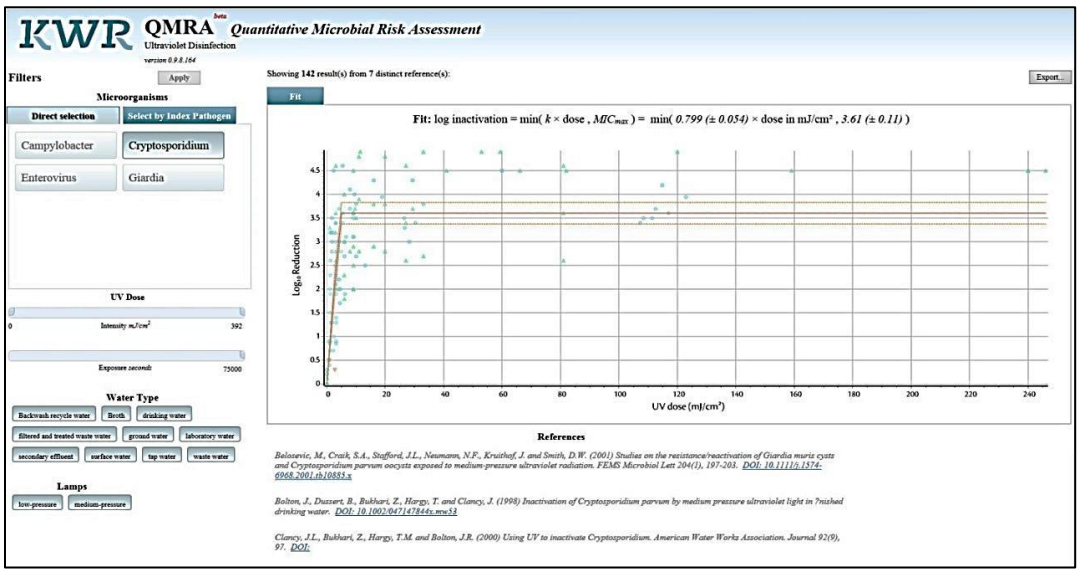


Figure 5: The log removal of Campylobacter using UV treatment through QMRA treatment calculator. Source: Watershare.eu.

[4] The Netherlands National Institute for Public Health and the Environment is a Dutch research institute that is an independent agency of the Dutch Ministry of Health, Welfare and Sport. Source: Wikipedia.

1.3 Problem description and knowledge gap

During the last twenty years, a lot of scientific papers and books have been published regarding the characteristics of CWs and their pathogen removal performance including Vymazal et al. (2006), Kadlec & Wallace (2008) and Vymazal & Kröpfelová (2008). What is common in these papers is the complexity and heterogeneity that characterizes these natural systems. The contaminant removal processes include physical, chemical and biological processes that can take place simultaneously and therefore it can be quite challenging to specify the main factors that are responsible for the removal of pathogens and draw valid conclusions. Moreover, even though the treatment calculator developed by KWR is a well-documented tool and there are various options for the selection of treatment like UV, ozonation or slow sand filtration, the option of a CW treatment is still missing. That is because up until now, a comprehensive and detailed dataset which includes the log reduction of pathogens from natural systems under the influence of specific parameters, and the development of a predictive model are missing. Therefore, the aim of this paper is to cover the knowledge gap by contributing to the efforts of KWR for the development of a knowledge base for pathogen removal through constructed wetlands.

1.4 Research objective and research question

As already mentioned, constructed wetlands are complex systems that have raised interest in the scientific community for many years now. Therefore, a more generic approach to the removal efficiency of these systems could possibly shed more light in the ''black box'' that is constructed wetlands. The objective of this paper is to contribute to the efforts of KWR for the development of a knowledge base for pathogen removal through constructed wetlands. In order to do that, a complete dataset with all the necessary information regarding pathogen removal will be created first. Furthermore, this dataset can be incorporated into the treatment calculator by the means of a predictive model. The model will give log removal values of pathogens based on specific parameters. Therefore, in order to cover the knowledge gap, the research question can be described as:

''*Quantifying the log reduction of pathogenic microorganisms by constructed wetlands as a basis for Quantitative Microbial Risk Assessment (QMRA) of water reuse applications''.*

2. Constructed wetlands

Constructed wetlands are alternative engineered systems that are based on the use of emerging plants such as bulrush or reeds for the purification of wastewater (Hunt et al., 2003; Kadlec & Wallace, 2009; Vymazal, 2010). The definition of a constructed wetland given by Nuttall et al. (1998), is described as a man-made system designed to replicate the operation of a natural wetland. The purpose of this system can either be for treating wastewater and stormwater or for different reasons like recreational activities or the implementation of nature-based solutions. Furthermore, they have the same capacity and functionality as natural wetlands, but without the limitations that are related to the disposal or regulation of effluents to natural ecosystems, site selection, flexibility in sizing and control of retention time (Vymazal & Kröpfelová, 2006). Therefore, these environmental-friendly treatment systems are considered to be a simple alternative method of treating wastewater that contributes to the applications of water reuse and water reclamation (Sultana, 2014).

Compared with conventional WWTPs, constructed wetlands are more attractive and easy to operate, with great applicability in small rural communities and especially in developing countries, (Jamieson et al., 2007; Kivaisi, 2001). Modern constructed wetlands are designed in such a way as to emphasize the features of wetland ecosystems. Various physical, chemical and biological processes take place for the purification of wastewater including sedimentation, precipitation, adsorption, absorption by plant tissues, and microbial transformations as can be seen in Figure 6 (Vymazal & Kröpfelová, 2006). Moreover, the removal of pollutants in CWs depends on the design characteristics of the wetland, the microbial community, as well as the various plant species that have been planted (Ibekwe et al., 2003).

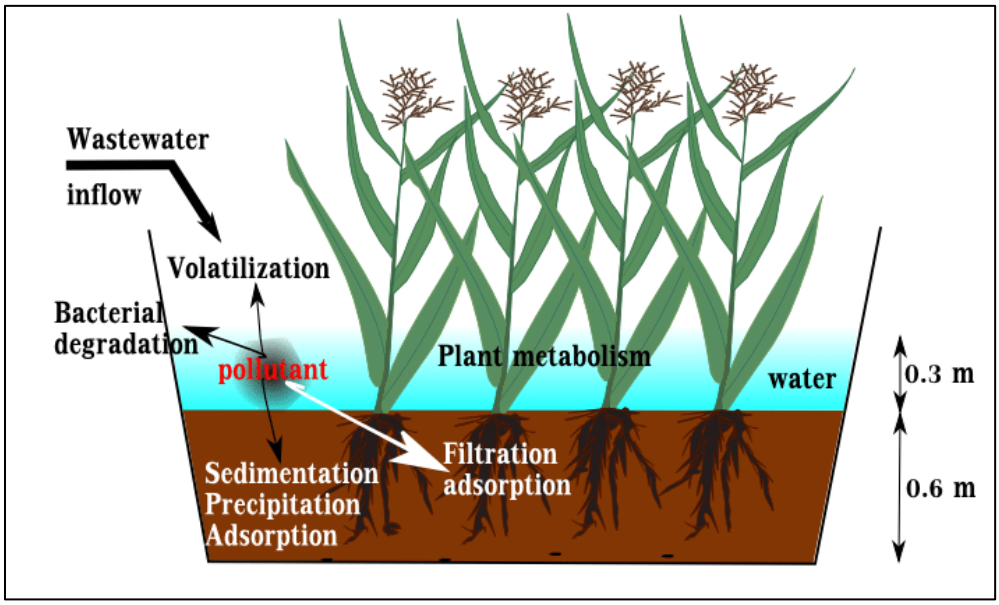


Figure 6: Processes of a constructed wetland. Source: Truijen and van der Heijden, (2013).

2.1 Hydraulic Retention Time and Hydraulic Loading Rate

2.1.1 Hydraulic Retention Time

Hydraulic Retention Time (HRT) is defined as the time a molecule stays inside a wetland on average, from the time it enters to the time it exits (Dotro et al., 2017). It is closely connected to the inactivation of pathogens since the more time a pathogen spends in the CW, the longer the contact time with biofilms and predators and therefore, the better the removal (Alexandros & Akratos, 2016). A lot of studies have concluded that there is a positive correlation between the microbial inactivation and HRT (Vymazal, 2005a; Tanner et al., 1995; Garcia et al., 2003).

2.1.2 Hydraulic Loading Rate

Hydraulic Loading Rate is another design factor that is correlated with pathogen removal (Tanner et al., 1995; Arias et al., 2003). Moreover, HLR is connected directly with HRT since they are inversely proportional. Subsequently, prolonged HRTs lead to a decrease in influent flow, thus enabling wastewater to flow slowly around substrate particles, leading to enhanced removal (Olsson, 2011).

2.2 Types of constructed wetlands

Constructed wetlands can be classified in three main categories depending on their water flow regime and these are:

- Free water surface (FWS) constructed wetlands that have areas of open water and behave similarly to natural wetlands.
- Subsurface horizontal flow constructed wetlands (SSHFCW), which typically consist of a gravel bed covered with wetland vegetation. The water flows constantly below the surface of the bed in a horizontal direction.
- Vertical flow constructed wetlands (VFCW), where the main difference compared with SSHFCWs is that water percolates vertically through sand or a gravel bed planted with vegetation. The influent enters through perforated pipes which are distributed over the surface in the form of a grid.

Each of these three categories demonstrates different layouts, media, efficiency and flow patterns (Kadlec & Wallace 2008).

2.2.1 Free water surface constructed wetlands

This type of CW consist of open water which flows horizontally through floating vegetation and emergent plant attached to parallel basins, canals or ditches. The flow regime can be regulated with the use of dikes and levees. During the passage of water through the surface, physical, biological, and chemical processes start to occur with sedimentation, filtration, oxidation, adsorption, and precipitation being the most important. The schematic representation of this type is shown in Figure 7 (Kadlec and Wallace, 2009).

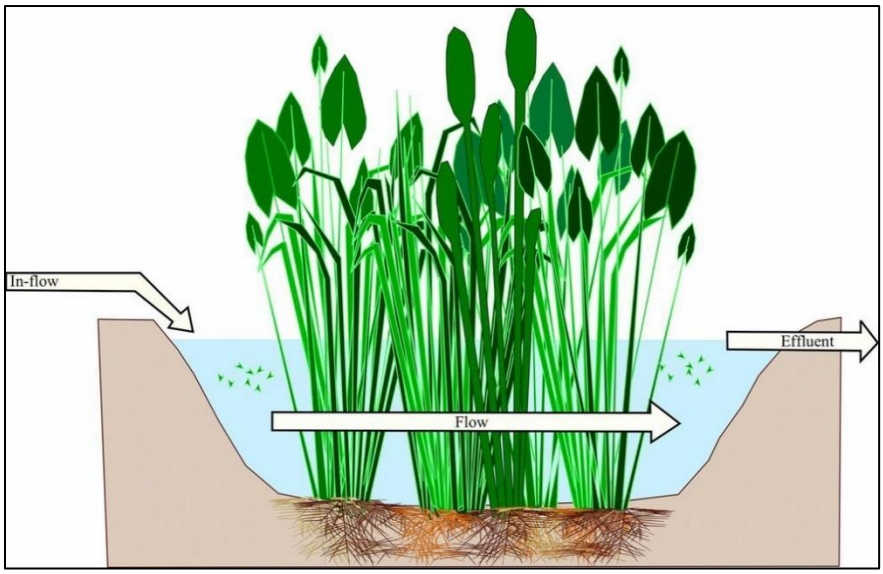


Figure 7: A schematic representation of a FWS CW. Source: White, (2013).

A typical FWS CW usually contains dense emerging macrophytes which are covering more than 50% of the surface. In many cases, indigenous plant species can appear in such wetlands mainly due to open space (Kadlec, 1995). Furthermore, the water flows horizontally through the surface with an average depth of 10-15 cm while the depth of the basin is between 40 and 60 cm (Vymazal, 2010).

FWS CWs are particularly effective in removing total suspended solids (TSS) through sedimentation and filtration processes (Kadlec & Knight, 1996). The influent that contains particulate and dissolved pollutants, diffuse over the surface and percolates through emerging or submerged vegetation. Moreover, these systems are efficient in Nitrogen removal through the processes of nitrification and denitrification (Kadlec et al., 2000).

Regarding their performance in the removal of pollutants, they are particularly effective and resilient through various weather conditions. However, low temperatures can create ice formation along the surface which can inhibit the hydraulic operation, since the rates of various removal processes can be significantly reduced (nitrogen conversion processes). This dysfunction can be attributed to the fact that the transfer of oxygen from the atmosphere to the body of water is reduced due to the creation of an ice layer on top of the surface. Subsequently, this results in reduced efficiency of oxygen-dependent processes. On the contrary, the removal of suspended solids is more efficient in cold than in summer months. Therefore, a common tactic is to store water during the winter months and then treat it during summer for maximum efficiency of the CW.

One of the main disadvantages of FWS CWs originates from their own function, as they imitate to a large extent to natural wetlands since a great variety of birds, insects, mollusks, and mammals find shelter in these areas (Kadlec & Knight, 1996). Therefore, the potential risk of recontamination from wildlife and the acute effects of human exposure to pathogenic microorganisms, make FWS unsuitable for secondary treatment. Instead, they find great usage when they are attached to a secondary or tertiary treatment for polishing of the effluent. Moreover, another disadvantage is the fact that they require an extensive land area to function properly (Kadlec & Wallace 2008).

2.2.2 Subsurface horizontal flow constructed wetlands

SSHFCWs are the most widespread subsurface wetland systems around the world. They consist of wetland vegetation attached on a bed that usually consists of gravel or sand. Typically, they comprised of an inlet zone, a clay or synthetic material, a suitable inert filler material, and an outlet piping with a water level control. The wastewater-which usually is primary effluent- flows horizontally beneath the surface of the media and percolates slowly through the rooted zone of the vegetation until it reaches the outlet area where usually it is collected and distributed through the level control tank (Kadlec & Wallace 2008). During its movement towards the outlet zone of the wetland, water meets a series of aerobic, anoxic and anaerobic zones that most of the time take place in the rhizosphere (Brix, 1987; Cooper et al., 1996). There, the bacteria that are attached to the plant's roots but also to the soil initiate aerobic and anaerobic processes of biodegradation. For the aerobic processes, the oxygen supply is coming directly from the atmosphere or from the plant's roots. Since as already mentioned, the wastewater is ''sealed'' beneath the surface, the transport of oxygen coming from the atmosphere is not enough to meet the needs of this process in the rhizosphere. Therefore, the anoxic and anaerobic biodegradation outweighs the aerobic process and it plays an important role in the purification of water in subsurface horizontal wetlands (Brix, 1990; Vymazal & Kröpfelová, 2006).

One of the biggest advantages of SSHFCW is the fact that because the water is flowing beneath the surface, it is not exposed to the environment. Therefore, there is limited risk associated with wildlife recontamination and exposure to pathogens. Additionally, because of the insulation of the surface, SSHFCW can operate better than FWS CW under low temperatures. On the contrary, clogging of the media substrate is one of the main drawbacks of these systems. Compared to the FWS systems, they are more expensive regarding maintenance and they are the number one option for secondary treatment in small communities and single-family homes (Wallace & Knight, 2006; Cooper et al., 1996). A schematic representation of a typical SSHFCW can be found in Figure 8.

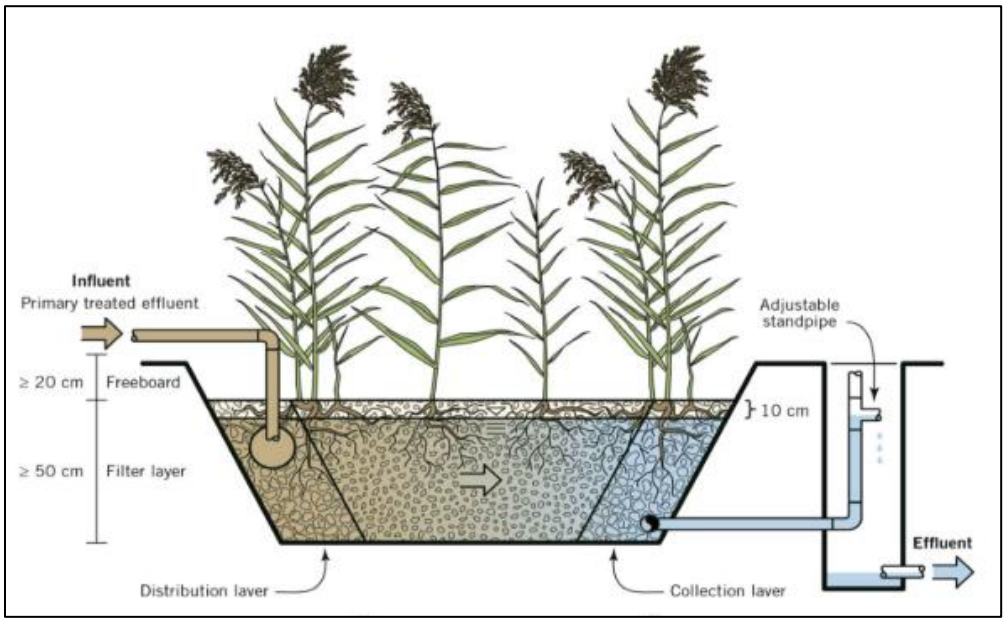


Figure 8: Schematic representation of a SSHFCW. Source: Dotro et al., (2017).

2.2.3 Vertical flow constructed wetlands

Vertical flow (VF) constructed wetlands-along with several modifications that exist-are particularly popular in Europe as they require smaller areas for their construction compared with the SSHFCW. These systems resemble the dosage scheme that is used in sand filters mainly because of the surface flooding of the bed (pulse loading) during a single-pass configuration (Kadlec & Wallace 2008). The main difference regarding the flow regime is the vertical flow of the wastewater penetrating through the soil layers of a basin. The water enters the system through perforated conduits which are distributed through the surface of the wetland in the form of a network. The ducts carry all along specific-sized holes that through them, a constant flow penetrates the surface, aiming to a uniform discharge as can be seen in Figure 9 (Kadlec and Knight, 1996; Reed et al., 1995).

VF wetlands have been used the last 25 years in various ways and modifications, depending on the desired quality of the effluent. Indicatively, in North America, they have been used as gravel filters with recirculation of water (Lemon et al., 1996), or as "tidal flow" systems in order to treat high-strength waste and oxidize ammonia (Behrends et al., 1996; Austin & Lohan, 2005). Moreover, there are "upflow'' and "downflow" systems of VF which are related to the direction of the flow of the wastewater. The former is preferred in situations where the oxygen transport has to be reduced (Kassenga et al., 2004), and the latter in cases where the oxygen transfer is imperative, in order to produce a nitrified effluent. This effective and simple technology, which is initiated by Dr. Kathe Seidel in 1960, can be combined with both FWS and SSHF wetlands to create treatment chains of nitrification-denitrification processes (Cooper et al., 1999). Furthermore, because of their unique ability to oxidize ammonia, VF CWs are especially used in situations where the levels of ammonia in the influent are too high.

One of the few drawbacks on the performance of these systems is clogging of the media filter (Langergraber et al., 2003; Winter & Goetz, 2003). Therefore, the selection of the most suitable media filter, the regulation of hydraulic loading rate and the uniform distribution of wastewater through the vegetated bed, are some of the most important things one has to be aware of.

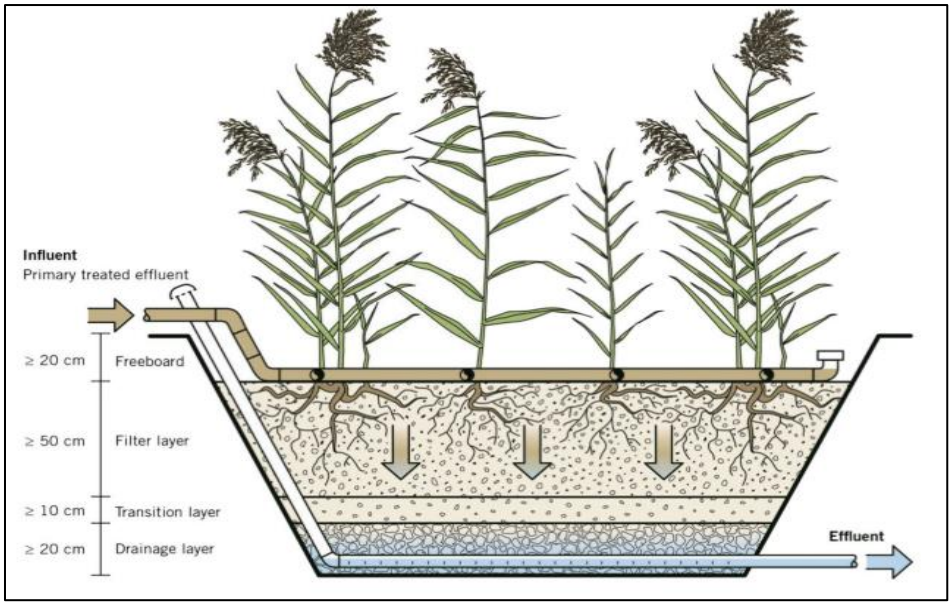


Figure 9: Schematic representation of a VFCW. Source: Dotro et al., (2017).

2.2.4 cNES applications

Many studies have shown that CWs have excellent flexibility and applicability when it comes to combinations with conventional technologies. Camacho et al. (2014), presented the ability of a HFCW and VFCW to work as a microbial fuel cell for wastewater treatment and electricity production. Moreover, a hybrid system[5] was combined with electrochemical oxidation for the treatment of unconventional surface water of the Wenyu river in China. The results were encouraging since the oxidation unit lead to an improvement in the average value of BOD_5/TN (Biological Oxygen Demand after 5 days/Total Nitrogen) coming into the CW, which enabled the latter to biodegrade the surface water under better conditions. Therefore, the combinations of the two methods paved the way for different approaches when it comes to treating unconventional surface water (Wang et al., 2014). Lastly, VFCW can function as biofilters and enhance the quality of a secondary treatment effluent. Sharma & Brighu (2014), showed that the use of CW as a WWTP polishing method can increase the removal efficacy of organic matter and Total Kjeldahl Nitrogen (TKN).

Another interesting effort regarding the combination of constructed wetlands with engineered systems is the AquaNES project which receives funding from the European Union's Horizon 2020 research and innovation program. The project takes place throughout the world and demonstrates the benefits of combining constructed wetlands as a pre-or post-treatment method with conventional engineered methods such as ozonation and bioreactor systems for wastewater treatment. This initiative has as an objective to showcase the usage of CWs as WWTP polishing and how they can influence the removal of micropollutants and pathogens. More information about AquaNES can be found on their website[6].

[5] Hybrid system is a combination of two or more types of CW (Vymazal, 2005).

[6] http://www.aquanes-h2020.eu

2.3 Pathogens in wastewater

Pathogenic microorganisms can be found in untreated domestic and municipal sewage but also in runoff waters as a result of animal (re)contamination. Regarding their size, they vary from infinitesimal viruses to parasitic worms that can be identified with a naked eye and their presence –as well as their concentration- are important factors in water quality. Table 1 demonstrates a classification of human pathogens (with domestic wastewater origin) in five groups namely: Viruses, Bacteria, Fungi, Protozoans and Helminths, along with their respective illnesses (Kadlec & Wallace 2008).

Table 1: Classification of human pathogens and their respective illnesses. Source: Modified from Kadlec & Wallace, (2008).

	Pathogen	Illness
Viruses	Adenovirus (31 types)	Respiratory disease
	Enterovirus (67 types)	Diarrhea, respiratory disease, polio
	Hepatitis A	Infectious hepatitis
	Norwalk agent	Gastroenteritis
	Rotavirus	Diarrhea
	Reovirus	Gastroenteritis
	HIV	AIDS
Bacteria	Campylobacter *jejuni*	Diarrhea
	Escherichia *coli*	Diarrhea
	Legionella *pneumophila*	Fever, respiratory tract infections
	Leptospira (150 spp.)	Leptospirosis
	Salmonella *typhi*	Typhoid fever
	Salmonella (~1,700 spp.)	Salmonellosis
	Shigella (4 spp.)	Diarrhea, dysentery
	Vibrio spp.	Cholera, diarrhea
	Yersinia spp.	Yersiniosis
Fungi	Aspergillus fumigatus	Aspergillosis
	Candida albicans	Fungal infections
Protozoans	Balantidium coli	Diarrhea, dysentery
	Cryptosporidium parvum	Diarrhea
	Entamoeba histolytica	Diarrhea, dysentery
	Giardia lamblia	Diarrhea
Helminths	Ascaris lumbricoides	Roundworm
	Clonorchis sinensis	Bile duct infection
	Diphyllobothrium latum	Fish tapeworm
	Enterobius vericularis	Pinworm
	Fasciola hepatica	Liver fluke
	Fasciolopsis buski	Intestinal fluke
	Hymenolepis nana	Dwarf tapeworm
	Necator americanus	Hookworm
	Opisthorchis spp.	Bile duct infection
	Schistosoma spp.	Schistosomiasis
	Taenia spp.	Tapeworm
	Trichuris trichura	Whipworm

2.3.1 Viruses

Viruses are submicroscopic non-living agents (20-200 nm) that are covered with a protein case for the protection of their genetic material. They cannot divide and replicate alone, but they can infect the living cells of host organisms and produce multiple copies of themselves reaching great numbers of population at the expense of the host organism (Weber & Legge, 2008).

2.3.2 Bacteria

The biggest group of microorganisms that can be found in human feces is bacteria, with concentrations of about 10^{11} organisms per gram and a size range of 0.1-5 μm (Leclerc et al., 1977). Furthermore, it is known that the majority of these prokaryotic organisms can live symbiotically with their hosts, although there are some species that can be extremely harmful to humans. Bacteria when they are found in their normal growth cycle, they can be removed or inactivated easily in wastewater (Weber & Legge, 2008).

2.3.3 Protozoa and Helminths

The next two groups are Protozoa and Helminths and they belong to the category of human parasites that come from wastewater-related infections. Two of the most common protozoan parasites that can cause diarrhea to an infected person are Giardia lamblia and Entamoeba histolytica. Moreover, Cryptosporidium belongs to the group of pathogenic water-born protozoan that it is very difficult to remove (Slifko et al., 2000). Regarding helminths, there is no taxonomy since the parasitic worm is a term that includes many different species of worms (Weber & Legge, 2008).

2.3.4 Animals as a source of microbial pathogens

Besides humans, another source of pathogenic organisms is the Animal Kingdom. Over the last three decades, it is estimated that almost 75% of emerging human diseases originate from animal pathogens due to zoonosis[7] (Brown, 2004). Livestock like cows, sheep, pig, and poultry produce high concentrations of bacteria from their feces. Maier et al., (2009), recorded concentrations of 10^5-10^7 fecal coliforms per gram and 10^6-10^8 fecal streptococci per gram in their book. Also, beavers, rodents and other warm-blooded animals can produce high concentrations of enteric pathogens. In particular, beavers are closely linked to the transmission of Giardia (Erlandsen, 1994).

2.3.5 Indicators and pathogens

The enumeration of human pathogens coming from wastewater and/or surface waters can be an expensive and time-consuming process. Therefore, numerous methods have been developed that first quantify groups of indicator organisms that are easy and inexpensive to monitor, and then correlate them with their respective index pathogenic organisms (Kadlec & Wallace 2008). Fecal and process indicators are the two common groups that are used for the enumeration of pathogens. Fecal indicators are a group of microorganisms that point the presence of fecal contamination presuming the existence of pathogens like E. *Coli*. Process indicators are organisms that indicate the effectiveness of a process like total coliforms (Fewtrell & Bartram, 2001).

[7] Zoonosis refers to animal source-infectious diseases caused by bacteria, viruses, and parasites that move into human hosts (Brown, 2004).

2.3.5.1 Coliforms

Although no indicator organism is a perfect indicator, the most popular group indicating fecal contamination, is that of coliform bacteria. This group consists of rod-shaped, strain gram-negative bacteria who live in optionally anaerobic conditions, and are capable of fermenting lactose within 48 hours at 35°C (Kadlec & Knight, 1996). Moreover, because some coliforms can be found in different environments and sources, a differentiation must be made for coliforms of fecal origin. Therefore, fecal coliforms (FC) are a distinct group of indicators that differ from total coliforms (TC) because they can ferment lactose with gas production faster (24 hours) and at a higher temperature (45°C) (Kadlec & Wallace 2008). One major drawback of TC is the fact that many of the enumerated bacteria can be found also in surface waters. Many bacterial species that are included in the TC indicator group do not come from human or warm-blooded animals. Therefore, TC cannot be a reliable indicator of human fecal contamination (Weber & Legge, 2008).

2.3.5.2 Escherichia Coli

A more specific indicator is Escherichia *Coli*, and it is commonly preferred for enumeration because it can easily be isolated. There are several different strains of this indicator that are actually pathogenic and can be extremely harmful to people. Nevertheless, E. *Coli* cannot be an exclusive indicator of human fecal contamination because it exists in other warm-blooded animals as well (Kadlec & Wallace 2008; Weber & Legge, 2008).

2.3.5.3 Fecal Streptococcus

Fecal streptococcus is another group that is commonly used as an indicator because it can be found in both human and warm-blooded animals like birds and mammals. FS usually exist in waters that are contaminated with feces, but they do not seem to replicate in natural or polluted waters. This indicator group is usually preferred from FC because they show greater resistance and have a longer lifespan (Kadlec & Knight, 1996).

2.3.5.4 Bacteriophages

The bacterial indicators group are considered to be an unsuitable indicator of viral contamination, mainly because viruses, in general, show greater persistence to chlorination and environmental deactivation processes (Gersberg et al., 1987). Therefore, for viral indication in wetland systems, the most common indicator is bacteriophages (a type of virus that infects bacteria) and particularly coliphage MS-2 (Gersberg et al., 1987; Kadlec & Knight, 1996). MS-2 phages are easier and faster to enumerate compared to target pathogenic viruses. They have almost the same size as enteroviruses[8] as well as they seem to be more resilient to ultraviolet light (UV), heat and disinfection compared to the majority of viruses.

[8] *Enterovirus* is a genus of positive-sense single-stranded RNA viruses associated with several human and mammalian diseases. Enteroviruses are named by their transmission-route through the intestine (van Regenmortel et al., 2000).

2.4 Pathogen removal processes and contributors

When it comes to wastewater treatment from constructed wetlands, there are various mechanisms of pathogen removal that take place simultaneously and sometimes act in combination. While on the conventional wastewater treatment systems the reduction processes take place in separate unit operations that are intended for a specific use, thus giving a maximum performance; CWs perform multiple processes at the same time in one or two reactors (Dotro et al., 2017).

Constructed wetlands are a well-established biofilter option that uses a series of physical, chemical and biological mechanisms for the reduction of pathogenic microorganisms of human origin. The processes to be described in this chapter consist of physical processes like filtration and sedimentation, chemical such as solar radiation (UV), and biological like predation and natural die-off. Table 2 demonstrates the main mechanisms for pollutant and pathogen removal in CWs. For the purposes of this paper, the focus will be on the pathogen removal processes that take place during the wastewater passage (Vymazal & Kröpfelová, 2008).

Table 2: Main mechanisms for pollutant and pathogen removal in treatment wetlands. Source: Modified from Dotro et al., (2017).

Parameter	Main removal mechanisms
Suspended solids	Sedimentation, filtration
Organic matter	Sedimentation and filtration for the removal of particulate organic matter, biological degradation (aerobic and/or anaerobic) for the removal of dissolved organic matter
Nitrogen	Ammonification and subsequent nitrification and denitrification, plant uptake and export through biomass harvesting
Phosphorus	Adsorption-precipitation reactions driven by filter media properties, plant uptake and export through biomass harvesting
Pathogens	Sedimentation, filtration, natural die-off, predation, solar radiation

2.4.1 The role of plants and temperature in the removal process

One major contributor to the removal processes is the wetland vegetation. Particularly, the various types and sizes of roots and rhizomes, create a suitable environment for the emergence of microbial biofilm. This, in turn, results in an increase in biological activity per unit area, indicating the superiority of these systems over open water systems such as ponds. In addition, with their root structure, they hinder part of the flow of water, thus minimizing hydraulic short circuits (Dotro et al., 2017).

The temperature along with seasonal effects are another major contributor to pathogen removal in CWs. It is well known that the majority of human pathogens are most active and function efficiently at temperatures similar to that of the internal human body (37 °C). Consequently, there is a valid assumption that pathogens at low temperatures are inactivated or forced to change to a spore or oocyst state (Webber & Legge, 2008). The

increased inactivation rates enhance the treatment of pathogens, although it is extremely difficult to remove pathogens when they surround themselves with durable coats of protein that allow them to survive in hostile environmental conditions. On the other hand, although the increased temperature can enhance the activity of nonpathogenic organisms like grazing protozoa, thus leading to increased predation rates, it can also increase the pathogenic activity in the wetland. Therefore, there is some skepticism as to what extent increased temperature is positively correlated with increase pathogen removal (Alexandros & Akratos, 2016; Ulrich et al., 2005; Arias et al., 2003).

2.4.2 Filtration and Sedimentation

The filtration mechanism is a process where the biological film that is formed in the media, creates "sticky traps" for particles and purifies the influent wastewater (Collivignarelli et al., 2018). These biofilms are extremely efficient in trapping sufficient concentrations of organisms since it is well known that bacteria have the ability to aggregate, forming shapeless and porous clusters with various sizes. (Flood, 2000; Stott & Tanner, 2005). Moreover, a bigger proportion of submerged surfaces, light exposure, and optimal plant density usually lead to higher efficiency of such biofilms (Kadlec & Wallace 2008).

Sedimentation is another proven mechanism of pathogen removal where the size of the particles and their respective density are the main regulators of its performance. Moreover, the agglomeration of bacteria on the media grains and sediments leads to a creation of a ''pathogen sink'' in the bottom layers of CWs (Alexandros & Akratos, 2016). Karim et al., (2004) noted that concentrations of Giardia cysts and Cryptosporidium oocysts were higher (1-3 orders of magnitude) in the sediment of a FWS CW than in the water column, concluding that sedimentation influences pathogen removal.

2.4.3 Solar radiation

Many studies have shown that solar radiation can be an effective factor in coliform bacteria removal, especially when there are high dissolved oxygen concentrations with water quality factors like optical absorbance and suspended solids content influencing the effectiveness of radiation. Moreover, coliform bacteria can be inactivated when they absorb wavelengths around mid-UV (290-320 nm) and near-UV (320-400 nm).

The downside of solar radiation is the fact that the majority of microorganisms cannot be found in a dispersed phase but usually, they are accumulated with wastewater particles that act as a shield to the UV wavelengths and results in reduced inactivation rates. Furthermore, the efficiency of solar disinfection depends on the amount of sunlight that reaches and penetrates the water. Therefore, in dense vegetation surfaces of wetlands, the sunlight gets intercepted resulting in reduced efficacy of solar disinfection. (Kadlec & Wallace 2008).

2.4.4 Predation

Predation in constructed wetlands is the biological process where pathogenic bacteria are eliminated by other organisms like protozoa, and bacteriophages (Wand et al., 2007; Kuschk et al., 2012). This mechanism is dependent on the key attributes of both the prey (such as population density and existing species) and the predator (morphology and physiology of the microorganism) (Shapiro et al., 2010; Alexandros & Akratos, 2016). Predation activities and particularly protozoan predation are cited as a main removal mechanism of bacteria in CWs in many studies (Wand et al., 2007; Stott et al., 2001).

2.4.5 Natural die-off

One of the main and most cited mechanisms that are responsible for pathogen removal is natural die-off. This process is closely related to parameters like the Hydraulic Retention Time of the system (HRT), predation conditions and famine of microorganisms (Green et al., 1997; Decamp & Warren, 1998; Alexandros & Akratos, 2016). Boutilier et al., (2009), concluded that natural die-off was the most important factor when it comes to coliform bacteria removal in a FWS CW. Moreover, Karim et al., (2004), in their study reported besides sedimentation, the natural die-off was also an important removal mechanism. More specifically, it was noted that natural die-off rates for bacteria were higher in the water column than in the sediment.

3. Methodology

To answer the research question, a systematic literature review, data extraction and meta-analysis and a selection of the most suitable models will be performed. The outcome of these steps will include (Figure 10):

- A complete database of case studies with their respective removal performance
- A model that can predict log removal of pathogens based on the most influential parameters
- A storyboard for the Graphical User Interface (GUI) of the treatment calculator.

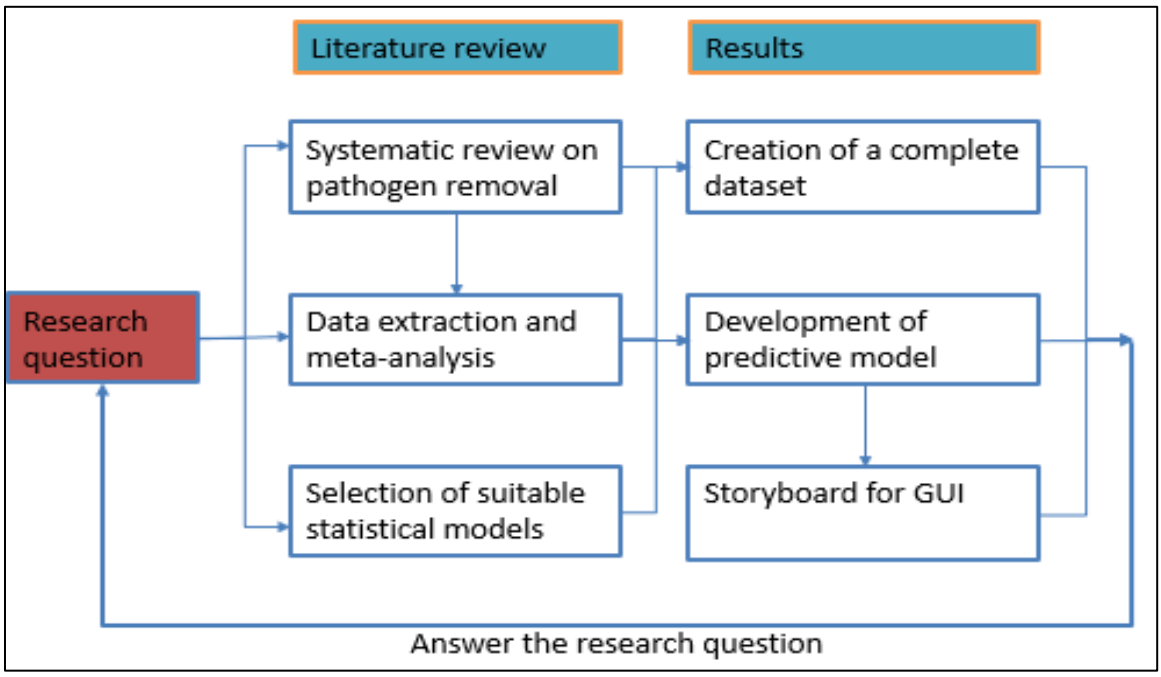


Figure 10: Schematic representation of the methodology in order to answer the research question

3.1 Systematic review of case studies

To identify relevant case studies regarding pathogen removal from constructed wetlands, a systematic review of the literature was performed adopting the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al., 2009). These guidelines were developed by an international group that included experienced writers and methodologists and describe a minimum set of evidence-oriented objects when reporting in systematic reviews and meta-analyses. The preferred search engines, ''Scopus'' and ''Pubmed'' were used. The combination of keywords, as well as the results from the databases regarding the total number of papers, are shown in Appendix 1. Moreover, the flow diagram in Figure 11, demonstrates the search process of the systematic review, along with the exclusion reasons for papers.

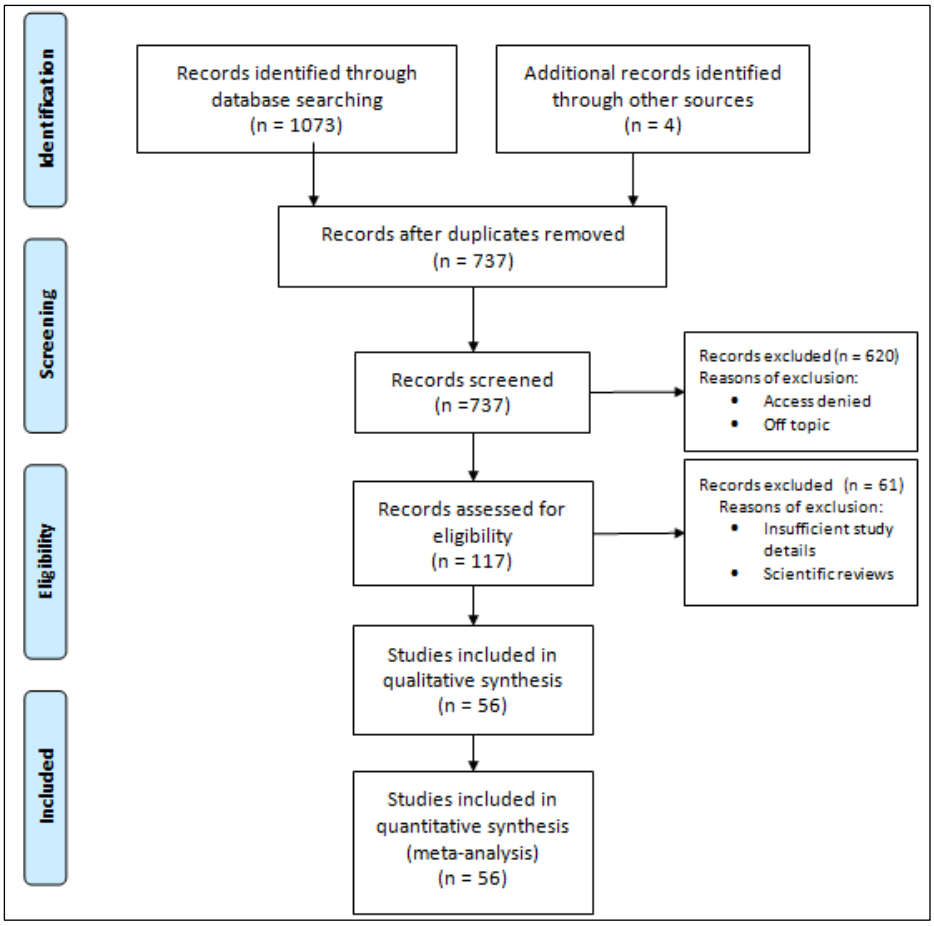


Figure 11: Flow diagram on the selection of studies adopted by the PRISMA guidelines. Source: Liberati et al., (2009).

3.2 Data extraction process

During the extraction process, a classification between the case studies under consideration was made in order to ease up the process of data extraction. The classes "Include 1'', "Include 2'' and ''Include 3'' were created, with the numbering system indicating the degree of difficulty on extracting data as well as the quantity of descriptive information of case studies. Therefore, the "Include 1'' and "Include 2'' categories included case studies were the extraction of data was easy and straightforward, in the form of tables or graphs like in Figure 12 and 13. The "Include 3'' category was excluded since pathogen removal data wasn't clearly reported or additional data on conditions was missing (Figure 11).

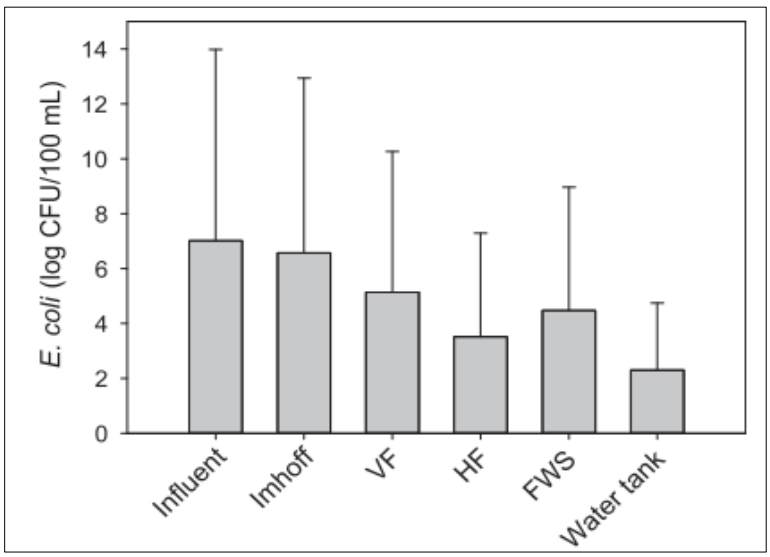


Figure 12: Example of reported data. Average concentration and standard deviations of E. coli along the different treatment system units. Source: Avila et al., (2013).

Bed	Hydraulic load		*E. coli*			
	HLR[a] (L/m² d)	*n*HRT[b] (days)	Influent concentration (MPN/100 mL)	Effluent concentration (MPN/100 mL)	Log_{10} reduction (MPN/100 mL)	Areal load removal rate (MPN/m² d)
H25	18	5.2	$7.4 \times 10^6 \pm 4.1 \times 10^6$ (42)	$2.7 \times 10^5 \pm 2.1 \times 10^5$ (42)	1.4 ± 0.41	$1.3 \times 10^9 \pm 7.5 \times 10^8$
H25p	18	5.6	$7.4 \times 10^6 \pm 4.1 \times 10^6$ (42)	$1.8 \times 10^5 \pm 1.9 \times 10^5$ (42)	1.5 ± 0.41	$1.3 \times 10^9 \pm 7.6 \times 10^8$
H50	36	5.2	$7.4 \times 10^6 \pm 4.1 \times 10^6$ (42)	$3.7 \times 10^5 \pm 2.5 \times 10^5$ (42)	1.3 ± 0.37	$2.5 \times 10^9 \pm 1.5 \times 10^9$
H50p	36	5.4	$7.4 \times 10^6 \pm 4.1 \times 10^6$ (42)	$3.0 \times 10^5 \pm 1.9 \times 10^5$ (42)	1.3 ± 0.37	$2.5 \times 10^9 \pm 1.5 \times 10^9$
HA	131	2.9	$7.6 \times 10^6 \pm 3.2 \times 10^6$ (26)	$6.8 \times 10^2 \pm 3.3 \times 10^3$ (26)	4.0 ± 0.65	$1.0 \times 10^{10} \pm 4.5 \times 10^9$
HAp	131	2.9	$7.6 \times 10^6 \pm 3.2 \times 10^6$ (26)	$2.9 \times 10^3 \pm 7.9 \times 10^3$ (26)	3.3 ± 0.72	$1.0 \times 10^{10} \pm 4.5 \times 10^9$
VA	97	na	$7.7 \times 10^6 \pm 3.2 \times 10^6$ (27)	$5.2 \times 10^4 \pm 1.2 \times 10^5$ (27)	2.1 ± 0.48	$7.3 \times 10^9 \pm 3.1 \times 10^9$
VAp	97	na	$7.6 \times 10^6 \pm 2.9 \times 10^6$ (25)	$5.2 \times 10^4 \pm 1.0 \times 10^5$ (25)	2.1 ± 0.48	$7.2 \times 10^9 \pm 2.9 \times 10^9$
VS1	95	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$1.2 \times 10^5 \pm 6.0 \times 10^5$ (42)	1.6 ± 0.75	$6.4 \times 10^9 \pm 3.2 \times 10^9$
VS1p	95	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$4.4 \times 10^4 \pm 2.4 \times 10^5$ (42)	2.1 ± 0.68	$6.6 \times 10^9 \pm 3.2 \times 10^9$
VS2	95	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$1.7 \times 10^5 \pm 5.8 \times 10^5$ (42)	1.5 ± 0.60	$6.4 \times 10^9 \pm 3.3 \times 10^9$
VS2p	95	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$1.0 \times 10^5 \pm 2.0 \times 10^5$ (42)	1.8 ± 0.56	$6.6 \times 10^9 \pm 3.2 \times 10^9$
VG	96	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$6.4 \times 10^5 \pm 1.4 \times 10^6$ (42)	0.9 ± 0.63	$5.5 \times 10^9 \pm 3.0 \times 10^9$
VGp	96	na	$7.2 \times 10^6 \pm 3.2 \times 10^6$ (42)	$8.9 \times 10^5 \pm 1.7 \times 10^6$ (42)	0.8 ± 0.50	$5.3 \times 10^9 \pm 2.7 \times 10^9$
R	146	2.5	$7.3 \times 10^6 \pm 4.1 \times 10^6$ (36)	$9.3 \times 10^3 \pm 1.3 \times 10^5$ (36)	2.8 ± 0.73	$1.1 \times 10^{10} \pm 6.8 \times 10^9$

na = not applicable.
[a] HLR = hydraulic loading rate (based in inflow rate).
[b] *n*HRT = Nominal hydraulic residence time (based on average of inflow and outflow).

Figure 13: Example of reported data. Mean hydraulic loading and E. coli concentration and removal rates (±one standard deviation) for the different wetland beds over the study period. Source: Headley et al., (2013).

3.1.2 Inclusion criteria

The primary objective of this systematic literature review was the extraction of results associated with various pathogen removal, preferably those related to the aforementioned viruses, bacteria, and protozoa groups. Furthermore, the secondary preferable objective was the optimal availability and quality of the technical features as well as the best possible description and quality of the experimental conditions of the papers. Therefore, the inclusion criteria comprised of case studies that described thoroughly the technical features of their experiment and included:

- The dimensions of the constructed wetland
- Hydraulic Loading Rate (HLR)
- Hydraulic Retention Time (HRT)
- Porosity (n) of the media grains
- A detailed description of CWs

As far as the experimental conditions, the preferred studies were those who included information regarding:

- Temperature (T)
- Type of influent wastewater (raw, domestic, pre-treated)
- Method of enumeration, (Colony Forming Units (CFU/100 ml), Most Probable Number (MPN))
- Physicochemical parameters like
    - Biological Oxygen Demand (BOD)
    - BOD_5
    - Chemical Oxygen Demand (COD)
    - Total Suspended Solids (TSS)

As can be seen from the flow diagram in Figure 11, after two series of screening, a total of 56 case studies qualified for both qualitative and quantitative analyses.

3.2.1 Spatial distribution of case studies

After the selection of the included case studies, a spatial distribution map was created to show the dispersion of studies throughout the world. Figure 14 shows that the selected studies come from various regions of the world with different climatic zones, which results in the variability on the performance of different CW types and strengthens the validity of the results for the implementation of a generic model.

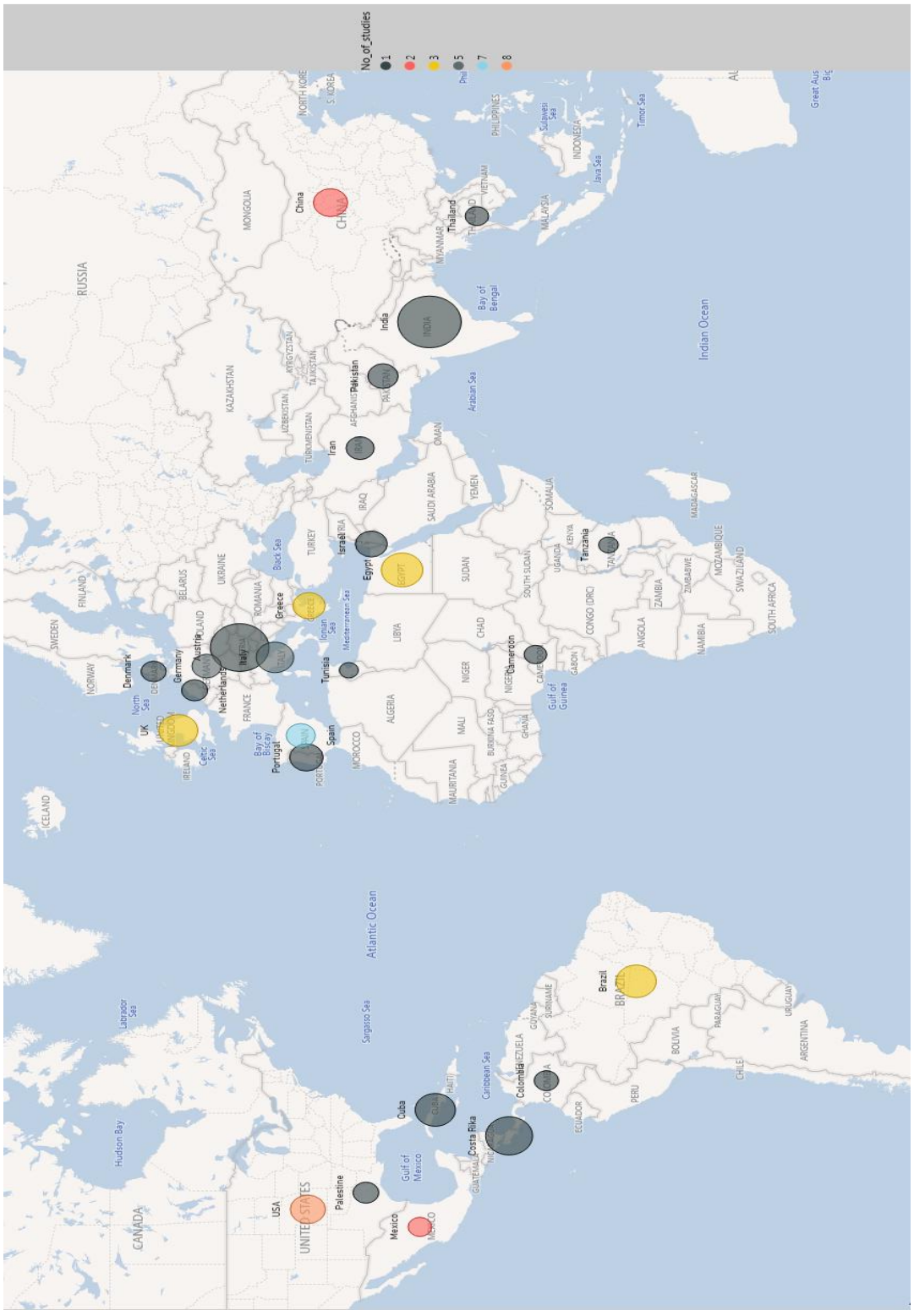
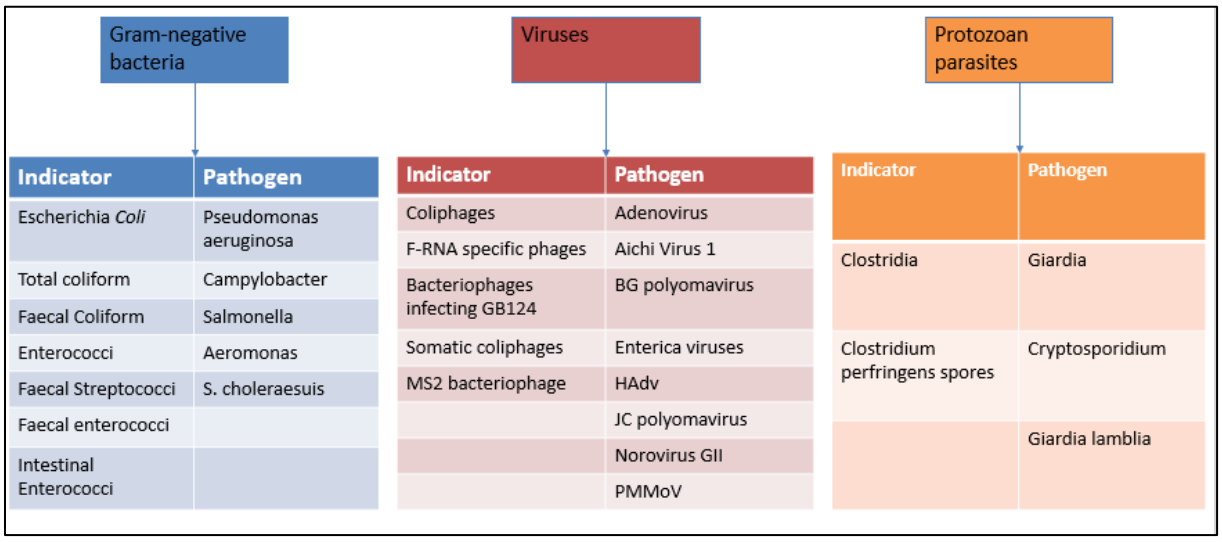


Figure 14: Spatial distribution of selected case studies. The creation of the map was made using Power Bi application.

3.2.2 Creation of dataset and classification of pathogens

After the data extraction was over, a comprehensive dataset with all the necessary values was created as can be seen in Appendix 2. This dataset includes comprehensive information about HRT, HLR, type of CW, and Concentration of influent and effluent for various pathogens. This information not only will help create a complete, up to date overview of the performance of constructed wetlands, but it will be also used as an input for the implementation of the predictive model in the next phase.

The next step was the classification of pathogenic microorganisms into three main categories named gram-negative, viruses, and protozoan parasites (Figure 15). As can be seen, the systematic literature review identified various microorganisms, both indicators, and pathogens for each category. Therefore, the grouping between indicators and pathogens for every category was necessary, in order to have a better understanding of their removal from CWs in our database, and facilitate the process of creating a robust, and predictive model. Regarding the latter, the grouping was considered an imperative decision since a ''strength in numbers'' approach was adopted. Briefly, although E. *coli* (which is a common indicator) was found in 37 of the 56 total papers, other indicators such as fecal coliforms, total coliforms, and fecal streptococci, did not have the same frequency (25, 28, and 10 respectively). Furthermore, from the viruses category, coliphages were found in only 8 papers in total, whereas from protozoan parasites category, Giardia had considerably less information, with only 5 papers. Therefore, we can see that there is not enough data (meaning not enough information) for all the important indicators or pathogens. Consequently, in order to develop a robust and reliable model that can have a generic application, the decision of grouping was considered a valid option since it was expected to have largest differences between categories, and rather smaller within categories.



Indicator	Pathogen
Escherichia *Coli*	Pseudomonas aeruginosa
Total coliform	Campylobacter
Faecal Coliform	Salmonella
Enterococci	Aeromonas
Faecal Streptococci	S. choleraesuis
Faecal enterococci	
Intestinal Enterococci	

Indicator	Pathogen
Coliphages	Adenovirus
F-RNA specific phages	Aichi Virus 1
Bacteriophages infecting GB124	BG polyomavirus
Somatic coliphages	Enterica viruses
MS2 bacteriophage	HAdv
	JC polyomavirus
	Norovirus GII
	PMMoV

Indicator	Pathogen
Clostridia	Giardia
Clostridium perfringens spores	Cryptosporidium
	Giardia lamblia

Figure 15: Grouping of pathogenic microorganisms in three main categories.

Moreover, Figure 16 demonstrates the range and average log removal for each category between indicators and pathogens. It can be seen (for gram-negative bacteria and protozoa) that indicator and pathogenic microorganisms have relatively similar levels of removal, which increases the reliability of the decision to group indicators with pathogens. For viruses, the indicators appear to be removed to a lesser extent, therefore they should be regarded as conservative indicators.

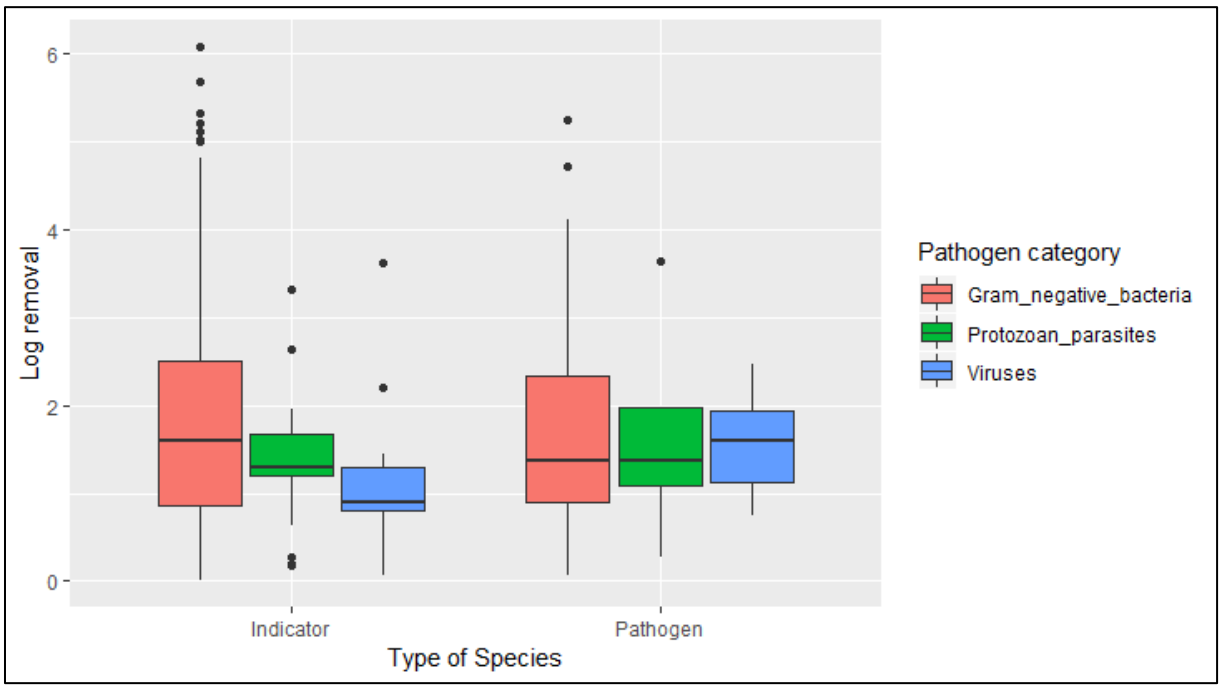


Figure 16: Log removal comparison between indicators and pathogens for gram-negative, protozoa, and viruses categories.

3.4 Selection of the most suitable models for the creation of a predictive model

Over the last decades, many mathematical models have been developed in an effort to describe the reduction of pollutants and predict the log removal of pathogens in CWs. The majority of them, are characterized as ''black box'' models, a very popular term that follows constructed wetlands the last few decades. These models usually are fed with observed data for the creation of mathematical equations (Dotro et al., 2017; Hamaamin et al., 2014). The main categories of these models are:

- Statistical models
- First-order kinetics models
- Process-based models

Only the first two categories are going to be described and implemented in this paper since the process-based models are extremely complex and difficult to find a generic utilization. Regarding the latter, the internal processes that occur in a CW are investigated for the derivation of energy and mass balance equations. Therefore, many sub-models have been created such as flow models, transport models, and biokinetics models in an effort to develop an integrated process-based model. Briefly, the most important models that have been developed are:

- The HYDRUS Wetland Module by Langergraber & Šimůnek, (2012) that was incorporated in the HYDRUS simulation software (Šimůnek et al., 2012) and
- BIO_PORE (Samsó & Garcia, 2013) using the COMSOL MultiphysicsTM platform.

A number of different approaches were applied aiming at selecting the most appropriate model that will best describe (statistically) the removal of pathogens. All of these processes had statistical modeling as a common principle and therefore, a regression analysis (linear and non-linear) was performed. The implementation of the regression analysis was done using the programming language R Studio software (Verzani, 2011). The goal was to determine (through regression analysis) a mathematical relationship between specific parameters with the highest possible statistical correlation and not to describe and analyze the internal processes and dynamics of CWs. Therefore, the objective was to create a model that can predict effluent concentrations (C *out*) as a function of influent concentrations (C *in*), hydraulic retention time (HRT), and hydraulic loading rate (HLR) since those were the parameters that could be found consistently in the literature. The ideal scenario would be the development of a predictive model for each pathogen category (gram-negative bacteria, viruses, and protozoa) per type of constructed wetland (FWS, SSHFCW, and VFCW). To do that, various approaches were used to identify the most appropriate model for each type of constructed wetland. The most appropriate model will be the one that outperforms the rest in terms of statistical residuals such as R square[9], RMSE[10], MAE[11] and correlation between observed and predicted values. Various approaches

[9] The coefficient of determination (R squared or R^2) is the proportion of variance in the dependent variable that can be explained and predicted from the independent variable (Steel & Torrie, 1960).

[10] The Root-Mean-Square Error (RMSE) is a common measurement of accuracy, to compare predicting errors between different models for a selected dataset. It is often used to demonstrate the differences between predicted values and observed values of a model. The RMSE represents the square root of the second sample moment of the differences between predicted values and observed values. It serves to aggregate the magnitudes of the errors in predictions into a single measure of predictive power (Hyndman & Koehler, 2006).

mean various models, and this is how they will be addressed from now on. Therefore, the models examined where:

1. Linear statistical model
2. Non-linear statistical model A
3. Non -linear statistical model B
4. Non -linear empirical model C
5. Non -linear empirical model D

Finally, after looking at the parameters of the dataset and the literature review about modeling removal processes for CWs, the parameters that were chosen for the implementation of the predictive model are:

- C *out*= Concentration effluent (CFU/100 ml)
- C *in* = Concentration influent (CFU/100 ml)
- HLR=Hydraulic Loading Rate (m/day)
- HRT=Hydraulic Residence Time (day)

Being able to predict the concentration of the effluent and by using the concentration of the influent as an input (along with the input of the remaining parameters) it will enable any user to calculate the log removal of pathogens (LRV) using equation 1.

$$\boldsymbol{LRV = log\left(\frac{C\ in}{C\ out}\right)} \quad \textbf{(1)}$$

3.4.1 Linear statistical model

Linear regression analysis was used in this model to predict the effluent concentration. Since in this case, the purpose was to find a relationship between a dependent variable (in this case C *out*) and multiple independent variables such as C *in*, HLR, and HRT, the process is called multiple linear regression.

Multiple linear regression is an approach that attempts to model a relationship between two or more explanatory variables (independent variables) and a response variable (dependent variable) by fitting a linear equation to the observed data (Mason & Perreault, 1991). Additionally, the statistical residuals (R^2, RMSE, and MAE) are applied in both single and multiple regression analysis. Furthermore, the implementation of the first model serves the purpose of getting a first impression between the different parameters that influence the effluent concentration. Moreover, multiple case studies have performed a regression analysis in order to create a predictive model, therefore the development of this model -as a first attempt to shade some light in the black box of CWs- was considered a reliable choice (Son et al., 2010; Ston et al., 2004; Knight et al., 2000).

The regression equation was created to predict effluent concentrations as a function of influent concentrations, hydraulic retention time, and hydraulic loading rate and took the form of:

$$\boldsymbol{C\ out = C\ in * HRT * HLR} \quad \textbf{(2)}$$

[11] The Mean Absolute Error (MAE) is a measure of the difference between a predicted variable X and an observed variable Y. MAE is the average vertical distance between each point (variable X) and the identity line (variable X) (Willmott & Matsuura, 2005).

The units of the two parameters of concentrations are in CFU/ 100 ml, whereas HRT is in days and HLR is in m/d. Furthermore, since the range of the concentration values is between 500-100000 (CFU/100 ml), whereas the range of HLR is between 0.01-0.5 (m/d), it was decided that equation 2 should be log-transformed in order to create a better relationship between the parameters. Therefore, after the transformation, equation 2 took the form of:

$$\mathbf{log\ (C\ out) = log(C\ in) + log(HRT) + log(HLR)} \quad \textbf{(3)}$$

The expected outcome is an equation in the form of:

$$\log(\mathrm{C}\ out) = a * \log(\mathrm{C}\ in) + b * \log(\mathrm{HRT}) + c * \log(\mathrm{HLR})$$

With *a*, *b*, and *c* being the regression coefficients that determine the individual relationship of each independent parameter with the dependent parameter.

3.4.2 Non-linear statistical model A

In this model, an assumption was made that the relationship between the dependent parameter and the three independent parameters is characterized by non-linearity. This means that the selected independent parameters have a power relationship with the dependent parameter which until now is unknown and therefore, the model will calculate the power using regression coefficients. Therefore, again a regression equation was created to predict effluent concentrations as a function of influent concentrations, hydraulic retention time, and hydraulic loading rate and took the form of:

$$\boldsymbol{C\ out = a * C\ in^{b} * HLR^{c} * HRT^{d}} \quad \textbf{(4)}$$

Again, the same process was followed regarding the transformation of the equation in Log values. Therefore, the expected outcome is an equation in the form of:

$$\boldsymbol{log(C\ out) = log(a) + b * log(C\ in) + c * log(HLR) + d * log(HRT)} \quad \textbf{(5)}$$

With *a*, *b*, *c*, and *d* being the regression coefficients that determine the individual relationship of each independent parameter with the dependent parameter.

3.4.3 Non-linear statistical model B

This model is closely related to the previous since the only difference is the absence of HRT in the equation. Multiple studies have been conducted for the removal of pollutants and pathogens from CWS. In their experiments, they use an equation through regression analysis where they try to determine if there are significant relationships between C *in*, C *out*, and HLR while disregarding completely the influence of HRT (Son et al., 2010; Ston et al., 2004; Knight et al., 2000). Therefore, the same approach was adopted in this model in an effort to examine the possibility of HRT not being such an influential parameter. The equation has the form of:

$$\boldsymbol{C\ out = a * C\ in^{b} * HLR^{c}} \quad \textbf{(6)}$$

After the log transformation:

$$\boldsymbol{log(C\ out) = log(a) + b * log(C\ in) + c * log(HLR)} \quad \textbf{(7)}$$

With *a*, *b*, and *c* being the regression coefficients that determine the individual relationship of each independent parameter with the dependent parameter.

3.4.4 Non-linear empirical model C

In this non-linear empirical model, we introduce the First-Order k-C* Model proposed first by Kadlec & Knight (1996). This approach is based on the idea that the removal of contaminants has an exponential relationship with either distance traveled, or with time mainly in horizontal subsurface flow constructed wetlands. Many papers have used this equation to describe the relationship between influent and effluent concentrations and the respective pollutant/pathogen removal (Kadlec, 1997; Kadlec, 1999; Avelar et al., 2014; Arias et al., 2003; Stone et al., 2004). The relationship between C *in* and C *out* can be described like this:

$$\frac{C\ out - C^*}{C\ in - C^*} = e^{\frac{-k_A}{q}} \quad \textbf{(8)}$$

where

C^*= Background concentration (CFU/100 ml)

K_A= Areal removal rate constant (m/d)
q= Hydraulic Loading Rate (m/d)

A lot of assumptions needed for the creation of the First-Order k-C* Model and these are:

- The constructed wetland is in a stationary state
- The precipitation cancels out the evapotranspiration, meaning that the wetland has a constant averaged flow
- No infiltration occurs
- No atmospheric depositions occur
- The shape of the CW is rectangular
- There are plug flow conditions (no back mixing)
- The cross-flow direction is not changing

Multiple scientific papers have used the First-Order k-C* model with further assumptions and implications being made like the default value of C* to zero (Vymazal, 2005a; Arias et al., 2003). This approach was adopted in this case since no background concentrations were found in the dataset. Therefore, after modifications, the equation used in this model has the form of:

$$C\ out = a * C\ in * e^{\frac{b}{HLR}} \quad \textbf{(9)}$$

With *a* being the regression coefficient and b representing the average constant value of k_A (decay coefficient). After exploring different ways to determine a value for the areal removal rate constant that could be representative for the whole model, it was decided to treat it as a regression coefficient and let the model give a constant value. Furthermore, after log-transforming equation 9, it took the form of:

$$log(C\ out) = log(a) + log(C\ in) + log(e) * \left(\frac{b}{HLR}\right) \quad \textbf{(10)}$$

3.4.5 Non-linear empirical model D

The final model is just an extension of the previous model in an effort to examine the influence of HRT in the removal of pathogens (Avelar et al., 2014; Kadlec, 1997; Kadlec, 1999; Selvakumar et al., 2007; Boutilier et al., 2009; Gonzales et al., 2018). It originates from the same initial equation with the same assumptions and demonstrates an exponential relationship between the parameters:

$$\frac{C\ out - C^*}{C\ in - C^*} = e^{\frac{-k_A}{q}} = e^{-k_V * t} \quad \textbf{(11)}$$

where

K_V= Volumetric removal rate constant (d^{-1})
t= Hydraulic Retention Time (d)

After using the same assumptions with the previous model and log-transforming, the equation has the form of:

$$log(C\ out) = log(a) + log(C\ in) + log(e) * (b * HRT) \quad \textbf{(12)}$$

With a being the regression coefficient and b representing the volumetric removal rate constant that the model will fit best.

4. Results

This chapter includes the results from the systematic literature review in the form of a complete dataset (Appendix 1), the development of a predictive model, as well as a storyboard of GUI. As far as the systematic literature review, there will be an overview of the removal efficiency of CW types as well as the performance of different pathogen categories and their respective characteristics in every type of CW. Furthermore, regarding the initial idea of creating a predictive model for each pathogen category per type of constructed wetlands, the results were not so promising. This is because as already discussed in chapter 3, data availability for specific pathogens was not at the required level to execute a regression analysis, even after the grouping of microorganisms.

4.1 Removal performance of constructed wetlands

Table 3 demonstrates the overall performance of different types of CWs. The majority of papers in the systematic literature review used SSHFCW for the removal of pathogens, whereas all of the three types seem to have a potential of pathogen removal to a certain level. Briefly, the FWS types have log removal values ranging from 0.07 to 5.32, the SSHFCW from 0.01 to 5.68 and finally the VFCW type from 0.35 to 6.08. Figure 17 shows that the log removal for the three types of CWs ranges between 1 and 6 log, whereas the average value ranges between 1 and 2 log.

Table 3: Overall performance of CW types. n= number of data points.

	Type of CW	FWS	SSHFCW	VFCW
General info	Number of papers	21	31	19
	Number of pilot scale	11	22	16
	Number of full-scale	10	9	3
Pathogen removal	Log removal	1.6 (n=122)	1.7 (n=246)	2.12 (n=93)
	Percentage removal	99.98	99.97	99.99
	Minimum Log removal	0.07	0.011	0.35
	Maximum Log removal	5.32	5.68	6.08
	Standar Deviation	1.12	1.12	1.38
Physicochemical characteristics	Air temperature °C	20 (n=34)	19.2 (n=95)	19.4 (n=31)
	Water temperature °C	17.3 (n=47)	21.6 (n=121)	17.9 (n=63)
	PH	7.43 (n=73)	7.03 (n=1223)	7.35 (n=57)
	COD removal (mg/L) %	66.6 (n=57)	69.5 (n=169)	83.4(n=58)
	BOD removal (mg/L) %	72.6 (n=32)	68.2 (n=83)	92.5 (n=24)
	BOD5 removal (mg/L) %	65.03 (n=11)	83.01 (n=88)	93.7 (n=18)
	TSS removal (mg/L) %	71.6 (n=56)	65.6 (n=136)	82.8 (n=50)
	HLR range (m/d)	0.058-5.1 (n=107)	0.005-2.59 (n=255)	0.0028-1.36 (n=86)
	HRT range (d)	0.29-7 (n=107)	0.028-15 (n=255)	0.01-9.1 (n=86)

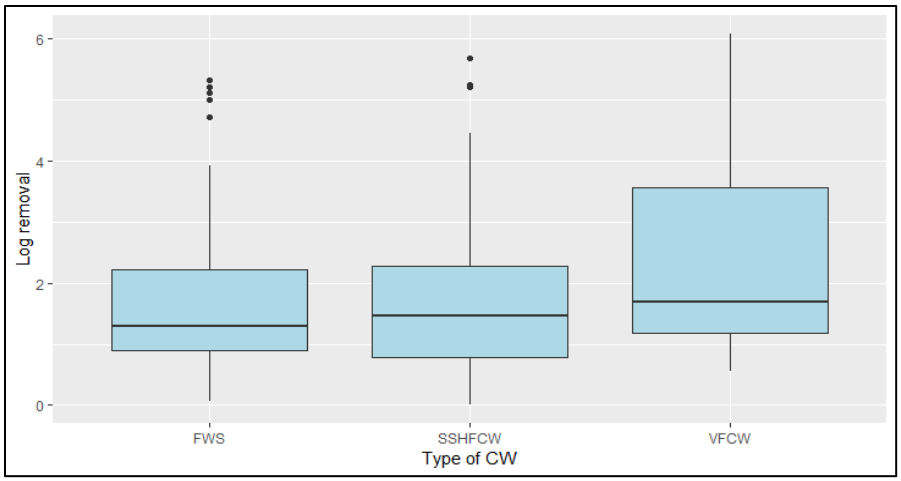


Figure 17: Log removal comparison between types of CW.

Moreover, as can be seen in Figures 18, 19, and 20, a preliminary examination of the parameters that are going to be used in the model was made. By plotting the influent and effluent concentration it can be observed that indeed there seems to be a linear relationship between the two parameters. Therefore, there is a strong belief that a robust predictive model can be created.

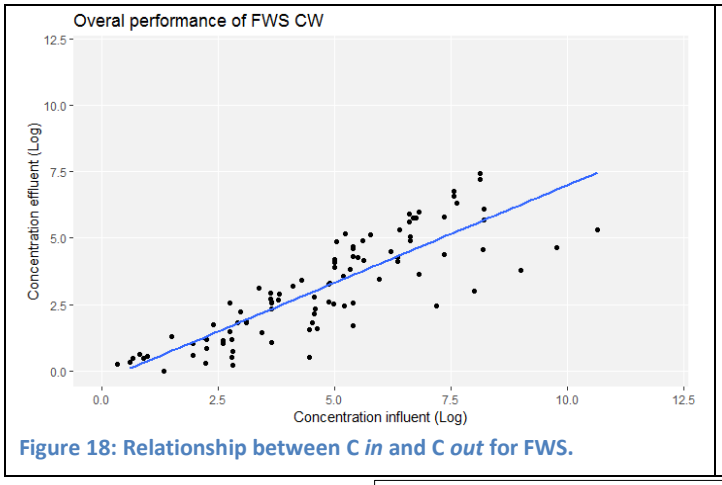


Figure 18: Relationship between C *in* and C *out* for FWS.

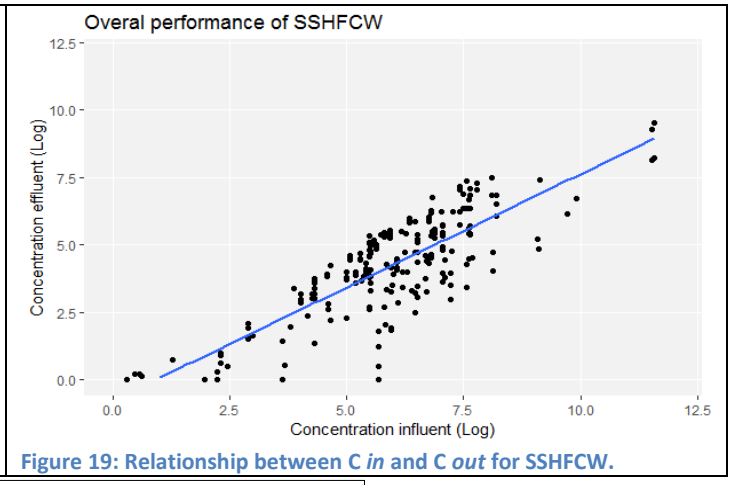


Figure 19: Relationship between C *in* and C *out* for SSHFCW.

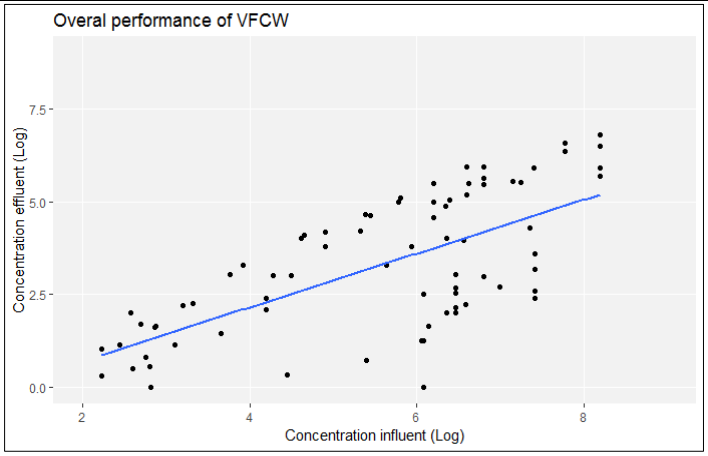


Figure 20: Relationship between C *in* and C *out* for VFCW.

4.2 Removal of pathogen categories

The category of gram-negative bacteria can be found in the majority of the papers collected (Table 4). This comes as no surprise since the specific category includes the most common indicators and index pathogens such as E. *Coli,* Total Coliforms, Fecal Coliforms, Campylobacter and Salmonella. Furthermore, as it was expected the same category is responsible for the minimum and maximum log removal values overall, when compared with the viruses and protozoan parasites categories. Table 5 gives an overview of the three pathogen categories regarding their removal per CW type.

Furthermore, Table 4 depicts the overall performance of the gram-negative bacteria while the rest of the categories can be found in Appendix 3. The Tables include information regarding the standard deviation of their removal, the number of pilot and full-scale applications as well as physicochemical characteristics such as air and water temperature, PH, HRT, and HLR range. In addition to that, there are values of COD, BOD and TSS removal percentage that can be indicators of pollutant removal for each type of CW and pathogen category. What can be extracted from these tables is the difference and the small number of data points between types of CWs regarding their reported physicochemical characteristics. Only the HRT and HLR were consistently reported and that is why they were the only two parameters selected. Additionally, Figure 21 shows that gram-negative bacteria show great variability in terms of log removal (0.01-6.08) compared to viruses and protozoan parasites categories where the ranges are between 0.02-3.62, and 0.18-3.63 respectively. Again, we can see the average log removal values ranging between 1 and 2 log for the three categories in all of the CW types except the viruses category in VFCW where the average log removal is less than 1 log. Furthermore, Figure 22 illustrates the average log removal of the most representative indicators and index pathogens as recorded during the systematic review.

Table 4: gram-negative bacteria removal performance. N= number of data points.

Pathogen category	Gram-Negative Bacteria		
General information			
Average removal (Log)	1.88		
Number of papers	52		
Number of pilot scale	34		
Number of full-scale	18		
Constructed wetlands			
Types of CWs	***FWS***	***SSHFCW***	***VFCW***
Pathogen removal	n=84	n=224	n=78
Log removal	1.83	1.70	2.27
Minimum Log removal	0.07	0.01	0.55
Maximum Log removal	5.32	5.68	6.08
Standar Deviation	1.19	1.14	1.45
Percentage removal	99.92	99.98	99.997
Physicochemical characteristics			
Air temperature °C	19.6 (n=27)	19.1 (n=84)	19 (n=27)
Water temperature °C	15.6 (n=29)	21.6 (n=108)	18.1(n=54)
PH	7.5 (n=52)	7.03 (n=108)	7.4(n=48)
COD removal (mg/L) %	66.2 (n=50)	67.5 (n=151)	82.6 (n=52)
BOD removal (mg/L) %	72.5 (n=26)	67.6 (n=81)	92.2 (n=22)
BOD5 removal (mg/L) %	64.8 (n=10)	84.6 (n=71)	92.46 (n=14)
TSS removal (mg/L) %	71.1 (n=49)	64.7 (n=127)	83.5(n=45)
HLR range (m/d)	0.058-5.1 (n=75)	0.005-2.59 (n=219)	0.0028-1.36 (n=72)
HRT range (d)	0.29-7 (n=74)	0.28-15 (n=219)	0.01-9.1 (n=72)

Table 5: Average log removal range for the three categories per CW type.

Pathogen category	Type of CW		
	FWS	SSHFCW	VFCW
Gram-Negative Bacteria	0.07-5.32	0.01-5.68	0.55-6.08
Viruses	0.06-2.88	0.02-3.62	0.57-1.75
Protozoan Parasites	0.18-2	0.28-3.63	0.35-2.72

The group of coliforms are those with the greatest variability. While the variability can be explained due to the different number of data points, another possible explanation is the difference between pilot and full-scale applications. The outliers in many of the species reinforce this assumption. Additionally, we see that there is great variability of pathogen log removal within their categories.

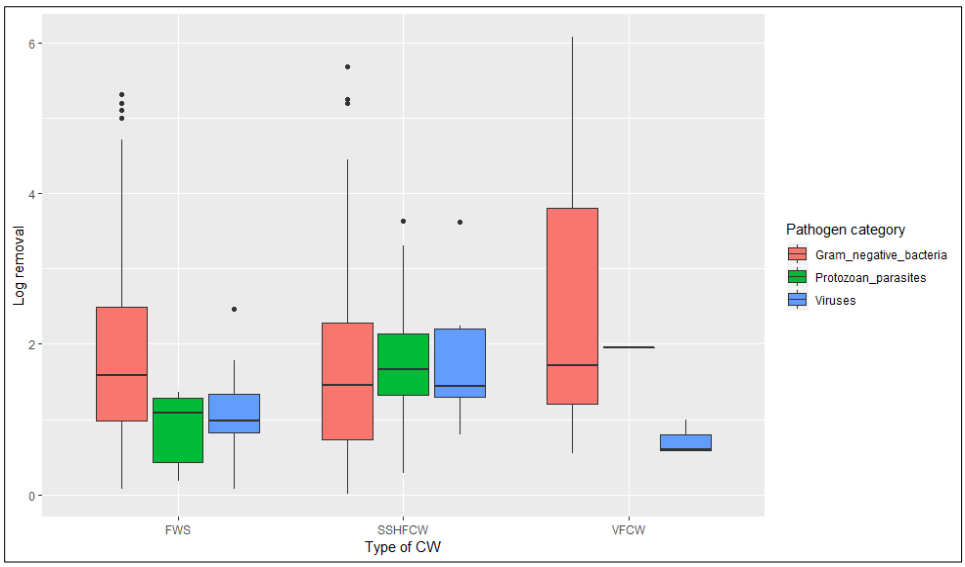


Figure 21: Comparison of pathogen categories per CW type.

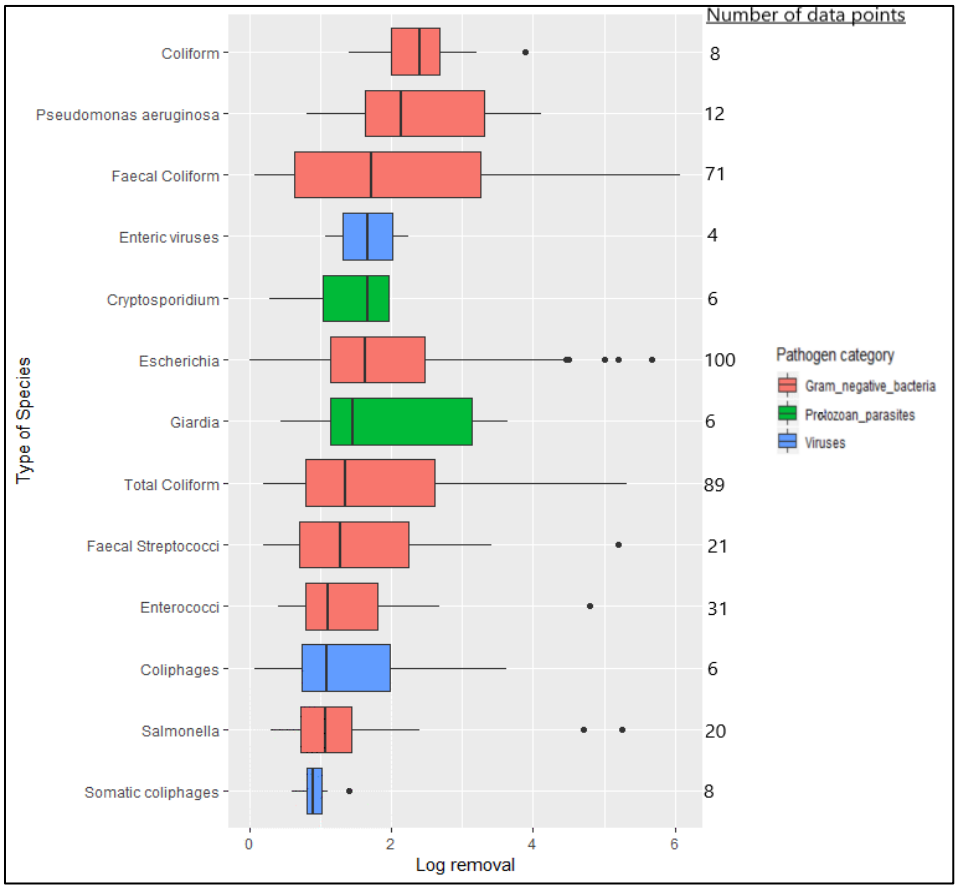


Figure 22: Log removal range for the most representative indicators and index pathogens, ranked based on average log removal.

4.2 Predictive model for different Pathogen Categories per CW type

One of the two objectives of this paper was to develop a model that can predict the removal of pathogenic microorganisms from constructed wetlands. The initial idea was to create a specific model for every pathogen category with the assumption that each category behaves differently per CW type. After going through the dataset, it was concluded that although the gram-negative Bacteria category (since it is the biggest group) had enough information, there were not enough data for the viruses and protozoan parasites categories for the implementation of a regression analysis. Therefore, a decision on how to carry out the development of the model had to be made. Subsequently, it was decided to initially carry out the process of creating a model for every type of CW, with only the gram-negative bacteria category. The model came up with the residuals of the regression analysis including the coefficients of every parameter for each of the five different models that were discussed in Chapter 3.4. The results of the models along with their statistical residuals can be found in Appendix 4, whereas Table 6 serves as an example here.

Table 6: Residuals of linear statistical model regression analysis for FWS CW.

FWS Linear statistical model	Coefficients	P value	R Squared	RMSE	MAE
Intercept	-0.2	0.6	0.67	1.16	0.93
C in	0.76	<2*e^-16			
HRT	-1.24	0.04			
HLR	-0.01	0.96			

Then, in order to see whether the developed model can be used as a generic model for all the three pathogen categories, it was decided to use the same coefficients (this time as known constant values) but with the viruses and protozoan parasites as input datasets. This time, all five models were not tested again, but the best-performing model was preferred. That way we could ascertain whether the new dataset fit the previously developed model by means of checking the new R Squared, RMSE, MAE as well as checking the correlation between the modeled values and the observed values of each dataset. Although the Viruses and Protozoan datasets are small, they are enough to draw conclusions on whether they fit the model. The results were very promising since both datasets fitted the model quite well as can be seen in Tables 7, 8, and 9. The R Squared values for all the pathogen categories in every CW type were pretty high, meaning that almost all of the variation of the (new) dataset can be explained by the predictive model. The RMSE and MAE values were quite low again in every category, meaning that the predicted data points of the model are really close to the observed points. Finally, the correlation between the observed and predicted values looks great for all of the three categories in every CW type since the closer the value is to 1, the stronger the correlation between the two variables. All the correlations for the different pathogen categories per CW type can be found in Appendix 5, whereas Figure 23 can be used as an example here.

Table 7: Performance between the three datasets for FWS CW type.

FWS CW	RMSE	R Squared	MAE	Correlation*
Gram-Negative Bacteria dataset	1.24	0.51	1.1	0.80
Viruses dataset	0.51	0.8	0.43	0.89
Protozoan dataset	0.42	0.78	0.33	0.88

* Correlation between observed and modeled values

Table 8: Performance between the three datasets for SSHFCW.

SSHFCW	RMSE	R Squared	MAE	Correlation*
Gram-Negative Bacteria dataset	1.19	0.78	0.90	0.93
Viruses dataset	0.94	0.97	0.89	0.98
Protozoan dataset	0.85	0.88	0.71	0.94

* Correlation between observed and modeled values

Table 9: Performance between the three datasets for VFCW.

VFCW	RMSE	R Squared	MAE	Correlation*
Gram-Negative Bacteria dataset	1.13	0.68	0.89	0.77
Viruses dataset	0.33	CBD	0.28	CBD
Protozoan dataset	0.62	CBD	0.62	CBD

* Correlation between observed and modeled values. CBD=Cannot be determined.

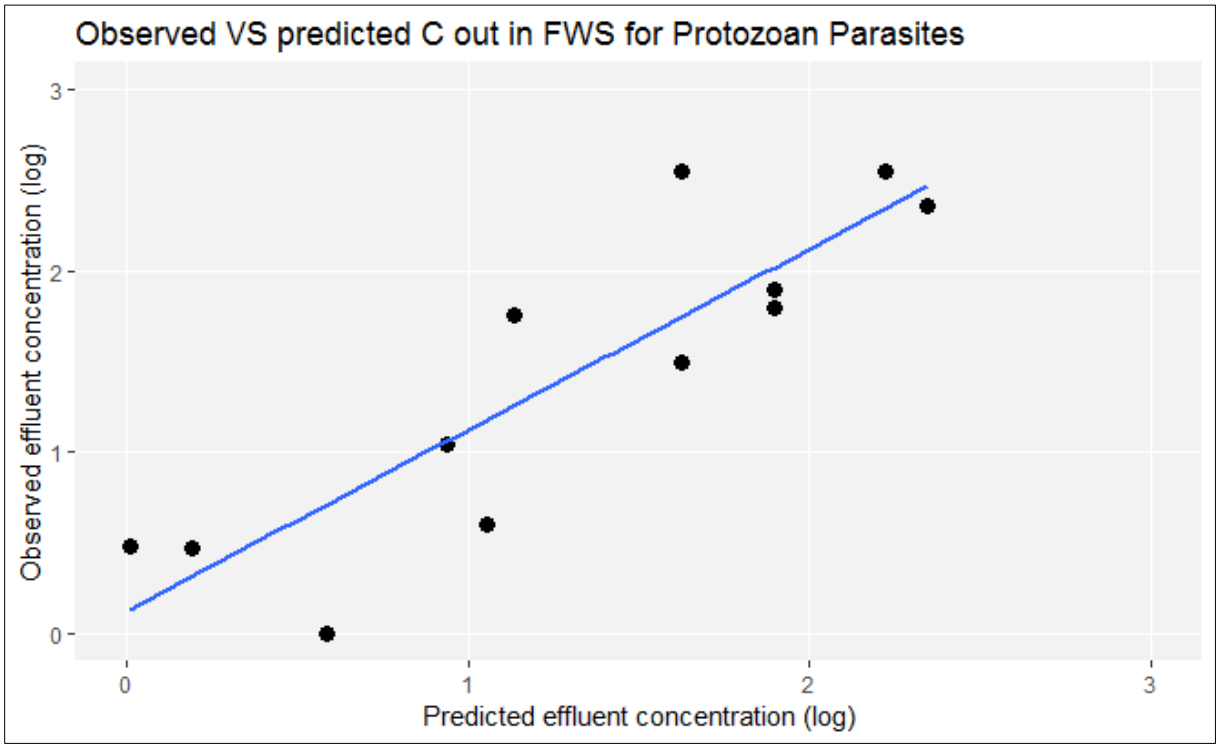


Figure 23: Comparison between observed and predicted effluent concentration for FWS using Protozoan Parasites dataset.

4.3 Predictive model per CW type

Since the results of the comparison between the different datasets showed that the three datasets do not differ much, it was decided that it would be best to combine the three pathogen categories in one dataset and create a predictive model for the three types of CW. Therefore, adopting the ''strength in numbers'' approach and including more data points in the regression analysis will lead to a more robust and reliable predictive model. After running the regression analysis through the different models in order to find the best-performing one, the results can be found in Appendix 6. Table 10 will serve as an example here.

Table 10: Performance comparison between the different models. Based on the statistical residuals the Non-linear statistical model A has the best score since it has the smallest RMSE and MAE value, the highest R^2, and one of the highest correlations. Therefore, it is ranked first as the most suitable model.

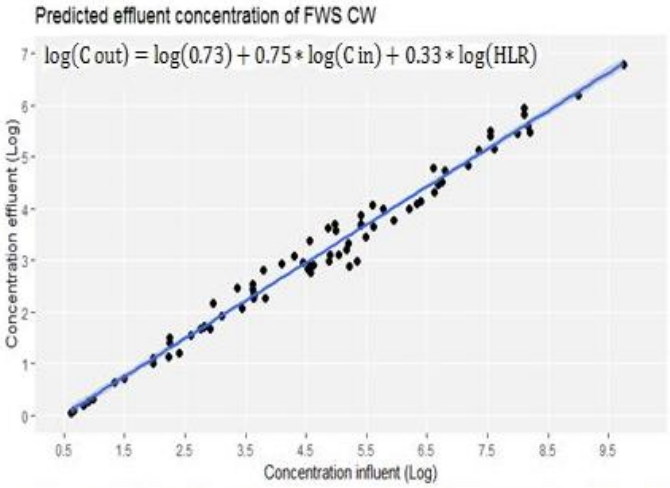
SSHFCW	*RMSE*	*R Squared*	*MAE*	*Correlation*	*Rank*
Linear statistical model	**0.98**	**0.88**	**0.75**	**0.94**	**2**
Non-linear statistical model A	**0.97**	**0.89**	**0.72**	**0.94**	**1✓**
Non-linear statistical model B	**1.14**	**0.85**	**0.90**	**0.92**	**5**
Non-linear empirical P-C-K model C	**1.02**	**0.88**	**0.72**	**0.93**	**3**
Non-linear empirical P-C-K model D	**1.04**	**0.88**	**0.82**	**0.94**	**4**

The results were very promising for all three types. Moreover, it was decided that the best model should not only have the best residuals but also provide the most information in terms of parameters. For example, a model that may not have the best possible residuals, but contains all three parameters will be preferred, rather than a model with the best possible residuals but only with 2 parameters instead of 3. After comparing all the models, the most suitable model for each type of Cw was selected:

- FWS CW: $\log(\mathrm{C}\ out) = \log(0.73) + 0.75 * \log(\mathrm{C}\ in) + 0.33 * \log(\mathrm{HLR})$ (13)
- SSHFCW: $\log(\mathrm{C}\ out) = \log(0.29) + 0.99 * \log(\mathrm{C}\ in) - 0.23 * \log(\mathrm{HLR}) - 1.49 * \log(\mathrm{HRT})$ (14)
- VFCW: $\log(\mathrm{C}\ out) = \log(0.008) + 0.94 * \log(\mathrm{C}\ in) + 1.2 * \log(\mathrm{HLR}) + 0.45^{12} * \log(\mathrm{HRT})$ (15)

Lastly, Figures 24, 25, and 26 demonstrate the final predictive model for every CW type along with their respective equation. Additionally, the correlation between observed and predicted effluent concentration can be seen in the same Figures.

[12] The coefficient of HRT is not statistically significant (see Appendix 6) which means that the particular model is not the best in terms of performance. However, this equation was preferred over the best performance-model because it contained more information, as well as the latter was considered unreliable.



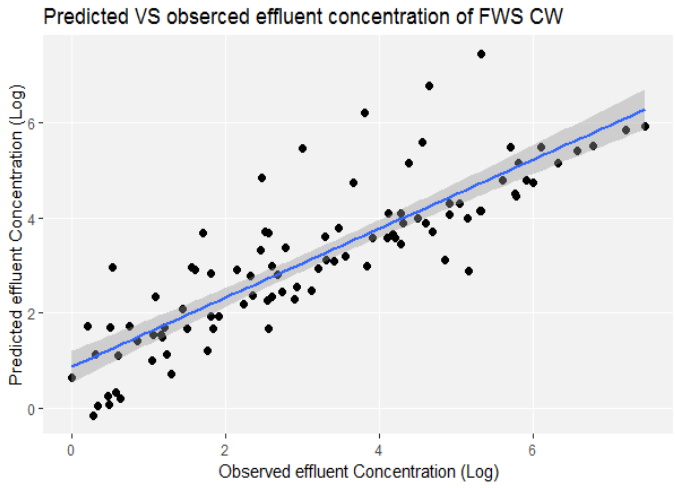
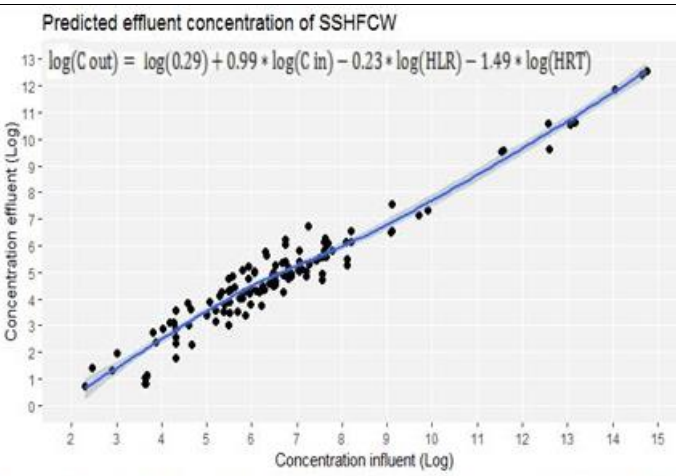


Figure 24: Left: Predicted effluent concentration of FWS CW. Right: Observed VS predicted effluent concentration.



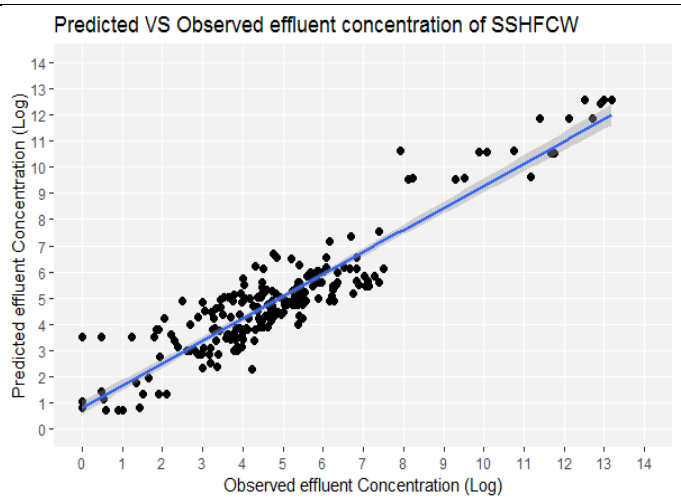
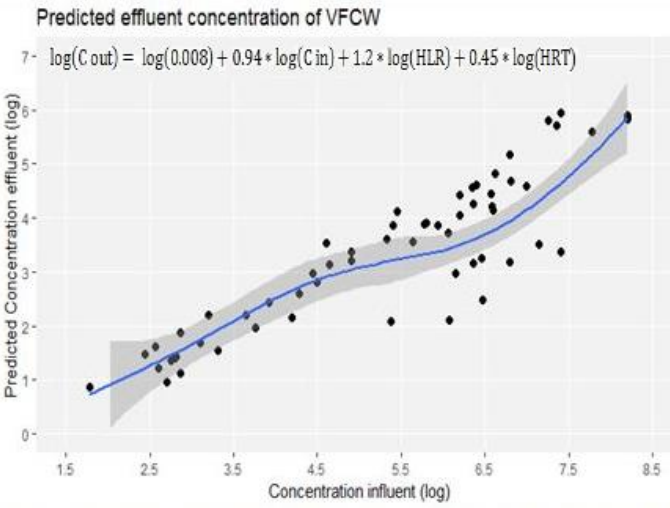


Figure 25: Left: Predicted effluent concentration of SSHFCW. Right: Observed VS predicted effluent concentration.



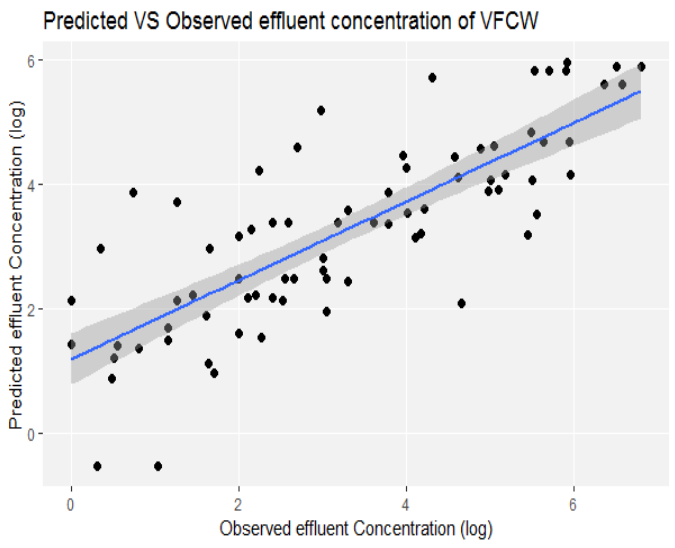


Figure 26: Left: Predicted effluent concentration of VFCW. Right: Observed VS predicted effluent concentration.

4.4 Storyboard

This section includes the incorporation of the outcome into the treatment calculator in the form of a Graphical User Interface (GUI). Through this platform, the user has the chance to make a variety of selections. The GUI showcases step by step the available options one may have to choose, in order to finally have an estimation of the reduction of a specific pathogen for a specific CW type. The steps to be described next, are indicative as the platform is in a beta phase, therefore, a lot of changes and modifications can be made in order to improve the user experience.

- **Step 1**: The user gets to choose between three options of WW treatment with the option of CW being the newest addition (Figure 27).
- **Step 2:** The user can choose from a variety of options regarding pathogens. The availability of options is dependent on the number of references available. This section includes both indicators and index pathogens (Figure 28).
- **Step 3:** This step includes the new features for choosing the desired setpoints for Cin, HRT, and HLR. Additionally, the selection of the desired type of CW is available (Figure 29).
- **Step 4:** This section consists of the final outcome in the form of a fitted line that gives an overview of log removal for a specific pathogen. It also includes the corresponding equation (Figure 30).
- **Step 5:** The final outcome can be exported in the form of attribute tables (Figure 31).

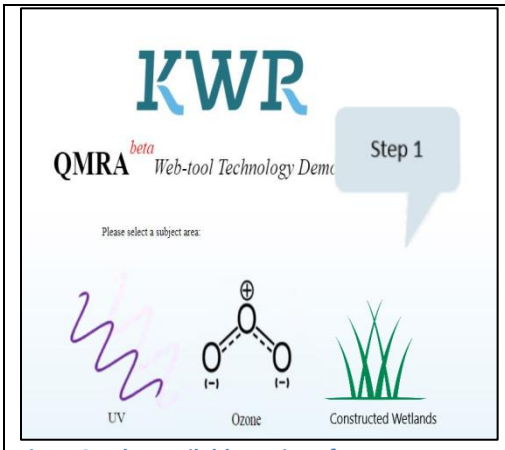


Figure27:The available options for treatment.

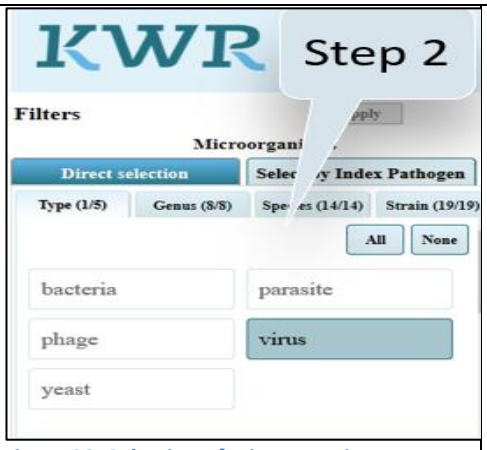


Figure 28: Selection of microorganisms

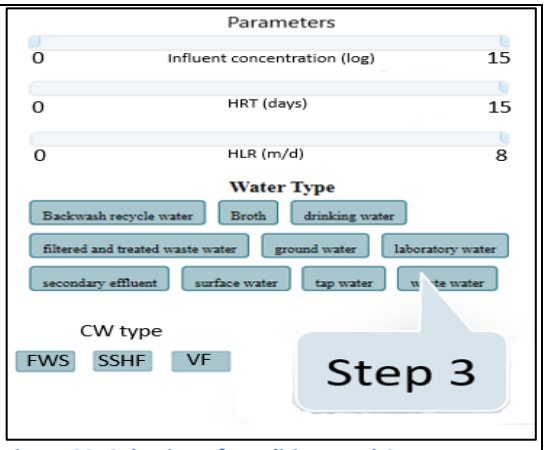


Figure 29: Selection of conditions and CW type.

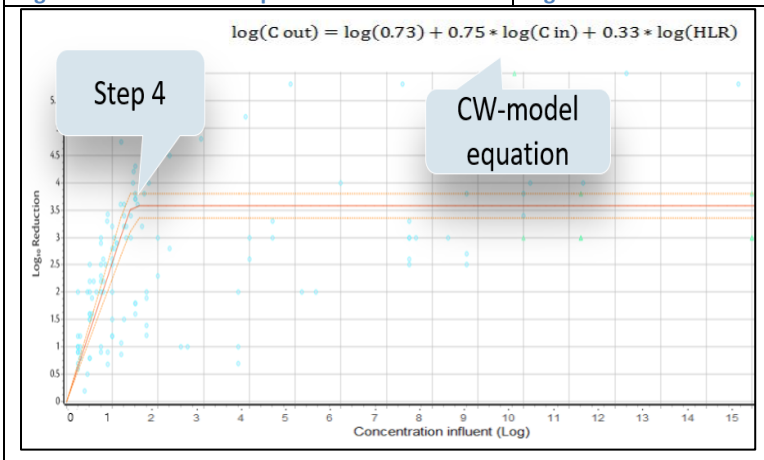


Figure 30: Log reduction of a specific pathogen along with the equation.

Figure 31: Ability to export results in the form of attribute tables.

The overview of the GUI can be illustrated in Figure 32, along with the corresponding references of data , following all the steps that were just discussed.

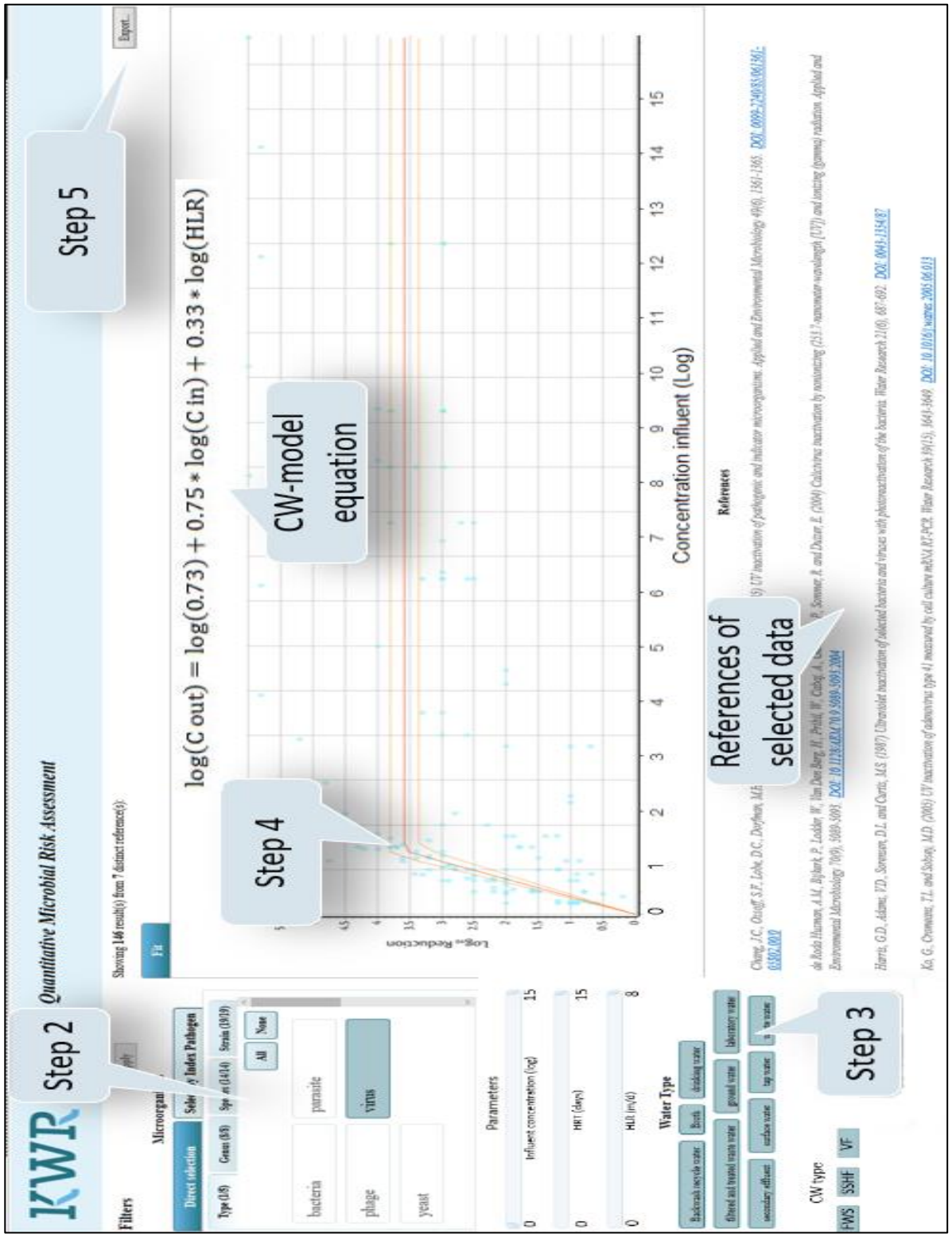


Figure 32: Complete overview of the GUI illustrating the required steps for the user.

5. Discussion

5.1 Dataset

The creation of a complete dataset enables any end user to have access to the overall performance of CW types. More importantly, the systematic review included case studies from all over the world, therefore the collected information on CWs as well as on pathogen removal can be considered reliable. Moreover, pilot-scale studies have been preferred the most, mainly because the majority of them were experimental case studies with an objective of investigating the removal efficiencies of pollutants and pathogens. Full-scale applications although used in many cases, they are not reported in scientific papers to the same extent.

The average log removal of the three types of CW was between 1 and 2 log, indicating that CWs cannot be a reliable option for pathogen removal as a standalone wastewater treatment method. However, they can be a credible choice for wastewater polishing when they are combined with conventional wastewater treatment systems. Furthermore, a linear relationship was identified by plotting the influent and effluent concentration prior to the development of the model that enabled us to anticipate a correlation between these two parameters. Therefore, it was expected that the influent concentration will be the most influential parameter for the calculation of the model, with HRT and HLR being secondary factors.

5.1.1 Pathogen categories

Regarding the removal of the three pathogen categories, Table 5 shows that the gram-negative Bacteria category (which includes the most common indicators) have the highest log removal in all of the three types of CW. On the other hand, the same group shows great variability which makes it difficult to draw tangible conclusions about the ease of removal of this category. This can best be seen in Figure 21 where the variability of all three categories is shown using boxplots. Protozoan Parasites show the less variability in FWS CW whereas the viruses category is the one with the smallest range in SSHFCW. Conclusions could not be drawn for the last two categories in VFCW types since their number of data points was low. In addition, the huge difference in variability between the three categories can be explained by the fact that the gram-negative bacteria category contains a huge database and to some extent, it is expected to see scattered removal values. On the other hand, although the other categories show less variability and it is easier to draw conclusions about their removal, their databases are quite small which should make any decision-maker skeptical about the reliability of the results.

5.2 Predictive model

5.2.1 Predictive model for pathogen categories

As already explained in the previous chapter, the predictive model for every pathogen category could not be implemented. The datasets for the viruses and protozoan parasites categories were too small. Therefore, the decision taken regarding the use of viruses and protozoan datasets with coefficients derived from the gram-negative bacteria model, resulted in great performances as can be seen in Tables 7, 8, and 9. At this point, it is worth noting that due to the small databases of viruses and protozoan parasites especially those related to VFCW type, the respective R^2 and correlation values could not be determined. The viruses database has 3 data points whereas the protozoan parasites only 2 with correlation values 0.97 and 1 respectively.

5.2.2 Predictive model per CW type

To get a better understanding of the three different models created in terms of what they actually represent and how each parameter incorporated in the model influences the final outcome a sensitivity analysis was performed, the results of which are shown in Figures 33-40. This time, the point of interest is the log removal and not the effluent concentration. The results of the Sensitivity Analysis (SA) can give an idea of which is the most influential parameter as well as what are the representative values of pathogen log removal one can encounter when performing experiments or treats wastewater with the use of CWs.

FWS Sensitivity analysis:

Equation: $\log(\mathbf{C\ out}) = \log(0.73) + 0.75 * \log(\mathbf{C\ in}) + 0.33 * \log(\mathbf{HLR})$ **(13)**

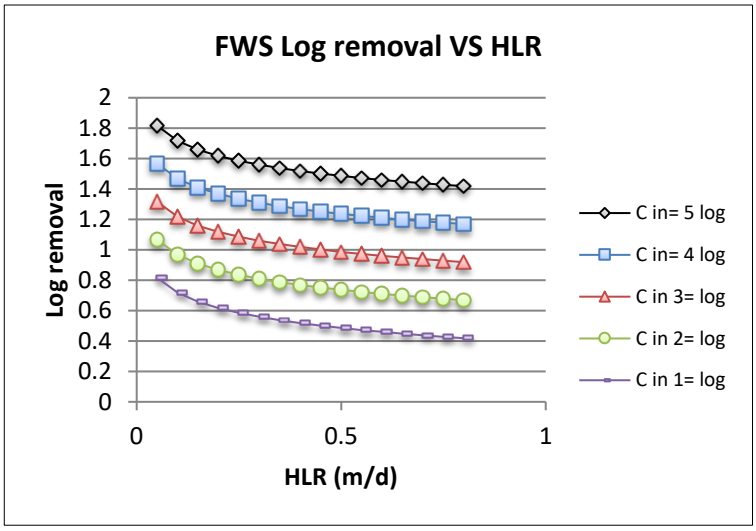


Figure 33: Sensitivity analysis of Log removal with influencing parameters of HLR in FWS.

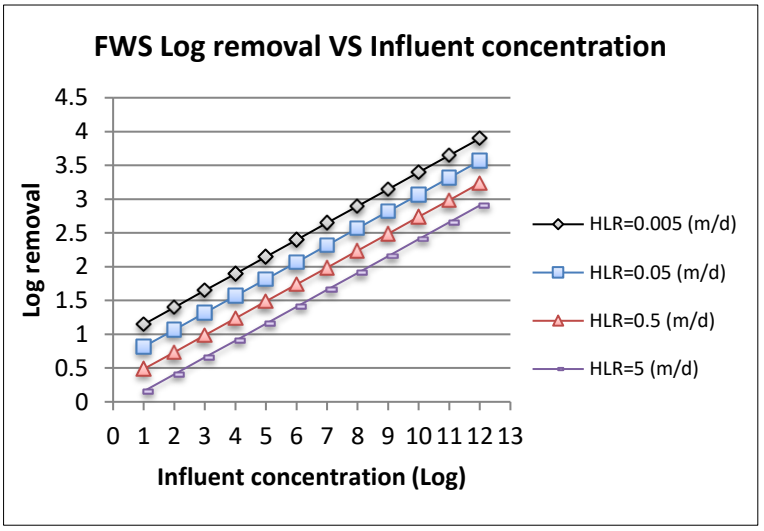


Figure 34: Sensitivity analysis of Log removal with influencing parameters of Cin in FWS.

As can be seen from Figures 33 and 34 regarding FWS, influent concentration is the one that most influences log removal compared to HLR, which also seems to play an important role. Furthermore, we can observe that the higher the Cin is, the better the FWS performs in log removal. On the contrary, as the HLR increases, the log removal levels decrease. Additionally, it can be observed that in the smaller values of HLR (between 0 and 0.5 m/d) the decrease is steeper, whereas in values higher than 1 m/d the log removal seems to reach a plateau. Therefore, in applications with high HLR (usually in full-scale CWs), the HLR does not play a crucial role.

SSHFCW Sensitivity analysis:

Equation: $\log(\text{C out}) = \log(0.29) + 0.99 * \log(\text{C in}) - 0.23 * \log(\text{HLR}) - 1.49 * \log(\text{HRT})$ (14)

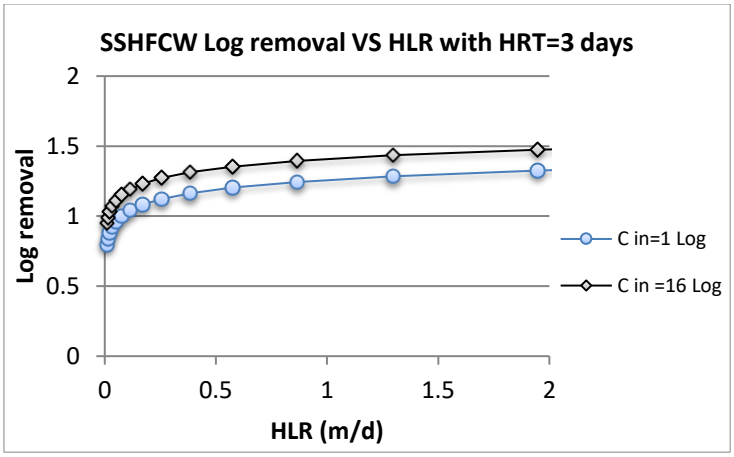


Figure 35: Sensitivity analysis of Log removal with influencing parameters of HLR in SSHFCW.

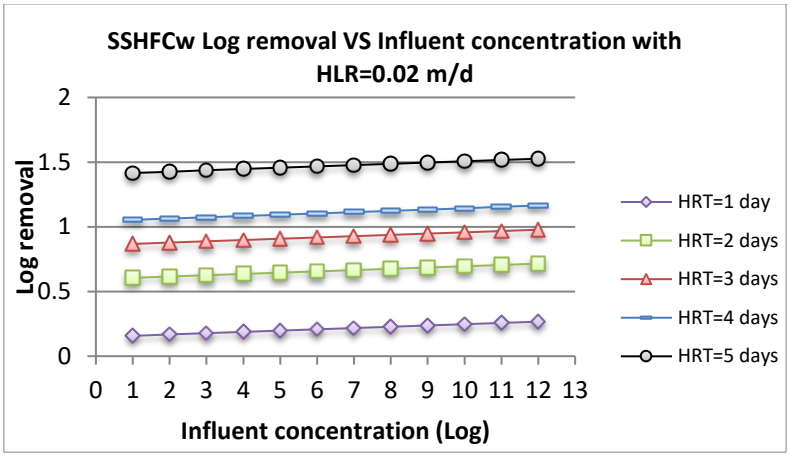


Figure 36: Sensitivity analysis of Log removal with influencing parameters of C *in* in SSHFCW.

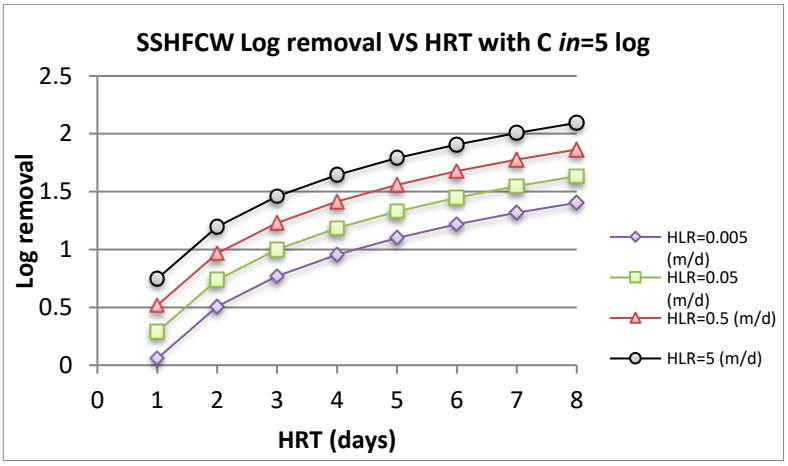


Figure 37: Sensitivity analysis of Log removal with influencing parameters of HRT in SSHFCW.

Regarding SSHFCW, the equation uses all three parameters. It can be observed from Figure 35 that HLR affects positively the log removal. Yet again, the smaller values are those who have a great impact, whereas when having HLR >1 m/d, log removal seems to level off meaning that in full-scale applications the log removal does not depend so much on HLR. Moreover, in the same plot, we can see that there is no much of a difference between Cin=1 log and Cin=16 log lines, which leads us on the next parameter. Figure 36 demonstrates the influence of influent concentration to log removal and it is easy to conclude that this is little to none. Furthermore, by comparing the different lines at the same graph, it looks like HRT plays a great role in the log removal of pathogens. This can be confirmed by looking at Figure 37 where here it is obvious that HRT is the most influential parameter. Overall, it can be concluded that all three parameters have a positive correlation with log removal.

VFCW Sensitivity analysis

Equation: $\log(\mathrm{C\ out}) = \log(0.008) + 0.94 * \log(\mathrm{C\ in}) + 1.2 * \log(\mathrm{HLR}) + 0.45 * \log(\mathrm{HRT})$ (15)

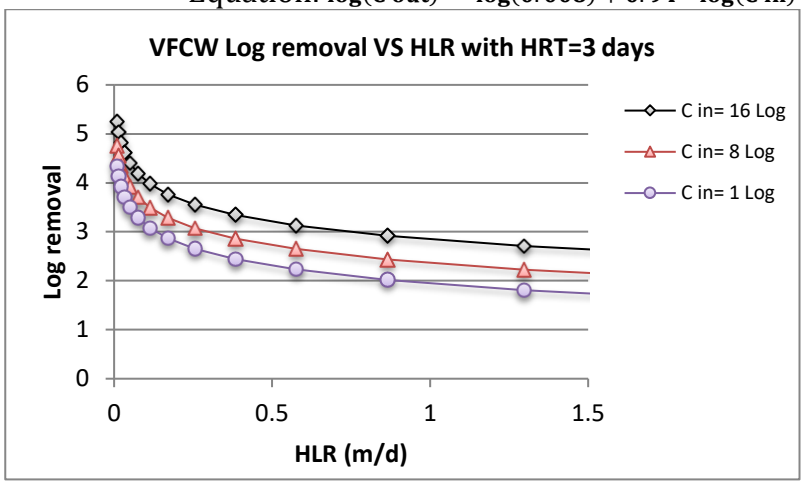


Figure 38: Sensitivity analysis of Log removal with influencing parameters of HLR in VFCW.

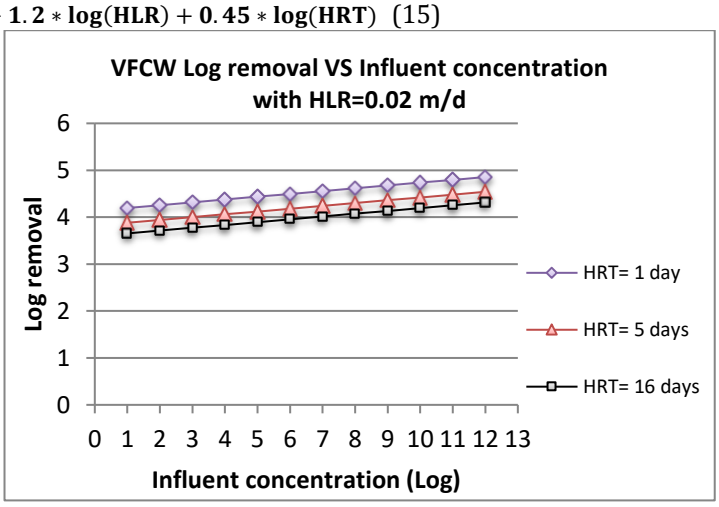


Figure 39: Sensitivity analysis of Log removal with influencing parameters of Cin in VFCW.

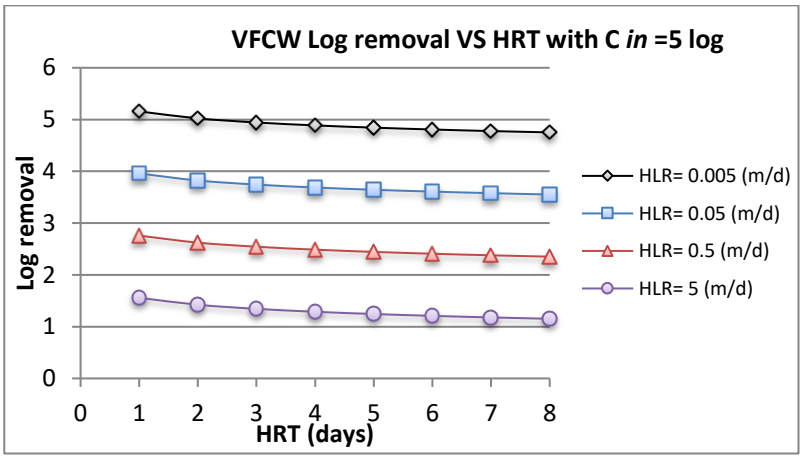


Figure 40: Sensitivity analysis of Log removal with influencing parameters of HRT in VFCW.

The SA of VFCW shows that HLR is the most influential parameter having a negative correlation with log removal (Figure 38). Furthermore, in Figure 39 we can observe that influent concentration plays a somewhat important role in pathogen removal while having a positive relation. As can be observed in the same figure, the difference between 1 day of HRT and 16 days is very small, which suggests that the effect of HRT in log removal is almost negligible. Figure 40 confirms that suggestion since log removal values slightly decrease with HRT between 1-3 days and after that, it remains at a stable level regardless of the increment of HRT.

Looking at the three models and their respective SA a lot of conclusions can be drawn regarding the three types of CW:

1. Regarding the FWS CW type, both HLR and Cin have a great impact on log removal, whereas HRT is not playing an important role since it is not included in the equation. Moreover, Figure 34 shows that FWS can be highly efficient in terms of log removal when there are high concentrations of pathogens in the influent wastewater.

2. Although SSHFCW uses all three of the initial parameters, Cin seems to play a negligible role in log removal whereas HLR is important only in small values. Moreover, HRT is the main determinant of the removal of pathogens meaning that the longer the pathogen stays inside the CW the better the chances of removal. Overall it can be observed that the removal capacity of SSHFCW is smaller compared to FWS.
3. The VFCW type shows a different profile compared to SSHFCW (although they are both subsurface systems) since here HRT is the negligible parameter while Cin appears to affect to a certain extent the pathogen removal. The crucial parameter in this type is HLR which shows that with a typical value ranging from 0.005-0.05 m/d there can be high levels of removal, which however decrease dramatically with the increase in HLR.

The sensitivity analysis gave an insight into the contribution of each parameter on the removal as well as the levels of the latter with default values given on Cin, HRT, and HLR. Moreover, the results can be described reliable, since the developed models although they originate from a specific dataset, they are not heavily dependent on the that. That is because, during the implementation of the regression analysis, a cross-validation technique[13] was performed. That means that the selected models have already tested in ''practice'' by means of using an independent dataset for the estimation of their predictive values. This is a common problem associated with predictive model development but in this case, the cross-validation technique gave promising results, therefore we can be confident that the results are reliable.

Overall, the creation of the dataset can be used as a benchmark for any new relevant research, since it covers a wide range of information on the properties and performances of constructed wetlands on a global scale. Moreover, it can be a key source of data in the scientific community, in terms of comparing experimental findings and drawing conclusions. Furthermore, the development of three different predictive models is a reliable solution in attempting to unravel the complexity that accompanies these systems, by the means of finding correlations between parameters that influence the removing efficacy of constructed wetlands.

5.3 Limitations

One of the major difficulties encountered in this paper was the lack of important and relevant data from the studies that were included in the analysis. Therefore, a need for more descriptive information from papers is considered imperative especially on papers that are directly related to constructed wetlands and pathogen removal. Another restriction was the difficulty in data extraction. A lot of these papers did not provide extractable information which meant that a lot of assumptions had to be made (e.g default value of porosity, calculation of missing parameters with nominal values). Furthermore, as it was expected there were a lot of limitations on the model itself. In specific, Non-linear empirical models C and D were used with the assumption that the rate constant had the same value for all three pathogen categories throughout the removal process since there was no other way to find decay coefficients for every pathogen category. That is of course not true since, under

[13] Cross-validation is a technique that is mainly used when someone wants to estimate the accuracy of a predictive model in practice. The goal of cross-validation is to test the model's ability to predict new values with data that was not used in estimating it, in order to avoid problems like overfitting and to give an insight on how the model will perform to an independent/unknown dataset (Picard & Cook, 1984).

normal conditions, each pathogen indicator has its own decay coefficient which is determined by various factors such as temperature and HRT. Therefore, this assumption automatically diminishes the reliability of these models.

5.4 Recommendation

Below is a list of recommendations for future research while using the developed dataset as a reference level.

1. An attempt to extend the current dataset by contacting authors of case studies with limited information and requesting additional data will be a good option.
2. The lack of information through the systematic literature review led to the non-use of temperature and porosity as key factors affecting pathogen removal. If enough data can be found, new regression analyses should be made but this time incorporating T and the size of the support medium as additional parameters.
3. Creation of a framework from policymakers which will oblige authors to provide adequate reporting of extractable data to allow for meta-analysis of future studies.

Furthermore, further studies can focus on the difference between pilot-scale and full-scale applications, since different environmental conditions and sizes prevail between them, possibly resulting in different correlations. That implication could possibly explain the huge variation on log removal from the gram-negative bacteria category in Figure 20, as scale-up could potentially play a significant role in the performance of CWs.

6. Conclusion

The purpose of this thesis was to quantify the log reduction of pathogenic microorganisms by constructed wetlands as a basis for QMRA of water reuse applications. The aim was achieved as the final outcome of the paper includes a comprehensive database that covers a wide range of information about constructed wetlands and log removal levels from specific pathogen categories. Moreover, it includes three predictive models that through statistical analysis are able to describe the relationships between specific parameters that were selected through systematic literature review, and finally provide predictive values of pathogen removal for three distinct types of CWs, thus answering the initial research question. An effort was made for the final outcome to cover the knowledge gap by contributing to the efforts of KWR to create a knowledge base associated with pathogen removal through CWs. The created models can be incorporated into the treatment calculator as demonstrated by the storyboard in Chapter 4. Subsequently, the latter can be integrated into the QMRAspot for the implementation of QMRA of specific index pathogens. KWR can use this added value by providing guidance and counsel to end-users like municipalities, local governments, water boards or water companies regarding the polishing ability of CWs for water reuse applications.

The three models created can pave the way for a better understanding of the design and operational parameters of CWs since the decision makers will now be able to know what the required values of HRT and HLR should be, in order to achieve a certain degree of pathogen removal from their CW. Overall, the final outcome does provide an efficient approach to the scientific community by taking a step closer to a better understanding of these ''black boxes'' and pointing out where future research needs to focus, in order to fine-tune and quantify the factors that influence the performance of constructed wetlands.

References

Akratos, C. S., & Tsihrintzis, V. A. (2007). Effect of temperature, HRT, vegetation and porous media on removal efficiency of pilot-scale horizontal subsurface flow constructed wetlands. *Ecological engineering*, *29*(2), 173-191.

Alexandros, S. I., & Akratos, C. S. (2016). Removal of pathogenic bacteria in constructed wetlands: mechanisms and efficiency. In *Phytoremediation* (pp. 327-346). Springer, Cham.

Arias, C. A., Cabello, A., Brix, H., & Johansen, N. H. (2003). Removal of indicator bacteria from municipal wastewater in an experimental two-stage vertical flow constructed wetland system. *Water Science and Technology*, *48*(5), 35-41.

Austin, D. C., & Lohan, E. (2005). Patent: Tidal vertical flow wastewater treatment system and method. *United States US*, *6*, B2.

Avelar, F. F., de Matos, A. T., de Matos, M. P., & Borges, A. C. (2014). Coliform bacteria removal from sewage in constructed wetlands planted with Mentha aquatica. *Environmental technology*, *35*(16), 2095-2103.

Ávila, C., Salas, J. J., Martín, I., Aragón, C., & García, J. (2013). Integrated treatment of combined sewer wastewater and stormwater in a hybrid constructed wetland system in southern Spain and its further reuse. *Ecological Engineering*, *50*, 13-20.

Behrends, L. L., Sikora, F., Coonrod, H., Bailey, E., & Bulls, M. (1996). Reciprocating subsurface-flow constructed wetlands for removing ammonia, nitrate, and chemical oxygen demand: potential for treating domestic, industrial, and agricultural wastewaters. In *Proceedings of the 69th Annual WEFTEC Conference. Water Environment Federation, Alexandria, Virginia*.

Bichai, F., & Smeets, P. W. (2013). Using QMRA-based regulation as a water quality management tool in the water security challenge: experience from the Netherlands and Australia. *Water research*, *47*(20), 7315-7326.

Boutilier, L., Jamieson, R., Gordon, R., Lake, C., & Hart, W. (2009). Adsorption, sedimentation, and inactivation of E. coli within wastewater treatment wetlands. *Water Research*, *43*(17), 4370-4380.

Brix, H. (1987). Treatment of wastewater in the rhizosphere of wetland plants–the root-zone method. *Water Science and Technology*, *19*(1-2), 107-118.

Brix, H. (1990). Gas exchange through the soil-atmosphere interphase and through dead culms of Phragmites australis in a constructed reed bed receiving domestic sewage. *Water Research*, *24*(2), 259-266.

Brown, C. (2004). Emerging zoonoses and pathogens of public health significance--an overview. *Revue scientifique et technique-office international des epizooties*, *23*(2), 435-442.

Camacho, J. V., Vico, M. D. C. M., Andrés, M., Rodrigo, R., Morales, F. J. F., & Cañizares, P. C. (2014). Energy production from wastewater using horizontal and vertical subsurface flow constructed wetlands. *Environmental Engineering and Management Journal*, *13*(10), 2517-2523.

Collivignarelli, M., Abbà, A., Benigna, I., Sorlini, S., & Torretta, V. (2017). Overview of the main disinfection processes for wastewater and drinking water treatment plants. *Sustainability*, *10*(1), 86.

Cooper, P. F., Job, G. D., Green, M. B., & Shutes, R. B. E. (1996). Reed beds and constructed wetlands for wastewater treatment. European water pollution control, 6(7), 49.

Cooper, P., Griffin, P., Humphries, S., & Pound, A. (1999). Design of a hybrid reed bed system to achieve complete nitrification and denitrification of domestic sewage. *Water science and technology*, *40*(3), 283-289.

Decamp, O., & Warren, A. (1998). Bacterivory in ciliates isolated from constructed wetlands (reed beds) used for wastewater treatment. *Water Research*, *32*(7), 1989-1996.

Dotro, G., Langergraber, G., Molle, P., Nivala, J., Puigagut, J., Stein, O., & Von Sperling, M. (2017). *Treatment wetlands* (Vol. 7). London, UK: IWA publishing.

Erlandsen, S. L. (1994). Biotic transmission—Is giardiasis a zoonosis. *Giardia: From Molecules to Disease. Wallingford, UK: CAB International*, 83-97.

Fewtrell, L., & Bartram, J. (Eds.). (2001). *Water quality: guidelines, standards & health*. IWA publishing.

Flood, J. (2000). Virus-sized particles can be entrapped and concentrated one hundred fold within wetland biofilms. *Advances in Environmental Research*, *3*(4), 403-411.

García, J., Vivar, J., Aromir, M., & Mujeriego, R. (2003). Role of hydraulic retention time and granular medium in microbial removal in tertiary treatment reed beds. *Water Research*, *37*(11), 2645-2653.

Gersberg, R. M., Brenner, R., Lyon, S. R., & Elkins, B. V. (1987). Survival of bacteria and viruses in municipal wastewaters applied to artificial wetlands.

Gonzales-Gustavson, E., Rusiñol, M., Medema, G., Calvo, M., & Girones, R. (2019). Quantitative risk assessment of norovirus and adenovirus for the use of reclaimed water to irrigate lettuce in Catalonia. *Water research*, *153*, 91-99.

González, M. B., Brugnoni, L. I., Flamini, D. O., Quinzani, L. M., & Saidman, S. B. (2018). Removal of Escherichia coli from well water using continuous laminar flow in a channel system containing PPy/Cu modified electrodes. *Journal of water and health*, *16*(6), 921-929.

Green, M. B., Griffin, P., Seabridge, J. K., & Dhobie, D. (1997). Removal of bacteria in subsurface flow wetlands. *Water Science and Technology*, *35*(5), 109-116.

Hamaamin, Y. A., Adhikari, U., Nejadhashemi, A. P., Harrigan, T., & Reinhold, D. M. (2014). Modeling Escherichia coli removal in constructed wetlands under pulse loading. *Water research*, *50*, 441-454.

Headley, T., Nivala, J., Kassa, K., Olsson, L., Wallace, S., Brix, H., ... & Müller, R. (2013). Escherichia coli removal and internal dynamics in subsurface flow ecotechnologies: Effects of design and plants. *Ecological engineering*, *61*, 564-574.

Hrudey, S. E., & Hrudey, E. J. (2004). *Safe drinking water*. IWA publishing.

Hunt, P. G., Matheny, T. A., & Szögi, A. A. (2003). Denitrification in constructed wetlands used for treatment of swine wastewater. *Journal of environmental quality*, *32*(2), 727-735.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.

Ibekwe, A. M., Grieve, C. M., & Lyon, S. R. (2003). Characterization of microbial communities and composition in constructed dairy wetland wastewater effluent. *Appl. Environ. Microbiol.*, *69*(9), 5060-5069.

Jamieson, R., Gordon, R., Wheeler, N., Smith, E., Stratton, G., & Madani, A. (2007). Determination of first order rate constants for wetlands treating livestock wastewater in cold climates. *Journal of Environmental Engineering and Science*, *6*(1), 65-72.

Kadlec, R. H. (1995). Overview: surface flow constructed wetlands. *Water Science and Technology*, *32*(3), 1-12.

Kadlec, R. H. (1997). Deterministic and stochastic aspects of constructed wetland performance and design. *Water Science and Technology*, *35*(5), 149-156.

Kadlec, R. H. (1999). The limits of phosphorus removal in wetlands. *Wetlands Ecology and Management*, *7*(3), 165-175.

Kadlec, R. H., & Knight, R. L. (1996). Treatment wetlands. Lewis Publ., Boca Raton. *Treatment wetlands. Lewis Publ., Boca Raton.*

Kadlec, R. H., & Wallace, S. (2008). *Treatment wetlands*. CRC press.

Kadlec, R. H., Knight, R. L., Vymazal, J., Brix, H., Cooper, P., & Haberl, R. (2000). Constructed wetlands for pollution control. IWA scientific and technical report 8.

Karim, M. R., Manshadi, F. D., Karpiscak, M. M., & Gerba, C. P. (2004). The persistence and removal of enteric pathogens in constructed wetlands. *Water research*, *38*(7), 1831-1837.

Kassenga, G., Pardue, J. H., Moe, W. M., & Bowman, K. S. (2004). Hydrogen thresholds as indicators of dehalorespiration in constructed treatment wetlands. *Environmental science & technology*, *38*(4), 1024-1030.

Kivaisi, A. K. (2001). The potential for constructed wetlands for wastewater treatment and reuse in developing countries: a review. *Ecological engineering*, *16*(4), 545-560.

Knight, R. L., Payne Jr, V. W., Borer, R. E., Clarke Jr, R. A., & Pries, J. H. (2000). Constructed wetlands for livestock wastewater management. *Ecological engineering*, *15*(1-2), 41-55.

Kuschk, P., Wiessner, A., Seeger, E. M., Kästner, M., Kappelmeyer, U., Paredes, D., & Shtemenko, N. I. (2012). The status of research on constructed wetlands. In *Environmental and food safety and security for south-east Europe and Ukraine*(pp. 155-171). Springer, Dordrecht

Langergraber, G., & Šimůnek, J. (2012). Reactive transport modeling of subsurface flow constructed wetlands using the HYDRUS wetland module. *Vadose Zone Journal*, *11*(2).

Langergraber, G., Haberl, R., Laber, J., & Pressl, A. (2003). Evaluation of substrate clogging processes in vertical flow constructed wetlands. *Water Science and Technology*, *48*(5), 25-34.

Lavrnić, S., & Mancini, M. L. (2016). Can constructed wetlands treat wastewater for reuse in agriculture? Review of guidelines and examples in South Europe. *Water Science and Technology*, wst2016089.

Leclerc, H., Mossel, D. A., Trinel, P. A., & Gavini, F. (1977). Microbiological monitoring—A new test for fecal contamination. In *Bacterial indicators/health hazards associated with water*. ASTM International.

Lemon, E., Bis, G., Rozema, L., & Smith, I. (1996). SWAMP pilot scale wetlands: Design and performance, Niagara-on-the-Lake, Ontario, Canada. In *Constructed Wetlands in Cold Climates: Design, Operation, Performance Symposium*.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine*, *6*(7), e1000100.

Liu, R., Zhao, Y., Doherty, L., Hu, Y., & Hao, X. (2015). A review of incorporation of constructed wetland with other treatment processes. *Chemical Engineering Journal*, *279*, 220-230.

López, D., Leiva, A. M., Arismendi, W., & Vidal, G. (2019). Influence of design and operational parameters on the pathogens reduction in constructed wetland under the climate change scenario. *Reviews in Environmental Science and Bio/Technology*, 1-25.

Maier, R. M., Pepper, I. L., & Gerba, C. P. (2009). *Environmental microbiology* (Vol. 397). Academic press.

Marecos do Monte, H., & Albuquerque, A. (2010). Analysis of constructed wetland performance for irrigation reuse. *Water Science and Technology*, *61*(7), 1699-1705.

Mason, C. H., & Perreault Jr, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of marketing research*, *28*(3), 268-280.

McAllister, T. A., & Topp, E. (2012). Role of livestock in microbiological contamination of water: commonly the blame, but not always the source. *Animal Frontiers*, *2*(2), 17-27.

Miller, G. W. (2006). Integrated concepts in water reuse: managing global water needs. *Desalination*, *187*(1-3), 65-75.

Nuttall, P.M., Boon, A.G., Rowell, M.R. (1998) Review of the Design and Management of Constructed Wetlands. CIRIA Publications, London, UK.

Olsson, L. (2011). Effect of design and dosing regime on the treatment performance of vertical flow constructed wetlands.

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, *79*(387), 575-583.

Reed, S. C., Crites, R. W., & Middlebrooks, E. J. (1995). *Natural systems for waste management and treatment* (No. Ed. 2). McGraw-Hill, Inc..

Rousseau, D. P. L., Lesage, E., Story, A., Vanrolleghem, P. A., & De Pauw, N. (2008). Constructed wetlands for water reclamation. *Desalination*, *218*(1-3), 181-189.

Samsó, R., & Garcia, J. (2013). BIO_PORE, a mathematical model to simulate biofilm growth and water quality improvement in porous media: application and calibration for constructed wetlands. *Ecological engineering*, *54*, 116-127

Saxena, G., Bharagava, R. N., Kaithwas, G., & Raj, A. (2015). Microbial indicators, pathogens and methods for their monitoring in water environment. *Journal of water and health*, *13*(2), 319-339.

Schijven, J., Rutjes, S., Smeets, P., & Teunis, P. (2014). QMRAspot: a tool for quantitative microbial risk assessment for drinking water: Manual QMRAspot version 2.0.

Selvakumar, A., Borst, M., & Struck, S. (2007). Microorganisms die-off rates in urban stormwater runoff. *Proceedings of the Water Environment Federation*, *2007*(5), 214-230.

Sgroi, M., Vagliasindi, F. G., & Roccaro, P. (2018). Feasibility, sustainability and circular economy concepts in water reuse. *Current Opinion in Environmental Science & Health*, *2*, 20-25.

Shapiro, O. H., Kushmaro, A., & Brenner, A. (2010). Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *The ISME journal*, *4*(3), 327.

Sharma, G., & Brighu, U. (2014). Performance analysis of vertical up-flow constructed wetlands for secondary treated effluent. *APCBEE procedia*, *10*, 110-114.

Šimunek, J., Van Genuchten, M. T., & Šejna, M. (2012). HYDRUS: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1263-1274.

Slifko, T. R., Smith, H. V., & Rose, J. B. (2000). Emerging parasite zoonoses associated with water and food. *International journal for parasitology*, *30*(12-13), 1379-1393.

Smeets PWMH. 2017. PCD 8, Protocol referentiedocument AMVD, January 2017. KWR, Nieuwegein.

Son, Y. K., Yoon, C. G., Kim, H. C., Jang, J. H., & Lee, S. B. (2010). Determination of regression model parameter for constructed wetland using operating data. *Paddy and Water Environment*, *8*(4), 325-332.

Steel, R. G. D., & Torrie, J. H. (1960). Principles and procedures of statistics. *Principles and procedures of statistics.*

Stone, K. C., Poach, M. E., Hunt, P. G., & Reddy, G. B. (2004). Marsh-pond-marsh constructed wetland design analysis for swine lagoon wastewater treatment. *Ecological Engineering*, *23*(2), 127-133.

Stott, R., & Tanner, C. C. (2005). Influence of biofilm on removal of surrogate faecal microbes in a constructed wetland and maturation pond. *Water Science and Technology*, *51*(9), 315-322.

Stott, R., May, E., Matsushita, E., & Warren, A. (2001). Protozoan predation as a mechanism for the removal of Cryptosporidium oocysts from wastewaters in constructed wetlands. *Water Science and Technology*, *44*(11-12), 191-198.

Sultana, M. Y. (2014). *Treatment of industrial and agro-industrial wastewater using constructed wetlands* (Doctoral dissertation).

Tanner, C. C., Clayton, J. S., & Upsdell, M. P. (1995). Effect of loading rate and planting on treatment of dairy farm wastewaters in constructed wetlands—II. Removal of nitrogen and phosphorus. *Water research*, *29*(1), 27-34.

Toet, S., Van Logtestijn, R. S., Kampf, R., Schreijer, M., & Verhoeven, J. T. (2005). The effect of hydraulic retention time on the removal of pollutants from sewage treatment plant effluent in a surface-flow wetland system. *Wetlands*, *25*(2), 375-391.

Truijen, G., & van der Heijden, P. G. (2013). *Constructed wetland and aquatic treatment systems for fish farms in Egypt: Desk study report*. Centre for Development Innovation, Wageningen UR.

Ulrich, H., Klaus, D., Irmgard, F., Annette, H., Juan, L. P., & Regine, S. (2005). Microbiological investigations for sanitary assessment of wastewater treated in constructed wetlands. *Water research*, *39*(20), 4849-4858.

Van Regenmortel, M. H., Fauquet, C. M., Bishop, D. H., Carstens, E. B., Estes, M. K., Lemon, S. M., ... & Wickner, R. B. (2000). *Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses*. Academic Press.

Verzani, J. (2011). *Getting started with RStudio*. " O'Reilly Media, Inc.".

Vymazal, J. (2005). Horizontal sub-surface flow and hybrid constructed wetlands systems for wastewater treatment. *Ecological engineering*, *25*(5), 478-490.

Vymazal, J. (2005a). Removal of enteric bacteria in constructed treatment wetlands with emergent macrophytes: a review. *Journal of Environmental Science and Health*, *40*(6-7), 1355-1367.

Vymazal, J. (2010). Constructed wetlands for wastewater treatment. *Water*, *2*(3), 530-549.

Vymazal, J., & Kropfelova, L. (2006). Can we predict processes in the filtration bed of horizontal-flow constructed wetlands according to dissolved oxygen concentrations at the outflow. In *Proceedings from the 10th International Conference on Wetland Systems for Water Pollution Control. Lisbon: Portugal*.

Vymazal, J., & Kröpfelová, L. (2008). *Wastewater treatment in constructed wetlands with horizontal sub-surface flow* (Vol. 14). Springer Science & Business Media.

Vymazal, J., Greenway, M., Tonderski, K., Brix, H., & Mander, Ü. (2006). Constructed wetlands for wastewater treatment. In *Wetlands and natural resource management* (pp. 69-96). Springer, Berlin, Heidelberg.

Wallace, S. D., & Knight, R. L. (2006). Small-scale constructed wetland treatment systems: Feasibility, design criteria, and O&M requirements, Final Report. *Water Environment Research Federation, USA*, 304.

Wand, H., Vacca, G., Kuschk, P., Krüger, M., & Kästner, M. (2007). Removal of bacteria by filtration in planted and non-planted sand columns. *Water research*, *41*(1), 159-167

Wang, C., Zhang, M., Ye, M., Wang, J., & Li, G. (2014). Pilot-scale electrochemical oxidation combined with constructed wetland system for unconventional surface water treatment. *Journal of Chemical Technology & Biotechnology*, *89*(10), 1599-1606.

Weber, K. P., & Legge, R. L. (2008). Pathogen removal in constructed wetlands. *Wetlands: Ecology, Conservation and Restoration*, 176-211.

White, S. A. (2013). Wetland technologies for nursery and greenhouse compliance with nutrient regulations. *HortScience*, *48*(9), 1103-1108.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*(1), 79-82.

Winter, K. J., & Goetz, D. (2003). The impact of sewage composition on the soil clogging phenomena of vertical flow constructed wetlands. *Water Science and Technology*, *48*(5), 9-14.

World Health Organization (WHO). (2006). Guidelines for the safe use of wastewater, excreta and greywater: Wastewater use in agriculture (Volume II). *World Health Organization, France. Available at www. whqlibdoc. who. int/publications/2006/9241546832_eng. pdf.(accessed11 April 2009)*.

World Health Organization. (2016). Quantitative microbial risk assessment: application for water safety management.

Links

www.watershare.eu.

Appendixes

Appendix 1

Search engine	Search terms	Searched fields	Number of papers	New papers
Schopus	'constructed wetlands'' AND "pathogen removal"	TITLE-ABS-KEY	50	50
	'constructed wetlands'' AND "pathogen index"	TITLE-ABS-KEY	1	0
	"constructed wetlands" AND "pathogen indicator"	TITLE-ABS-KEY	7	6
	'constructed wetlands" AND "microbes removal"	TITLE-ABS-KEY	1	0
	'constructed wetlands" AND "micro organism" OR "microorganism" AND removal	TITLE-ABS-KEY	611	551
	'constructed wetlands" AND "micro organism" OR "microorganism" AND inactivation	TITLE-ABS-KEY	17	6
	"constructed wetlands" AND "micro organism" OR "microorganism" AND immobilization	TITLE-ABS-KEY	12	3
	"constructed wetlands" AND "micro organism" OR "microorganism" AND "microbial removal"	TITLE-ABS-KEY	10	0
	"constructed wetlands" AND "pathogen inactivation"	TITLE-ABS-KEY	3	1
	"constructed wetlands" AND inactivation AND bacteria	TITLE-ABS-KEY	17	5
	"constructed wetlands" AND inactivation AND virus	TITLE-ABS-KEY	8	3
	"constructed wetlands" AND inactivation AND protozoa	TITLE-ABS-KEY	3	1
	"constructed wetlands" AND index AND bacteria	TITLE-ABS-KEY	54	33
	"constructed wetlands" AND "pathogen removal" AND bacteria	TITLE-ABS-KEY	25	0
	"constructed wetlands" AND "pathogen removal" AND virus	TITLE-ABS-KEY	7	0
	"constructed wetlands" AND "pathogen removal" AND protozoa	TITLE-ABS-KEY	4	0
	"constructed wetlands" AND index AND bacteria OR virus OR protozoa	TITLE-ABS-KEY	57	3
Pubmed	(constructed wetland and pathogen removal)	All fields	72	50
	((constructed wetland* AND (immobilization OR inactivation) AND water AND (microorganism OR bacteria OR protozoa OR virus OR spore)))	All fields	20	8
	((constructed wetland) AND log removal) AND bacteria	All fields	28	11
	((constructed wetland) AND log removal) AND virus*	All fields	5	0
	((constructed wetland) AND log removal) AND microorganism*	All fields	6	2
	((constructed wetland) AND log removal) AND bacteria OR constructed wetland AND log reduction	All fields	24	4
	(constructed wetland*) AND pathogen inactivation AND treatment	All fields	4	0
Sum of Schopus				662
Sum of Pubmed				75
Total number of papers				737

Appendix 1: Total number of papers and the combination of keywords that was used at Schopus and Pubmed databases.

Appendix 2

Title	Type_of_CW	HLR	HRT	[BOD5_remc	COD_removal	TSS_total	Medium	Dimensions_of_	Water_type	Group_Genus	Group	Type_of_s	Method_of_quan	Conce	Concer	Concx	Units	sign	Log_removal
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	Escherichia	Gram_negat	indicator	TBX medium	=	2.44	CHU/100 mL	=	1.29	
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	Faecal Coliform	Gram_negat	indicator	membrane filtratio	=	2.86	CHU/100 mL	=	1.61	CFU/100 mL	=	1.25
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	Intestinal Enter	Gram_negat	indicator	SB agar	=	1.79	CHU/100 mL	=	0.48	CFU/100 mL	=	1.31
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	Somatic coliph	Viruses	indicator	triplicates by envir	=	4.61	CHU/100 mL	=	4.01	CFU/100 mL	=	0.6
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	F-RNA specific	Viruses	indicator	triplicates by envir	=	3.2	CHU/100 mL	=	2.2	CFU/100 mL	=	1
Presence of	VFCW	1.08	0.83	98.034996	92.66247379	0.32954	gravel	L= 29 m, W= 40	SST effluent	phages infectir	Viruses	indicator	triplicates by envir	=	2.57	CHU/100 mL	=	2	CFU/100 mL	=	0.57
Comparative	VFCW	0.0028	4		91.82222222	0.01307	sand+ gravel	25-cm diamete	raw wastew a	Escherichia	Gram_negat	indicator	membrane filtratio	=	6.47	CFU/mL	=	2	CFU/mL	=	4.47
Comparative	VFCW	0.0028	4		91.82222222	0.01307	sand+ gravel	25-cm diamete	raw wastew a	Total Coliform	Gram_negat	indicator	membrane filtratio	=	7.41	CFU/mL	=	2.4	CFU/mL	=	5.01
Comparative	VFCW	0.0028	4		91.82222222	0.01307	sand+ gravel	25-cm diamete	raw wastew a	Faecal Coliform	Gram_negat	indicator	membrane filtratio	=	6.08	CFU/mL	=	0	CFU/mL	=	6.08
Comparative	VFCW	0.0028	4		89.22222222	0.01103	sand+ gravel	25-cm diamete	raw wastew a	Escherichia	Gram_negat	indicator	membrane filtratio	=	6.47	CFU/mL	=	3.04	CFU/mL	=	3.43
Comparative	VFCW	0.0028	4		89.22222222	0.01103	sand+ gravel	25-cm diamete	raw wastew a	Total Coliform	Gram_negat	indicator	membrane filtratio	=	7.41	CFU/mL	=	3.6	CFU/mL	=	3.81
Comparative	VFCW	0.0028	4		89.22222222	0.01103	sand+ gravel	25-cm diamete	raw wastew a	Faecal Coliform	Gram_negat	indicator	membrane filtratio	=	6.08	CFU/mL	=	2.52	CFU/mL	=	3.56
Comparative	VFCW	0.0028	4		91.84444444	0.01865	sand+marble c	25-cm diamete	raw wastew a	Escherichia	Gram_negat	indicator	membrane filtratio	=	6.47	CFU/mL	=	2.54	CFU/mL	=	3.93
Comparative	VFCW	0.0028	4		91.84444444	0.01865	sand+marble c	25-cm diamete	raw wastew a	Total Coliform	Gram_negat	indicator	membrane filtratio	=	7.41	CFU/mL	=	2.59	CFU/mL	=	4.82
Comparative	VFCW	0.0028	4		91.84444444	0.01865	sand+marble c	25-cm diamete	raw wastew a	Faecal Coliform	Gram_negat	indicator	membrane filtratio	=	6.08	CFU/mL	=	0	CFU/mL	=	6.08
Comparative	VFCW	0.0028	4		92.59259259	0.01861	sand+marble c	25-cm diamete	raw wastew a	Escherichia	Gram_negat	indicator	membrane filtratio	=	6.47	CFU/mL	=	2.66	CFU/mL	=	3.81
Comparative	VFCW	0.0028	4		92.59259259	0.01861	sand+marble c	25-cm diamete	raw wastew a	Total Coliform	Gram_negat	indicator	membrane filtratio	=	7.41	CFU/mL	=	3.18	CFU/mL	=	4.23
Comparative	VFCW	0.0028	4		92.59259259	0.01861	sand+marble c	25-cm diamete	raw wastew a	Faecal Coliform	Gram_negat	indicator	membrane filtratio	=	6.08	CFU/mL	=	1.26	CFU/mL	=	4.82
Halophytes a	VFCW	0.095	3.5		78.125		2 layers of grai	L= 1.20 m, W= 0.	primary treat	Total Coliform	Gram_negat	indicator	IDEXX Quanti-Tra	=	6.2	MPN/100mL	=	5	MPN/100mL	=	1.2
Halophytes a	VFCW	0.095	3.5		78.125		2 layers of grai	L= 1.20 m, W= 0.	primary treat	Escherichia	Gram_negat	indicator	IDEXX Quanti-Tra	=	4.2	MPN/100mL	=	2.1	MPN/100mL	=	2.1
Halophytes a	VFCW	0.095	3.5		79.01785714		2 layers of grai	L= 1.20 m, W= 0.	primary treat	Total Coliform	Gram_negat	indicator	IDEXX Quanti-Tra	=	6.2	MPN/100mL	=	5.5	MPN/100mL	=	0.7
Halophytes a	VFCW	0.095	3.5		79.01785714		2 layers of grai	L= 1.20 m, W= 0.	primary treat	Escherichia	Gram_negat	indicator	IDEXX Quanti-Tra	=	4.2	MPN/100mL	=	2.4	MPN/100mL	=	1.8
Emerging org	VFCW	0.044	6.34	91.2	82.94573643	0.09649	1 layer of sand	D=0.8	combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	6.6	CFU/100 mL	=	5.95	CFU/100 mL	=	0.65
Emerging org	FWS	0.06	2.3	94.4	88.75968992	0.11404	siliceous gravel+stones		combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	5.95	CFU/100 mL	=	3.47	CFU/100 mL	=	2.48
Emerging org	FWS	0.02	5.1	95.2	81.78294574	0.05263	siliceous gravel		combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	3.47	CFU/100 mL	<	1.6	CFU/100 mL	<	1.87
Integrated re	VFCW	0.044	6.34	97.455471	91.46567718	0.02787	1 layer of sand	D=0.8	combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	6.6	CFU/100 mL	=	5.18	CFU/100 mL	=	1.424
Integrated re	FWS	0.06	2.3	98.727735	94.43413729	0.02787	gravel		combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	5.176	CFU/100 mL	=	3.55	CFU/100 mL	=	1.626
Integrated re	FWS	0.02	5.1	98.21883	90.72356215	97.9094	gravel		combined se	Escherichia	Gram_negat	indicator	Chromogenic Me	=	3.55	CFU/100 mL	=	4.5	CFU/100 mL	=	-0.95

Appendix 2: A sample of the complete dataset after the systematic literature review was complete.

Appendix 3

Appendix 3: Viruses and protozoan parasites removal performance. N=number of data points.

Pathogen category	Protozoan Parasites			
General information				
Average removal (Log)	1.46			
Number of papers	11			
Number of pilot scale	6			
Number of full-scale	5			
Constructed wetlands				
Types of CWs	*FWS*	*SSHFCW*	*VFCW*	
Pathogen removal	n=20	n=16	n=11	
Log removal	0.95	1.92	1.50	
Minimum Log removal	0.18	0.28	0.35	
Maximum Log removal	2.00	3.63	2.72	
Standar Deviation	0.52	0.91	0.80	
Percentage removal	94.25	99.86	99.17	
Physicochemical characteristics				
Air temperature °C	ND	18 (n=3)	ND	
Water temperature °C	16.5 (n=8)	23.9 (n=4)	9.14 (n=2)	
PH	7.1 (n=9)	7.1 (n=4)	7.05 (n=2)	
COD removal (mg/L)	69.7 (n=5)	86.1 (n=9)	84.8 (n=2)	
BOD removal (mg/L)	74.1(n=4)	94.4 (n=22)	95.97 (n=2)	
BOD5 removal (mg/L)	ND	87.5 (n=9)	ND	
TSS removal (mg/L)	71.5 (n=5)	80 (n=3)	91.2 (n=2)	
HLR range (m/d)	0.04-0.62 (n=13)	0.03-0.62 (n=16)	0.48 (n=2)	
HRT range (d)	0.8-15 (n=13)	0.8-15 (n=16)	0.83 (n=2)	

Pathogen category	Viruses		
General information			
Average removal (Log)	1.33		
Number of papers	10		
Number of pilot scale	5		
Number of full-scale	5		
Constructed wetlands			
Types of CWs	*FWS*	*SSHFCW*	*VFCW*
Pathogen removal	n=18	n=10	n=4
Log removal	1.27	1.59	0.98
Minimum Log removal	0.06	0.02	0.57
Maximum Log removal	2.88	3.62	1.75
Standar Deviation	0.66	0.90	0.47
Average Percentage removal	98.71	99.79	94.50
Physicochemical characteristics			
Air temperature °C	23.7 (n=6)	17.8 (n=4)	ND
Water temperature °C	22.6 (n=10)	21.8 (n=5)	15.4 (n=3)
PH	7.31 (n=10)	7.1 (n=4)	7.83 (n=3)
COD removal (mg/L)	ND	70.2 (n=3)	92.6 (n=3)
BOD removal (mg/L)	ND	ND	ND
BOD5 removal (mg/L)	ND	69.2 (n=3)	98.03 (n=3)
TSS removal (mg/L)	ND	80 (n=3)	67.04 (n=3)
HLR range (m/d)	0.014-5.1 (n=15)	0.014-5.1 (n=9)	1.08 (n=3)
HRT range (d)	0.29-15 (n=14)	0.29-15 (n=9)	0.83 (n=3)

Appendix 4

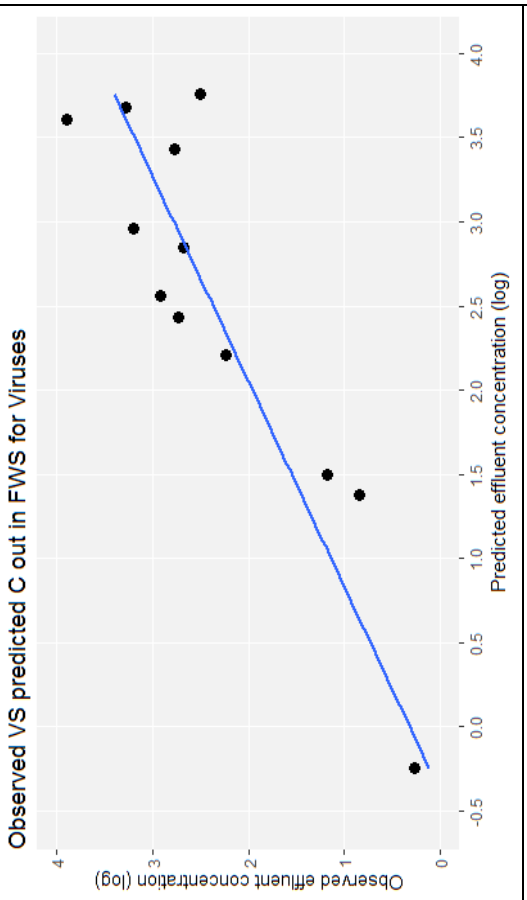
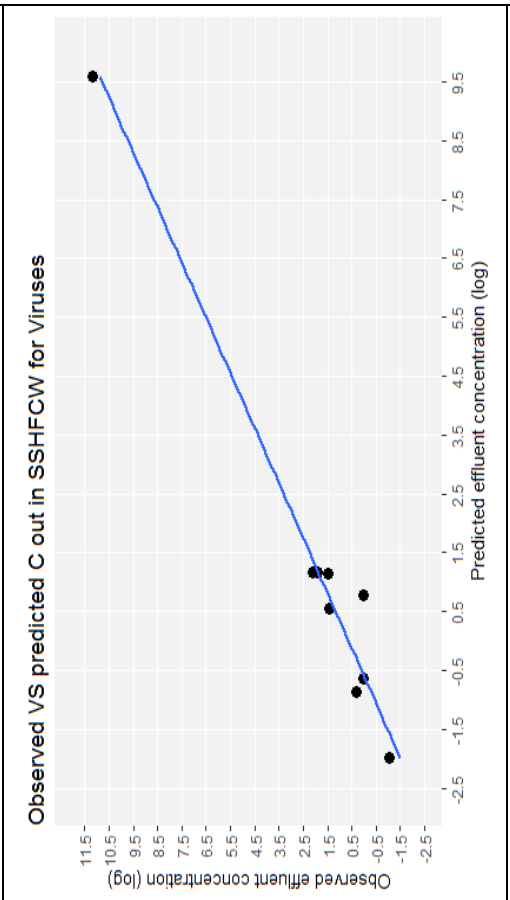
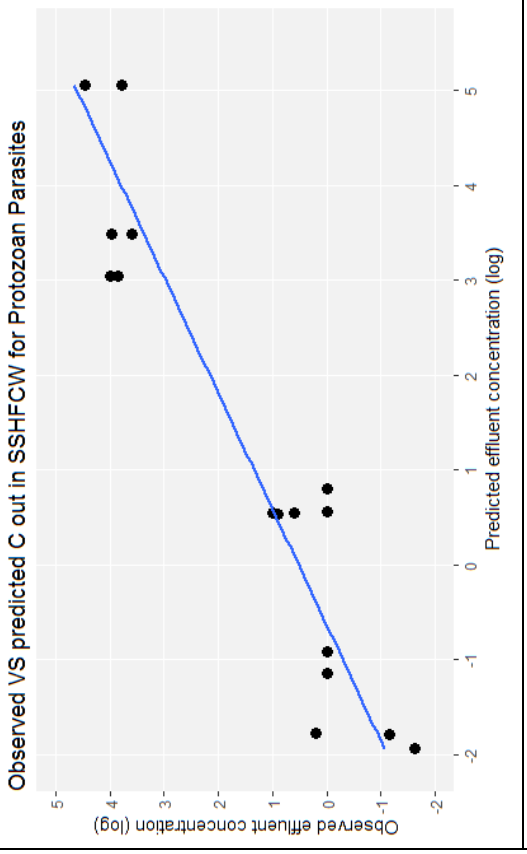
Linear Statistical Model log (C out)= a+b*log(C in)+c*log(HRT)+d*log(HLR)	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	-1.51	1.73*e^-06	0.86	0.99	0.75
	b	1.02	<2*e^-16			
	c	-1.62	7.2*e^-14			
	d	-0.31	0.006			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	-0.78	0.12	0.6	1.22	0.98
	b	0.93	2.64*e^-15			
	c	0.45	0.13			
	d	1.19	4.54*e^-07			
Non-linear Statistical Model A log(C out)=log(a)+b*log(C *in*)+c*log(HLR)+d*log(HRT)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.81	0.01	0.55	1.18	1.04
	b	0.76	<2*e^-16			
	c	-0.01	0.96			
	d	-1.24	0.04			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.21	0.0013	0.78	1.19	0.90
	b	1.02	<2*e^-16			
	c	-0.31	0.0066			
	d	-1.62	7.12*e^-14			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.45	0.05	0.72	1.06	0.85
	b	0.93	2.64*e^-15			
	c	1.19	4.54*e^-07			
	d	0.45	0.13			
Non-linear Statistical Model B log(C out)=log(a)+b*log(C in)+c*log(HLR)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.69	0.01	0.51	1.24	1.1
	b	0.77	<2*e^-16			
	c	0.40	0.08			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.17	0.004	0.75	1.30	0.91
	b	0.95	<2*e^-16			
	c	-0.25	0.05			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.44	0.05	0.68	1.13	0.89
	b	0.94	2.89*e^-15			
	c	1.04	7.27*e^-07			

Non-linear Statistical Model C log(C out)=log(a)+log(C *in*)+(b/HLR)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.14	1.8*e^-08	0.59	1.24	0.92
	b	0.004	0.36			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.12	<2*e^-16	0.77	1.27	0.86
	b	0.006	7.7*e^-10			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.16	9.97*e^-09	0.66	1.13	0.92
	b	-0.007	3.04*e^-10			
Non-linear Statistical Model D log(C out)=log(a)+log(C in)+(b*HRT)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.22	0.0018	0.56	1.27	0.94
	b	-0.16	0.16			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.39	<2*e^-16	0.78	1.18	0.95
	b	-0.18	<2*e^-16			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.08	0.0019	0.64	1.19	1.02
	b	0.02	0.74			

Appendix 4: Residuals of statistical regression analysis through different models per type of CW using the gram-negative bacteria dataset.

Appendix 5

Appendix 5: Observed VS predicted Cout using viruses and protozoan parasites datasets. CDB=Could not be determined.

viruses			**CDB**
protozoan parasites		**CDB**	For FWS see Figure 22

Appendix 6

Linear Statistical Model log (C out)=a+ b*log(C in)+c*log(HRT)+d*og(HLR)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	-0.15	0.5	0.74	0.99	0.78
	b	0.75	<2*e^-16			
	c	-0.8	0.00178			
	d	0.17	0.3388			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	-1.23	3.91*e^-06	0.88	0.98	0.75
	b	0.99	<2*e^-16			
	c	-1.49	<1.04*e^-14			
	d	-0.23	0.0241			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	-0.82	0.0678	0.63	1.19	0.94
	b	0.94	<2*e^-16			
	c	0.45	1.08*e^-07			
	d	1.20	0.1150			
Non-linear Statistical Model A log(C out)=log(a)+b*log(C *in*)+c*log(HLR)+d*log(HRT)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.85483	0.000103	0.78	0.85	0.66
	b	0.75431	<2*e^-16			
	c	0.17099	0.338839			
	d	-0.80	0.017812			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.29	0.000161	0.89	0.97	0.72
	b	0.99	<2*e^-16			
	c	-0.23	0.02			
	d	-1.49	1.04*e^-14			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.43	0.02	0.78	0.92	0.76
	b	0.94	<2*e^-16			
	c	1.2	1.08*e^-07			
	d	0.45	0.11			
Non-linear Statistical Model B log(C out)=log(a)+b*log(C in)+c*log(HLR)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.73	9.12*e^-05	0.78	0.84	0.66
	b	0.75	<2*e^-16			
	c	0.33	0.0481			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.14	0.000439	0.85	1.14	0.90
	b	0.96	<2*e^-16			
	c	-0.12	0.012764			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.41	0.029	0.73	1.002	0.83
	b	0.95	<2*e^-16			
	c	1.05	1.7*e^-07			

Non-linear Statistical Model C log(C out)=log(a)+log(C *in*)+(b/HLR)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.19	1.78*e^-12	0.77	1.1	0.81
	b	0.002211	0.652			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.12	<2*e^-16	0.88	1.02	0.72
	b	0.006	2.32*e^-10			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.17	1.44*e^-09	0.74	0.97	0.78
	b	-0.007	4.77*e^-11			
Non-linear Statistical Model D log(C out)=log(a)+log(C in)+(b*HRT)	FWS	Coefficients	P value	R Squared	RMSE	MAE
	a	0.26	4.72*e^-06	0.77	1.1	0.77
	b	-0.12	0.103			
	SSHFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.36	<2*e^-16	0.88	1.04	0.82
	b	-0.16	<2*e^-16			
	VFCW	Coefficients	P value	R Squared	RMSE	MAE
	a	0.10	0.000845	0.38	1.61	1.28
	b	-0.007	0.91			

Appendix 6a: Residuals of statistical regression analysis through different models per type of CW.

FWS	RMSE	Rsquared adjusted	MAE	Correlation	Rank
Linear statistical model	0.99	0.74	0.78	0.86	5*
Non-linear statistical model A	0.85	0.78	0.66	0.86	2*
Non-linear statistical model B	0.84	0.78	0.66	0.85	1✓
Non-linear empirical P-C-K model C	1.1	0.77	0.81	0.84	4*
Non-linear empirical P-C-K model D	1.1	0.77	0.77	0.84	3*

VFCW	RMSE	Rsquared adjusted	MAE	Correlation	Rank
Linear statistical model	1.19	0.63	0.94	0.79	4*
Non-linear statistical model A	0.92	0.78	0.76	0.79	1*✓
Non-linear statistical model B	1.002	0.73	0.83	0.78	3
Non-linear empirical P-C-K model C	0.97	0.74	0.78	0.81	2
Non-linear empirical P-C-K model D	1.61	0.38	1.28	0.66	5*

Appendix 6b: Performance comparison between the different models per CW. The * sign means that the equation includes a parameter/coefficient that is not statistically significant.