Development and implementation of a sustainability assessment tool for wastewater asset decision-making

by

Jiean Ling

Submitted for the Degree of Practitioner Doctorate in Sustainability (EngD)

Centre for Environment and Sustainability

Faculty of Engineering and Physical Sciences

University of Surrey

Academic supervisors: Prof Richard Murphy, Dr Devendra Saroj
Industrial supervisor: Dr Eve Germain-Cripps

Declaration of Originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

Some parts of the thesis contain content (such as data) of my own publication, which has been declared and cited appropriately. Several documents in appendices were also previously presented in the doctorate confirmation report, which were part of the assessment of the academic progress at the University of Surrey.

Date: ____ 04/12/2021

Acknowledgement

I want to thank all my supervisors Richard, Devendra and Eve for your continuous support and guidance throughout this research. It has been a great journey with you to see me grow despite some challenges. I also want to thank Jaqi for helping me navigate the project and also my life at the beginning of my project. I have my special gratitude for Deidre (and Laura) for your unwavering dedication to all PDS students and your support for me will never be forgotten.

Thank you to all my friends, and especially my housemates. We had made a great 'home' that is loving and fun. The last 4 years has been so eventful because of you and the time we have spent. Although it is sad we will part, I hope you will all find what you need and want in the next chapter of your lives. I also want to thank Ceri and Yulia for being my EngD buddies. I will never forget all those times we helped each other and ranted about what it took to be a doctorate student. I am grateful for my dear partner, Tom, who has been a rock for me during difficult times. You are the most loving and kindest person I've met. Despite all the challenges, we have become closer to each other. I will always remember our memories and can't wait to create more.

At last, I had unlimited gratitude for my dear parents, for your unconditional love and support, which has empowered me to better and love myself. There were so many nights during the pandemic that I wish I can teleport myself to home and just simply have dinner together. And you have always been the light in the darkest nights. And I love you all.

I have never imagined I could go this far, to complete a doctorate. But I would never ask for my life to turn out any differently. I have never become so committed to one thing and pour my dedication into it. I have grown stronger, wiser with more grit that will serve my life further. Although I am glad I have come to the final moment of the marathon, I wish I will embark on a new exciting journey in the field of sustainability.

Jiean

Abstract

The UK wastewater sector is facing increasing challenges such as regulatory pressure, population growth and climate change when making investment decisions. This has resulted in a growing demand for assessment tools that informs the selection of suitable wastewater treatment processes and technologies from the sustainability perspective. The objective of this research was to develop a sustainability assessment tool for a water company in the UK to compare wastewater treatment processes and inform its investment decisions given its unique combination of challenges and needs.

The development of the assessment tool encompassed several phases with an underlying pragmatic research paradigm. The first stage utilised an exploratory case study to understand the current decision drivers and the organisational context. The case study specifically involved a round of semi-structured interviews with stakeholders and thematic analyses. The findings then informed the methodological design of the assessment tool. A suite of assessment criteria and indicators were selected based on literature review and findings from the case study. Multi-Criteria Decision Analysis (MCDA) was selected as the assessment methodology as it was considered suitable and useful to incorporate the Three-Pillars model of sustainability. The assessment methodology was applied to two pilot studies to test its feasibility and robustness. Once the methodology had been confirmed, the assessment methodology was built into a 'tool' with a user interface, culminating in a round of usability testing with end-users in the organisation to examine its overall utility and ease of use. The results of testing suggest the assessment tool is easy to use and understand and offered useful insights into the sustainability credentials of wastewater treatment alternatives. This research also proposed a provisional framework for developing a multi-criteria sustainability tool in a corporate environment, which can be extrapolated for wider applications. The research also demonstrated the significance of pragmatic research in developing a practical solution for industrial-based research whilst highlighting the potential synergy between sustainability assessment, MCDA and decision support systems. The findings and insights of this research will accelerate the practical integration of MCDA into the corporate decision-making process for performing sustainability assessments.

Table of Contents

List of Figures

List of Tables

List of appendices

List of Acronyms

AD	Anaerobic digestion
AM	Arithmetic mean
AHP	Analytical Hierarchy Process
AIJ	Aggregation of Individual Judgements
AIP	Aggregation of Individual Priorities
AMP	Asset Management Plan (period)
ASP	Activated Sludge Process
B-ASP	Ballasted Activated Sludge Process
BOD	Biological oxygen demand
Capex	Capital expenditure
CAPS	Chemically Assisted Primary Settlement
CBA	Cost-Benefit Analysis
CI	Composite indicator/index
C.I.	Consistency Index
C.R.	Consistency Ratio
CHP	Combined heat and power
COD	Chemical oxygen demand
DAF	Dissolved Air Flotation
De-ammo	Mainstream De-ammonification process
Defra	Department for Environment, Food and Rural Affairs
DSS	Decision support system
DST	Decision support tool
EF	Ecological footprint
ELECTRE	ÉLimination et Choix Traduisant la REalité
EU	European Union
G-ASP	Granular Activated Sludge Process
GHG	Greenhouse gas
GLA	Great London area
GM	Geometric mean
LCA	Life Cycle Assessment

LCC	Life Cycle Costing
MAVT	Multi-Attribute Value Theory
MCDA	Multi-Criteria Decision Analysis
MCDST	Multi-Criteria decision support tool
MDG	Millennium Development Goals
NIS	Negative ideal solution
OECD	Organisation for Economic Co-operation and Development
Ofwat	The Water Services Regulation Authority
Opex	Operational expenditure
PAC	Polyaluminium chloride
p.e.	Population equivalent
PIS	Positive ideal solution
PR19	Price Review 2019
R.I.	Random Consistency Index
SAW	Simple Additive Weighted (model)
SBR	Sequencing Batch Reactor
SDG	Sustainable Development Goals
SMART	Simple Multi-Attribute Rating Technique
STW	Sewage treatment works
THP	Thermal hydrolysis process
TN	Total nitrogen
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
TP	Total phosphorus
TSS	Total suspended solid
VFA	Volatile fatty acids

Chapter 1 Introduction

1.1 Research background

1.1.1 Rationale

Wastewater treatment processes have been going through development in recent years, in response to population growth in our cities, tighter environmental permits and the imperative to reduce the energy and cost of the operations. Novel treatment processes and technologies are being developed, tested, and deployed. However, it is important that other sustainability criteria are considered alongside the financial costs when comparing new and current treatment processes and technologies. This project aims to ensure sustainability assessments are performed so that decisions on future investments can be made considering the holistic aspects of sustainability. From a business perspective, this project seeks to support strategic decision-making and the future direction of sustainable wastewater treatment in Thames Water Utilities.

This research project is part of the Practitioner Doctorate in Sustainability programme of the Centre for Environment and Sustainability (CES) at the University of Surrey, in partnership with Thames Water Utilities. The project was initiated by the Research, Development and Innovation Department in Thames Water Utilities and final research outputs or deliverables are owned by Thames Water Utilities. The nature of the research requires the doctoral practitioner to undertake their sustainability research inside the company to deliver the research project, with consideration over the specific needs and requirement of the organisation.

1.1.2 Thames Water Utilities

Thames Water Utilities is the largest water and sewerage company in the UK providing essential services for 15 million customers in London and the Thames Valley region (Thames Water Utilities Limited, 2019a) (Figure 1). Its operation revolves closely around the local hydrological cycle, supplying water to customers from mostly surface water sources and discharging back to the local environment once used water is safely treated (Figure 2). The business covers a wider wastewater service area than its water supply area. Approximately 2.7 billion litres of water is supplied to its customers and 4.4 billion litres of wastewater is treated in 351 sewage treatment works (STW) on a daily basis (Thames Water Utilities Limited, 2018a). Wastewater treatment is thus a critical process that recycles used water from a large number of customers and prevents it from adversely impacting the local environment.

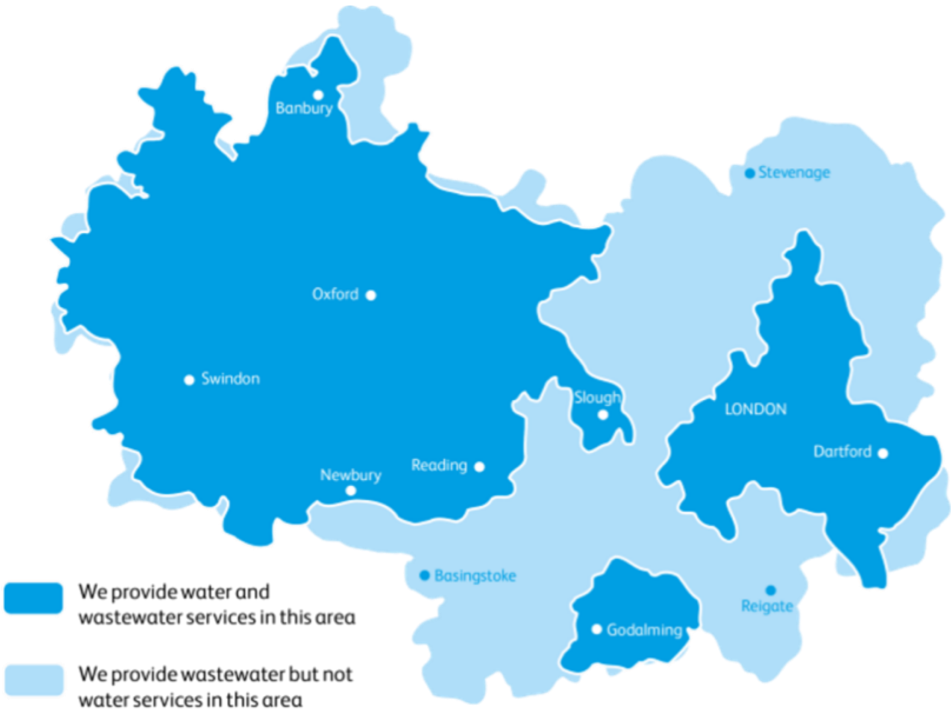


Figure 1. Water and wastewater service area provided by Thames Water (Thames Water Utilities Limited, 2019a)

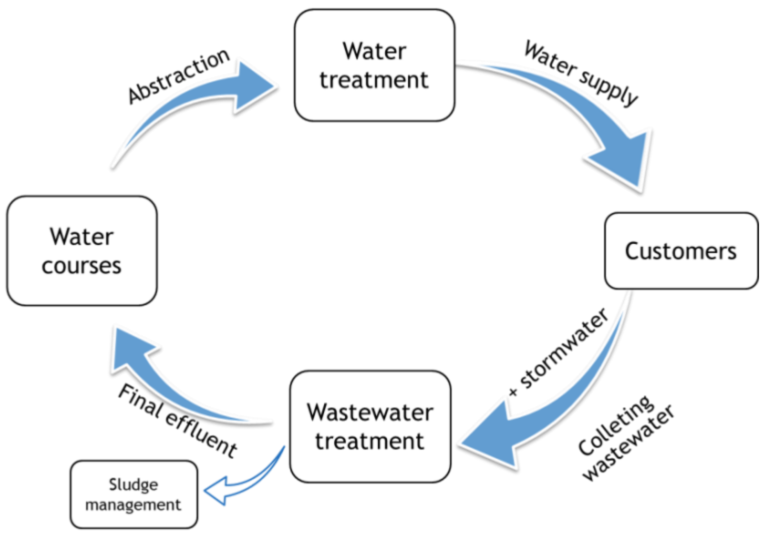


Figure 2. Illustration of water and wastewater cycle as part of company operation

1.1.3 Wastewater treatment process and operation

Wastewater treatment aims to remove hazardous substances and pathogens from sewage and restore water quality. Wastewater is a collective term including surface runoff (i.e. stormwater) and used water from human activities. The majority of wastewater is derived from domestic uses including blackwater (i.e. toilet flushing, human faeces and urine) and greywater (i.e. household washing and bathing water). Wastewater from urban households is usually transported to a centralised STW through sewer network and treated with a combination of physical, biological and chemical treatment processes. The major pollutants to be removed by the treatment process are solids, organic matter and nutrients such as phosphorus and nitrogen. Biological Oxygen Demand (BOD) is used to quantify the level of oxygen required to biologically decompose organic pollutants, indicating the level of organic contamination in the wastewater (Akcin *et al.*, 2005). Excessive nitrogen and phosphorus level in the effluent can contribute to eutrophication and the proliferation of algal blooms (Conley *et al.*, 2009; Bunce *et al.*, 2018). The ecological quality of watercourses in the region is subject to statutory monitoring and scrutiny by the Environment Agency. Therefore, meeting local effluent standards is a key operational objective of the treatment processes.

The operation of wastewater treatment processes requires a continuous input of resources (Figure 3). Although microbial activities play a leading role in removing the majority of pollutants especially in the secondary treatment process, energy is required for pumps, aeration and additional treatment technologies. Chemicals dosing such as coagulants are commonly added to the process for phosphorus removal, and sometimes for enhancing solids removal (Bunce *et al.*, 2018). New processes and technologies are being trialled by Thames Water Utilities to improve treatment efficiency and meet tighter effluent standards. For example, Anammox (Anaerobic ammonium oxidation) bacteria provides a 'shortcut' to convert ammonium to nitrogen gas compared to the conventional Nitrification-Denitrification process. This biological process has the potential of significantly reducing energy requirements for aeration, chemical consumption, and hence operational costs (Thames Water Utilities Limited, 2017c). Anammox has already been implemented at full scale for its feasibility in sidestream treatment[1] and then trialled at the pilot scale for secondary mainstream treatment[2] to determine its feasibility.

There are also opportunities for energy and resource recovery in wastewater treatment. Sewage sludge is a by-product of the wastewater treatment process that contains calorific value and nutrients. Anaerobic digestion (AD) is a widely practised process to treat sludge while producing methane-rich biogas which can be utilised for power generation through a combined heat and power (CHP) unit. Advanced AD technology, such as the Thermal Hydrolysis Process (THP), has been implemented by Thames Water Utilities to increase the digestibility of sewage sludge (Thames Water Utilities Limited, 2016). THP demonstrated multiple benefits including increased biogas yield, lower total solid mass and hence lower cost related to sludge disposal (Pickworth et al., 2006; Mills et al., 2014). Approximately 286 GWh of electricity was generated from sewage sludge last year by Thames Water Utilities in 2017 and this was used to offset part of operational energy consumption at its STWs (Thames Water Utilities Limited, 2017b). Digested sludge can be further dewatered to form biosolids. This is a sterilised and nutrient-rich resource that can be traded as fertiliser or soil conditioner for agriculture and land applications (Mills *et al.*, 2014; Thames Water Utilities Limited, 2018b).

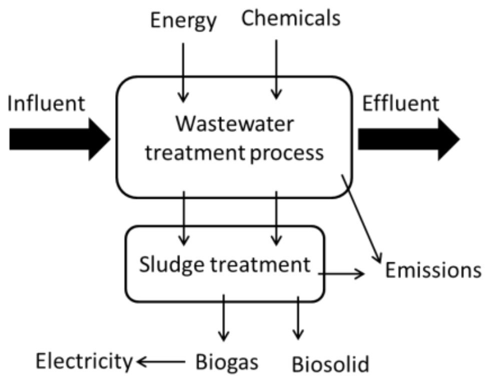


Figure 3. A system view of resources input and output for a typical sewage treatment work

There have been several other commercial technologies developed to improve the efficiency and effectiveness of various stages of treatment processes. However, in order to fully understand the 'net' benefit and impact of these technologies at a STW site level or the company level, there is a growing need for undertaking analyses to assess and compare the impact of different treatment processes and technologies before making investment decisions.

[1] Sidestream refers to the liquor from sludge dewatering and handling

[2] Mainstream refers to the main wastewater treatment process including pre-treatment, secondary treatment and tertiary treatment.

1.1.4 The regulatory context

As a privatised utility company, multiple stakeholders play a critical role in influencing and driving the wastewater strategy of Thames Water Utilities (such as those in Figure 4). Governmental bodies and international authorities provide the regulatory and legislative frameworks within which wastewater companies operate. Specifically, all sewerage companies in the UK were required to comply with the EC Wastewater Treatment Directive (1991) in terms of treatment potential and effluent qualities (Table 1). After Brexit, it is likely that the environmental standards will remain mostly unchanged in the short term (Jacquelyn and Forwood, 2021) but with the possibility of becoming more stringent. In addition, the Environmental Agency also has established quality consents for the effluent to safeguard the ecological status of regional catchments. Financial penalties and reputational damages can be incurred if companies fail to comply with the consents. The Department for Environment, Food and Rural Affairs (DEFRA) has also published a 25 Year Environment Plan that sets a greater ambition on safeguarding UK's critical environmental resources and promoting a natural-capital approach (HM Government, 2018). This plan implies the trend of more stringent discharge consents and greater scrutiny on the ecological status in the future.

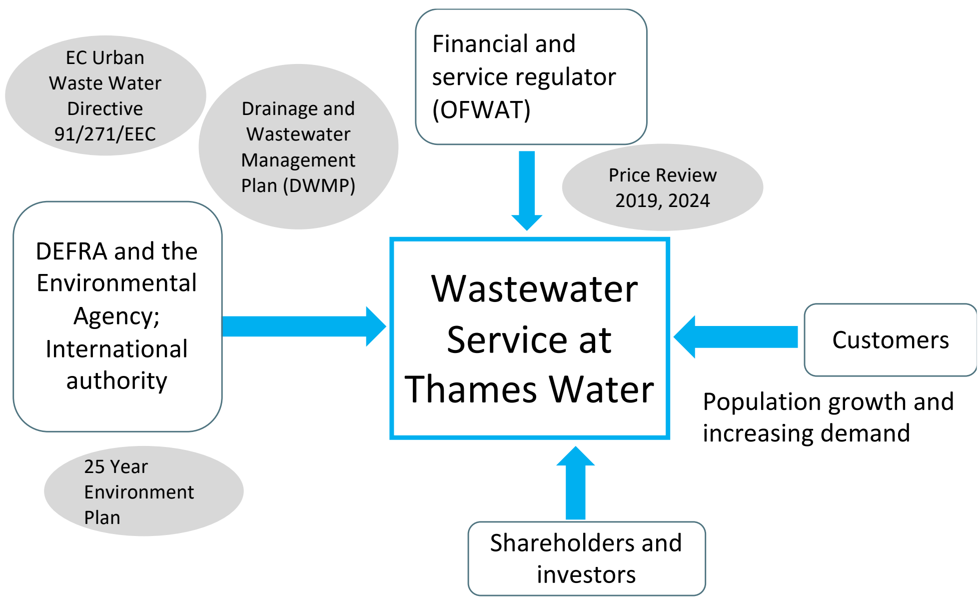


Figure 4. Overview of key external stakeholders of the UK wastewater industry and examples of key legislative frameworks and policies.

Table 1. Wastewater treatment and discharge standards set by the Urban Waste Water Treatment Directive (European Commission, 1991). The standard varies depending on the population equivalent (p.e.) of the STW.

Parameter	Concentration (mg/l)	Removal (%)
BOD	25	70-90
COD	125	75
Total nitrogen (10,000<100,000 p.e.)	15	70-80
Total nitrogen (>100,000 p.e.)	10	
Total phosphorus (10,000<100,000 p.e.)	2	80
Total phosphorus (>100,000 p.e.)	1	

The Water Services Regulation Authority (Ofwat) was established and appointed by the UK Government as a service and financial regulator to review the business plan of privatised water and sewerage companies in England and Wales. Ofwat aims to protect the interest and rights of customers and ensure companies are environmentally cognisant and financially resilient. Ofwat provides an evaluative framework and criteria upon which water companies devise their business plans on a 5-year interval. For the most recent Price Review 2019 (PR19), Ofwat highlighted four overarching themes that guide the business plans from 2020 to 2025 (Ofwat, 2017):

a) Customer services and engagement: Ofwat expects companies to make performance commitments that reflect customers' priorities. The Outcome Delivery Incentive scheme is imposed to financially award companies that outperform its performance targets.
b) Financial, corporate and operational resilience: this includes reducing disruption in daily operation and improving resilience of water supply and wastewater infrastructure while ensuring the financial viability of the services. This requires companies to adopt a system-wide approach to understand and manage risks and opportunities.
c) Affordability: Companies are required to reflect affordability in business planning and support customers who struggle to pay.
d) Innovation: Ofwat encourages companies to look at a wider business boundary and explore new opportunities that deal with the challenges they face.

The Drainage and Management Plan (DWMP) is the new industry-wide framework for the UK water industry to achieve an integrated and long-term planning and management of the collective system of drainage, flooding and wastewater asset for a minimum of 25 years ahead. The development and emergence of DWMP followed the implementation of its sister framework Water Resource Management Plan (WRMP) to complement existing water resource management in the UK. Both DWMP AND WRMP are statutory frameworks that require water companies to present their plan every 5 years (alongside the Price Review business plan) to safeguard the environmental quality, secure long-term water and wastewater services and demonstrate value-for-money for customers and communities (ATKINS, 2019). There are generally 5 stages in the DWMP framework (a detailed procedural process can be found in ATKINS, 2018, P11):

1) Strategic scope and context: to identify long-term challenges and objectives
2) Initial risk-based catchment screening: identify risks and vulnerability in each catchment
3) Catchment risk and vulnerability assessment: analyse and quantify the (environmental, wastewater and drainage) risks that have been identified
4) Options development and appraisal: develop a list of potential options to address the risk and vulnerability
5) Programme appraisal: validate and integrate solutions into a business plan

Since the proposal of DWMP in 2019, Thames Water Utilities has completed multiple rounds of catchment risk and vulnerability analyses and is currently in the process of evaluating a list of different options to address the identified risks (Thames Water Utilities Limited, 2022). Once the final options have been decided following the consultation with customers and business partners, a plan will be put together to be reviewed alongside the next Price Review 2024.

Besides the regulatory pressure, climate change and population growth are emerging long-term investment challenges for Thames Water. There was a working group (London 2100) in the company dedicated to model those impacts and understand the potential risks (Thames Water Utilities Limited, 2017b). The population of the Greater London Area (GLA) is expected to increase from 8.75 million to 15 million by the end of the century (Thames Water Utilities Limited, 2017b). This poses great pressures on meeting the increasing demand for wastewater services due to insufficient capacity and ageing assets. Figure 5 shows that the process capacities of all STWs in the GLA are under pressure to different extents based on this projected population growth (Thames Water Utilities Limited, 2017b). Moreover, some large sites of STWs have very limited space for expansion. The shortfall to meet future demand is compounded by the impact of climate change. Climate change projection shows that an increasing level of winter rainfall is very likely in the coming decades (Fung *et al.*, 2018). Excessive runoff and stormflow can add more pressure on treatment capacity and subsequently increase the risks of pollution incidents and sewer flooding (Zouboulis and Tolkou, 2015).

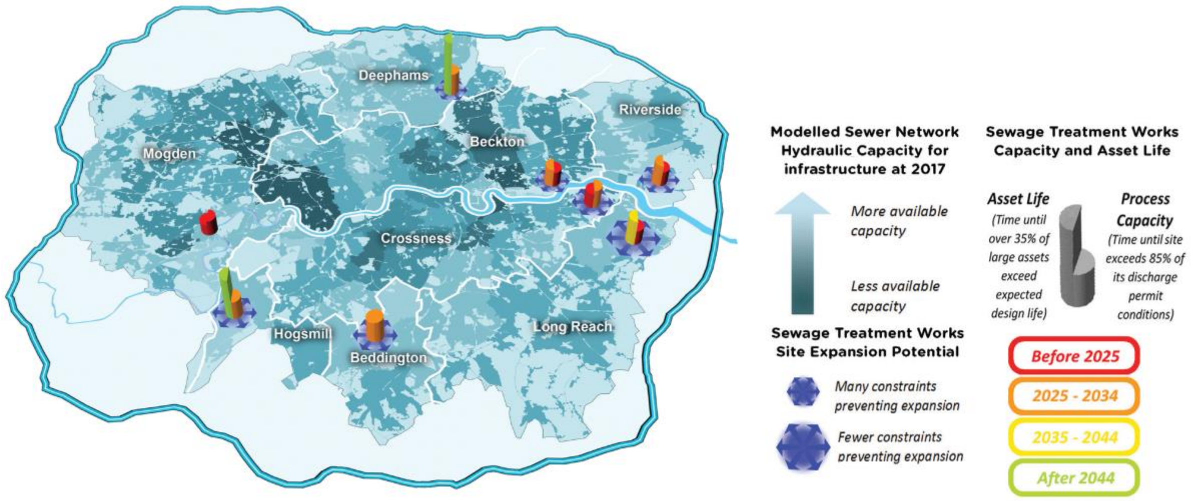


Figure 5. The profile of eight STWs in London highlights that the asset life, process capacities and spare land for expansion of many STWs are under pressure to various extents (Thames Water Utilities Limited, 2017b).

In the light of the emerging long-term challenges and increasing pressures from regulatory bodies, water companies are faced with multiple and complex objectives when selecting wastewater technologies. Despite there are existing work group in the company such as the the London 2100 team and DWMP to investigate the potential impact of population growth and climate change, there is still the potential need for assessment tools that systematically analyses the complexity of the problem while providing practical decision support in terms of selecting the optimal wastewater and sludge treatment processes or technologies at the site level.

1.2 Research objectives and scope

The aim is to develop an assessment tool to identify the best wastewater and sludge treatment technologies from the sustainability perspective, in this case in Thames Water Utilities. This is subdivided into two primary objectives:

1. Develop an assessment approach and tool that evaluates wastewater and sludge treatment technologies/processes from a sustainability perspective.
2. Trial the approach to test its feasibility and usefulness.

The scope of the research is limited to the wastewater sector in the company, typically including the wastewater treatment processes and sludge treatment. The development of this assessment tool did not intend to replace the current decision process. Instead, it sought to inform the current management and decision system as a decision support tool by evaluating and identifying the best alternative.

During the second year of the research project, DWMP was published as a new statutory framework that all water companies were required to comply. The emergence of DWMP influenced the formulation of this research in the following ways. First, DWMP accentuated the demand for resilience asset management planning in the long term. This echoes the purpose of this research to inform and improve wastewater asset decision-making. However, the initial need and purpose of this project were reviewed following the emergence of DWMP. Despite its importance and relevance, it was not practically possible to address the entire DWMP framework in the scope and the life cycle of this research project, and therefore, the purpose of this research was re-defined to provide supplementary analysis for evaluating wastewater treatment options (with lower complexity and risks) that are outside the optioneering scope of DWMP. While DWMP intends to evaluate risk at the catchment and regional scale, this research aims to provide decision aid for wastewater asset planning at the site level. Second, the framework of DWMP strongly encourages the engagement of stakeholders in its process. This was also absorbed and reflected by the decision made on methods selection in this research.

1.3 Identifying the implementation point

Before the design of the interface of the assessment tool, the implementation point has been identified in the decision-making process for wastewater asset investment. There were 5 general phases in the investment delivery process in the water company:

1. identification of needs,
2. project definition,
3. procurement,
4. the construction
5. delivery of the project.

The assessment tool was designed to support the function in the project definition phase where potential options are compared and evaluated. There were two reasons to select this phase for implementation: first, the project definition phase is when all the options are comparatively screened and where a sustainability assessment tool can potentially provide direct decision aid; second, given that the project was initiated within the Asset Planning and Investment department, the tool needed to be aligned with the function and resources in the department. The tool has also been considered to be designed and implemented for the first phase 'identification of needs'; however, there has been already extensive works done by the company in terms of risk and needs modelling (such as in the Wastewater and Drainage Management Plan).

In the project definition phase, the novelty added by the tool is a 'solution-optioneering' function which aims to support the decision-makers to systematically evaluate potential solutions from a holistic list of decision criteria pertaining to sustainability. The output of the assessment will inform the selection of the most desirable wastewater treatment options and provide insights on their sustainability credential to complement other existing analyses (such as risk analysis) in the project definition phase.

1.4 Chapter overview

An overview of individual chapters is shown in Figure 6. Chapter 2 provides a literature review on the concept and approaches of sustainability assessment. Chapter 3 describes the research paradigm, methodology and methods selected for this research in order to develop and implement a sustainability assessment tool. Chapter 4 presents the design, methods and results of a qualitative case study to understand the existing decision drivers and context in Thames Water Utilities. Subsequently, the findings of Chapter 4 provide the underlying foundation and evidence for the development of the assessment tool presented in Chapters 5 to 7. Specifically, Chapter 5 provides a detailed literature review on different multi-Criteria Decision Analysis (MCDA) models to propose a suitable multi-criteria methodology for the assessment tool. Chapter 6 presents the process of developing assessment indicators and weightings for the assessment tool with a pilot study for trialling. Chapter 7 presents some improvements based on the insights from the first pilot study and the second pilot study to further validate the assessment tool. Chapter 8 presents the assembled assessment tool with the user interface and a round of usability testing with the end-users in the company. Chapter 9 discusses the overall significance of the research findings and limitations. For consistency, Thames Water Utilities is referred to as ‘the water company’ in the rest of the thesis.

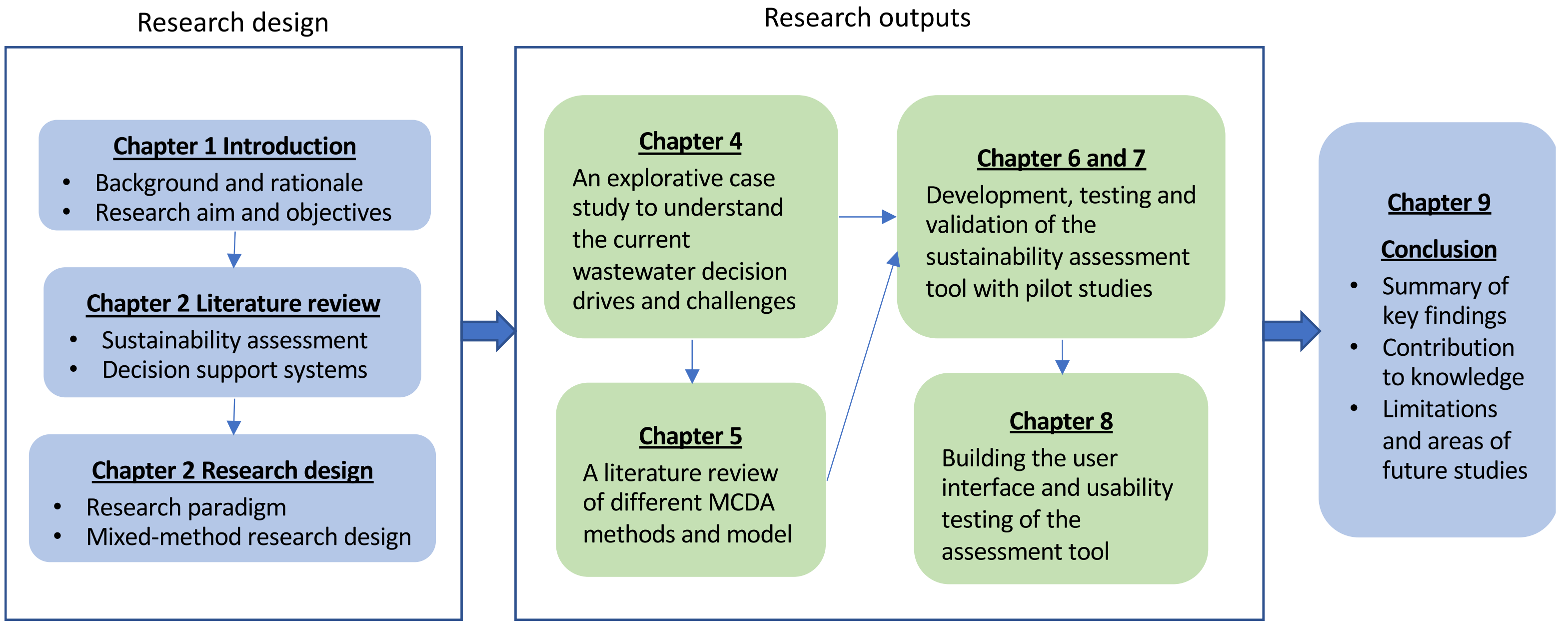


Figure 6. A summary of the content of each chapter and the structure of the thesis.

Chapter 2 Literature review

This chapter provides a review of the literature on two topics of key relevance to the overall research aim '*To develop a sustainability assessment tool to provide decision support for Thames Water*'. These are (1) Sustainability and sustainability assessment; (2) Decision support systems. The objective of this chapter is to establish an understanding of the current state of knowledge to inform the methodological design of this research.

2.1 Concepts of sustainability

The concept of 'Sustainable Development' or 'Sustainability' has gained increasing traction for more than two decades since the publication of the World Commission on Environment and Development (WCED) report '*Our Common Future*' by the United Nations. This report, commonly referred to as the 'Brundtland Report (1987)', provides a widely recognised definition: '*Sustainable development is development which meets the needs of the present without compromising the ability of future generations to meet their own needs'*. This definition placed a strong proposition that the current way of life is unsustainable due to the adverse impacts caused by accelerating population growth, rapid industrialisation, depletion of non-renewable resources and the deterioration of the environment (Meadows *et al.*, 1972). The definition also emphasised the idea of 'intergenerational justice or equity'. The significance of this equity arises because future generations are excluded from participating in the decisions in the current generation but have to inherit the impacts of those decisions on the future ecological and economic capacity (Padilla, 2002).

The launch of the Brundtland Report served as an initiation that invited changes in the global policy landscape. Efforts have been made to interpret sustainability in different contexts (Mebratu, 1998) because its general definition is considered to be too ambiguous and broad to apply to policy-making (Dixon and Fallon, 1989; Mebratu, 1998; Glavič and Lukman, 2007; Giovannoni and Fabietti, 2013; Holden *et al.*, 2014). Holden et al., (2014) suggest that danger of such 'fuzziness' is that it makes sustainability irrelevant in practice (Holden *et al.*, 2014). Redclift (2005) describes the definition in the Brundtland Report as "*deceptive*" because it "*obscures underlying contradictions and complexities*" and hinders its applications. Lafferty (2004, p.24) suggests that the notion of sustainability is similar to democracy: "*it is universally* *desired, diversely understood, difficult to achieve and won't go away*". That is to say, the overarching interpretation of sustainability is continuously debated but this in itself is also a contribution to the shared goal of driving changes. Although the definition from the Brundtland Report emphasised the global and collective acknowledgement of the urgency for change, the challenge remains on how to translate the concept of sustainability/sustainable development into operational principles and a functional framework that is relevant to the organisational context (Johnston et al., 2007).

2.1.1 The Three Pillars of sustainability

One of the widely recognised interpretations of sustainability is based on the Three-Pillar Model or the 'Triple Bottom Line' (Basiago, 1998; Pope *et al.*, 2004; Gibson, 2006; Purvis *et al.*, 2019). It represents the shared priority between the environmental, social and economic dimensions of sustainability (Figure 7a). This model is often referred to as the sustainability Venn diagram. A more sophisticated depiction of the three pillars is the concentric model (Figure 7b) which depicts a nested relationship between them (Montiel, 2008; Ivory and Brooks, 2018). Ainger and Fenner (2013) also positioned the infrastructure sector as an interface between the environment and society because it provides critical service to society and seeks to enable economic activities to operate within environmental limits.

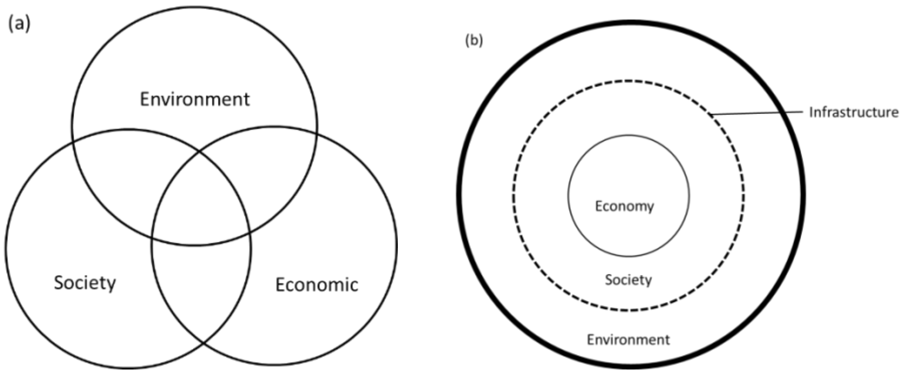


Figure 7. The fundamental 'triple bottom line' concept of sustainability. (a) represents equal importance of the three pillars and (b) represents a 'nested' interrelation (Illustration adapted from Ainger and Fenner, 2013)

Although the conceptualisation of sustainability using the Three Pillars model has been widely applied due to its simplicity, its theoretical origin was unclear (Purvis *et al.*, 2019). As such, the interpretations and applications of this model have been diverse. Such diversity of interpretation is manifested through different definitions of each pillar in the context and the relationship between those pillars. There have been debates over which model provides a more accurate interpretation. The Venn diagram (Figure 7a) suggests that sustainability is the integration of the three pillars which presents sustainability as residing at the intersection between them (Mebratu, 1998). This implies the need to balance the trade-off between "seemingly equally desirable goals" in those pillars (Purvis *et al.*, 2019). However, Elkington *et al.* (2006) argued this is an "*oversimplified delineation*" of the three pillars that fails to capture the complexity among them. The model depicts a "delusion" that human-made capital can compensate for the loss in the natural capital to maintain sustainability. In other words, environmental quality and economic growth are equally valued and complementary (Cole, 1999). The concept of 'capital' refers to the Five Capital Model which includes natural, social, human, financial and manufactured capitals (Porritt, 2012). The definitions of the five capitals are shown in Table 2. Although The 'capitals' often refer to stocks or factors of production in an economy (Ivory and Brooks, 2018), in the context of sustainable development are different types of resources that are available to society at a macro-level (Parkin *et al.*, 2003). And as such, sustainable development can be also defined as a process to achieve sustainable management of those different capital flows (Parkin *et al.*, 2003). It can be argued that some types of capital are derived from two fundamental capital: natural capital and human capital (Parkin *et al.*, 2003). Furthermore, human capital can also be seen as a sub-set of the natural capital because humans exist in the biosphere and rely on the resources and services it provides (Ivory and Brooks, 2018). This reflects the nested systems in the concentric model where each pillar is a subordination of another (Figure 8).

Table 2. The Five-Capital model and the descriptions of each capital (Ekins *et al.*, 1992; Serageldin and Steer, 1994; Parkin *et al.*, 2003)

Name of capital	Descriptions
Natural capital (Environmental or ecological capital)	The 'stock' of resources (e.g. the biota and biomass) and services (e.g. environmental processes) provided by the environment
Human capital	"*Health, knowledge, skills, motivation, and spiritual ease*" of people
Social capital	Cooperative systems where people live and work together such as families, businesses, schools etc.
Manufactured capital	Physical systems and infrastructures built by humans such as roads, houses and machines
Financial capital	A representative of value in other capitals and it has no intrinsic value in itself.

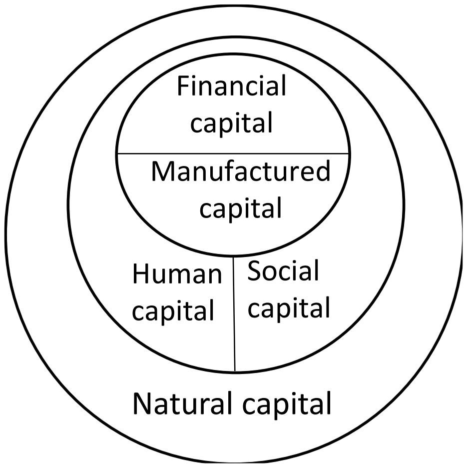


Figure 8. The allocation of the five capitals into the nested sustainability diagram (adapted from Ivory and Brooks, 2018).

The Three-Pillar model (Venn diagram) has been branded as a 'weak' sustainability perspective because the integration of the three pillars is perceived as a 'reductionist' approach. Biely *et al.* (2018) argued that a weak sustainability perspective is invalid because theoretically, it does not lead to a genuine state of sustainability. This is based on the proposition that the stock of natural capital is finite and its loss can be irreversible and irreplaceable. On the contrary, a strong sustainability model requires the condition that human capital cannot substitute natural capital (Redclift, 2005) and it opposes the idea of simplifying the Three-Pillar through integration. The concentric model (Figure 7b), which depicts a strong sustainability perspective, emphasises that the three systems are not directly interchangeable with each other. As such, strong sustainability implies greater importance on the natural capital due to the incommensurability (Ekins *et al.*, 2003). This means monetary compensation and technological innovations should not be the excuse or solution for the loss of natural capital.

Other scholars have questioned the purpose of the debates between weak sustainability and strong sustainability (Jamieson, 1998; Ang and Passel, 2012). First, the differentiation between these two perspectives is fundamentally based on the notion of capital. As capital is defined as the stock that enables the flows of goods and services, there is a strong connotation of "*human needs and interests*" when defining natural capital (Jamieson, 1998). However, ecologists would classify all living matters as part of natural capital because they are integral to the totality of ecosystem. Thus, the proposition of either a strong or weak sustainability perspective is only valid if the concept of natural capital is consistently defined. The second element in the debate is whether the substitution between natural and human-made capitals should be allowed. Holland (1997) argued that the substitutability of capital should depend on its purpose and context. In other words, there is an extent to which substitutability is allowed, rather than the polarities between full substitution and zero substitution. Overall, it is acknowledged that the debate over the interpretation of sustainability will not be resolved easily and soon (Jamieson, 1998). However, this has not diluted the importance of it as a concept and the urgency for its operationalisation.

2.1.2 Sustainable Development Goals

The United Nations has been a key catalyst for accelerating the realisation and implementation of sustainable development globally. The Millennium Development Goals (MDGs), adopted in 2000 along with the United Nation Millennium Declaration, condensed a global action plan and framework to tackle major worldwide challenges. MDGs include 8 goals and 12 targets ranging from reducing hunger, poverty to environmental degradation with measurable and timebound objectives. The onset of MDGs marked a historical moment of uniting and mobilising global efforts to achieve those priorities. Good progress has been made against poverty and diseases (Sachs, 2012). A summary of the progress in terms of health-related targets can be found in the report by World Health Organization (2015, p. 5). However, there seems to be a lack of ambition and efforts in other priorities such as inequality and unemployment (Fukuda-Parr, 2016). Additional, progress has been uneven between developing and developed countries (Caprani, 2016).

Preceded by MDGs in 2015, the Sustainable Development Goals (SDGs) are a set of 17 goals and 169 targets with a stronger focus to integrate into the three pillars of sustainability. The specific goals and descriptions of targets can be found on the UN website (United Nations, n.d.). Compared to MDGs, the development of SDGs was based on consultation with more than 100 countries, compared to MDGs which were mostly determined by the Organisation for Economic Co-operation and Development (OECD) members which are developed countries. The departure of SDGs also marked a shift in emphasis from addressing developing countries to all countries (Allen *et al.*, 2018). SDGs provide greater coverage of global issues and contribute toward a shared and balanced focus on the multiple facets of sustainability for the entire planet (Sachs, 2012). However, the implementation of SDGs also poses challenges. The SDGs framework can be complex as it contains a great number of goals, targets and indicators (Caprani, 2016). The comprehensiveness of SGDs has attracted comments such as "*where everything is a top priority, which means nothing is a priority*" (Easterly, 2015). This potentially hinders a systematic and coherent implementation of SDGs and gives rise to the challenge of informing national practice (Allen *et al.*, 2018). Additionally, the trade-off and synergy between the three pillars of sustainability remain elusive and no consensus has been reached (Sachs, 2012). The Implementation of SDGs at the national level also faces various challenges such as the selective prioritisation of goals by the local government (Carius *et al.*, 2018) and constraints of financial resources (Saxena *et al.*, 2021). Although SDGs marked an unprecedented effort to accelerate the implementation of sustainable development, the progress has been uneven (United Nations, 2019) and it remains challenging to implement the framework at a local and regional scale.

2.1.3 Sustainability principles

Given the focus of this research on wastewater treatment and sustainability-informed decision making in the water company, this section reviews literature associated with the principles in the context of infrastructure projects. Table 3 condenses some of the key principles after Ainger and Fenner (2014) and Hugé *et al.* (2011). Absolute principles reflect the fundamental and widely acknowledged characteristics of sustainable development whilst operational principles are specific guidelines for infrastructure management and planning. These principles are elaborated in the following sub-sections.

Table 3. The absolute and operational principles for sustainable infrastructure. Adapted from Ainger and Fenner (2014) and Hugé *et al.* (2011)

Absolute principle	1. Environment-within limits	2. Social and economic Development	3. Intergenerational stewardship	4. Complexity of infrastructure systems
Operational principles	1a. Measure against environmental limits	2a. Measure against socio-economic goals	3a. Plan for long term	4a. Integrate needs, roles and disciplines
		2b. Engage with stakeholders and decision-makers	3b. Consider a whole life cycle perspective	4b. Consider uncertainty

2.1.3.1 Principle 1: Environment – within limits

The first absolute principle is based on the concept of limits. The concept of limits has been widely discussed in the context of economic and ecological sustainability. Georgescu-Roegen (1973) Suggests there is a physical limit to the economic system. Based on the second law of thermodynamics, in a closed system, available energy is constantly converted to a state with less available energy as the system entropy increases (Rees, 1990). There are two sources of available energy with low entropy: the terrestrial stock of resources and solar energy (Georgescu-Roegen, 1973). Since the inflow of solar energy may not be sufficient to support the exponential growth of economic activities, increasing reliance has to be placed on the use of terrestrial resources to sustain economic throughput. And this will result in a net energy deficit based on the entropy law, which implies an "eventual limit" to economic growth (Ekins, 1993). The ecological limits of the Earth system (sometimes referred to as biophysical limits) can be seen as the ceiling for a "*safe operating space*" for humanity (Rockström *et al.*, 2009; Steffen *et al.*, 2015). The concept of 'Planetary Boundaries' was proposed by Rockström et al. to conceptually, and quantitatively, identify the "tipping points" of the global environmental system. The tipping points refer to the thresholds where human activities have caused systematic, irreversible and catastrophic responses in the Earth system. In order to operationalise the concept of limit, Ainger and Fenner (2014) suggest that it is important to provide measurements against environmental limits. Although the work of Planetary Boundaries was mostly concerned with the global environmental limits, it is a good practice to include the measurement of environmental impacts throughout the whole life stages of an infrastructure project.

2.1.3.2 Principle 2: Social and economic development

The second absolute principle is based on social and economic development. The concept of development and growth has been an inseparable part of sustainable development, which has been widely seen as an overarching policy goal. For example, in the policy paper '*A Better Quality of Life*' (DEFRA, 1999), the UK government proposes four objectives to realise its sustainable development strategy:

1) '*Social progress that recognises the needs of everyone'*

2) *'Effective environmental protection'*
3) *'Prudent use of natural resources'*
4) *'Maintenance of high and stable levels of economic growth and employment'*

These objectives were superseded by the revised priorities in the new Sustainable Development Strategy paper (HM Government, 2005) as:

1) *'Sustainable consumption and production'*
2) *'Climate change and energy'*
3) *'Sustainable communities at local level and'*
4) *'Natural resource protection and environmental enhancement'*

In the more recent UK policy paper (HM Treasury, 2011), the main economic objective was "to achieve strong, sustainable and balanced growth", which entailed ambitions such as building a competitive tax system, attracting investment and exports etc. Other social-related policy papers include the education policy paper to support young people to acquire better access to education and opportunities (Department for Education, 2015), the public health strategy paper (HM Government, 2010). These policy papers present a snapshot of the multidimensionality in the social-economic objectives and the requirement to achieve them simultaneously. To unpack the socio-economic progress holistically, Raworth (2012) proposed a 'Doughnut' model that added social progress insides the environmental limits (Figure 9). The model attempted to quantify the progress made in social criteria such as gender equality, food security, education, health care etc. as part of a sustainable development framework. Similar to the previous principle (1a), the need for social-economic development necessitates the use of indicators to measure and monitor the progress and impacts. An example with a comprehensive indicator system to monitor sustainable development can be found in the report (DEFRA, 2013) containing 12 categories and 23 indicators.

Another operational principle (2b) is to engage with stakeholders throughout the project phase. Given that stakeholders may have different needs and priorities, it is necessary to create a mutual understanding of the project goal, scope and impacts. Millard (2011) and Gallagher et al. (2018) suggests that a participatory learning culture can positively influence the implementation of the three pillars of sustainability. Bal et al. (2013) also demonstrated that in construction projects with a diverse range of stakeholders, engagement is particularly critical for delivering excellent project outcomes and sustainable development objectives.

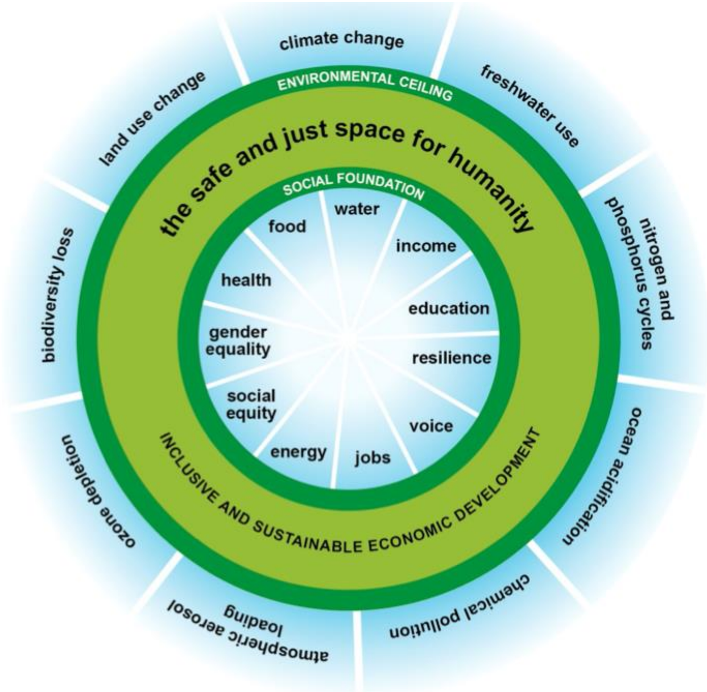


Figure 9. The Doughnut Economics model built on the foundation of Planetary Boundaries (Raworth, 2012)

2.1.3.3 Principle 3: Intergenerational stewardship

This principle reflects one of the key messages from the definition of sustainable development in the Brundtland Report. There is a strong notion of care and responsibility of the current generation and recognition of rights to future generations. Compounded by the irreversibility of losses of some critical natural capital and ecological function, the current decisions should avoid harming the natural legacy of future generations (Padilla, 2002). To be able to understand the negative and positive impacts of an infrastructure project, planning for the long term is a key operational principle (3a). Another operational principle (3b) is to include all life cycles stages of a project so the total sustainability benefits and impacts can be understood. This covers the multiple stages of design, construction, operation and end of life in a project. The significance of the inclusion of a whole life cycle approach is to evaluate the holistic impact of the project and reveal the major areas of concern. Some assessment approaches that adopt a life cycle perspective are Whole Life Costing (WLC) and Life Cycle Assessment (LCA), which is discussed in section 2.2.2. in this chapter.

2.1.3.4 Principle 4: Complexity of infrastructure system

Infrastructure systems often exhibit complexity. Such complexity can be divided into functional complexity and relational complexity (Fratini *et al.*, 2012). Functional complexity refers to complexity in the physical and technical objects in the infrastructure system and relational complexity arises from different views and perspectives between actors in the decision-making process. To deal with both types of complexity often requires a range of skills and a transdisciplinary approach. Transdisciplinary approaches involve collaborations and mutual learning processes between different stakeholders and experts (Wiek and Walter, 2009), accentuating the importance of stakeholder engagement. This is reflected by the operational principle (4a): integrate needs, roles and disciplines to tackle a complex infrastructure project. The other operational principle (4b) is to consider uncertainty in the project. The three sources of uncertainty in infrastructure planning are described in Table 4. All these sources of uncertainty are relevant to the context of wastewater asset planning. For example, uncertainty can arise due to of impacts of climate change and extreme weather on the wastewater treatment capacity (Unpredictability). There is also uncertainty in asset decision making when there is a lack of performance data of new wastewater treatment technologies which have not been trialled (Incomplete information or knowledge). Additionally, different stakeholders of the project may uphold their own opinions and priorities in the decision-making process (Different knowledge frames). Asset operators may have a strong priority in operational safety and maintenance whilst the finance managers are inclined towards investment cost and efficiency.

Table 4. Descriptions and mitigation strategies of different sources of uncertainty

Sources of uncertainty	**Unpredictability/variability**	**Incomplete information or knowledge**	**Different knowledge and problem framing**
Descriptions	This denotes the randomness or inherent variability in systems that are constantly changing, evolving and adapting to new conditions. This is also known as ontological uncertainty (Walker *et al.*, 2003; Brugnach *et al.*, 2008)	This can refer to the lack of information or imperfect knowledge. The situation of inadequate information can be due to "*inexactness, unreliability and border with ignorance*" (Funtowicz and Ravetz, 1990). Uncertainty due to imperfect or inadequate knowledge is known as epistemic uncertainty(Walker *et al.*, 2003; Brugnach *et al.*, 2008; Salet *et al.*, 2013)	this refers to differences in subjective understanding and structuring of problems between decision-makers (Brugnach *et al.*, 2008; Scholten *et al.*, 2015)

2.2 Sustainability assessment

2.2.1 What is sustainability assessment?

Sustainability assessment is an appraisal methodology that has gain practical prominence in supporting decision-making such as policy development and product/service appraisals in businesses (Sala *et al.*, 2015). The overarching objective of sustainability assessment is to identify plans and actions that can contribute towards sustainable development (Verheem, 2002; Bond and Morrison-Saunders, 2011; Devuyst, 2016). Waas *et al.*, (2014) defined sustainability assessment based on three fundamental purposes as "*Interpretation, Information-structuring and Influence*" when comes to decision-making. Firstly, sustainability assessment aims to contribute to a better understanding of the problem in a specific context (i.e. interpretation), generating information for the decision-making (i.e. information-structuring) and create sustainability objectives and imperatives (i.e. influence). These elements constitute the operationalisation of the assessment and also serve as a learning process for decision-makers by creating new insights and actions (Waas *et al.*, 2014; Sala *et al.*, 2015).

Sustainability assessment is recognised as a broad approach under the family of integrated assessment and impact assessment which is underpinned by the Three-Pillar model of sustainability (Pope *et al.*, 2004; Devuyst, 2016). Gibson (2012) differentiated sustainability assessment from traditional impact assessment because sustainability assessment is designed to address several sustainability imperatives. He argued sustainability assessment should accelerate a paradigm shift from merely "*minimising negative impacts*" to "*encourage positive steps towards greater community and ecological sustainability*". Additionally, unlike other well-established impact assessment frameworks such as Environmental Impact Assessment, sustainability assessment does not entirely rely on a prescriptive format and each assessment process should be tailored to fit the context (Bond, Morrison-Saunders and Howitt, 2012). It is important to establish the understanding of the decision problem and objectives as part of a robust sustainability assessment process (Hacking and Guthrie, 2008; Bond, Morrison-Saunders and Pope, 2012).

There are general procedures for conducting sustainability assessments. First, it is important to understand the underlying perspective and values when approaching sustainability. This requires a transparent presentation of the sustainability perspective (e.g. strong or weak sustainability?) and guiding principles (Sala *et al.*, 2015; Mostafa and Negm, 2018). Second, the practical process of sustainability assessment should be translated based on the context to which it is to be embedded. This entails the identification of the key actors, the assessment driver or its purpose, the scope for assessment, the complexity of the decision and timescale of impacts to be considered (Ashley *et al.*, 2003; Sala *et al.*, 2015; Mostafa and Negm, 2018). Third, a suitable methodology choice is made in accordance with the decision context. Specifically, the methodology choice here refers to the collection of methods (models or tools and indicators) that underlies the computation in the assessment (Sala *et al.*, 2015). A list of ideal qualities and characteristics of sustainability assessment is summarised as below (adapted from Ling *et al.*, 2021):

- Comprehensiveness: sustainability assessment should cover a holistic scope integrating the Three-Pillar model of sustainability (Hacking and Guthrie, 2008; Cinelli *et al.*, 2014; Waas *et al.*, 2014)
- Supporting decision: sustainability assessment should be incorporated in the decision-making process by offering new insights, information and perspective that lead to actions towards sustainable development (Sala *et al.*, 2015). It should be able to aid complex decisions and evaluating different alternatives with trade-offs for decision-makers. (Waas *et al.*, 2014)
- Stakeholder engagement: continuous engagement communication with stakeholders is recommended to understand decision openly and improve the decision quality (Barrett and Grizzle, 1999; Lai *et al.*, 2008; Bond, Morrison-Saunders and Pope, 2012; Cinelli *et al.*, 2014; Sala *et al.*, 2015; Wong-Parodi *et al.*, 2020)
- Pluralism: compared to the prescriptive process of other impact assessments, each sustainability assessment process should be designed and tailored to the specific context (Bond, Morrison-Saunders and Pope, 2012; Sala *et al.*, 2015). The methodological choice of sustainability assessment should also be linked to the social and institutional context where the sustainability problem originates.

- Transparency: the assessment process should be transparent in terms of the data source, methodological design and justification so it allows criticism and improvement (Ashley *et al.*, 2003; Sala *et al.*, 2015)
- Intergenerational equity: wider and long-term impacts should also be assessed to ensure the decision demonstrates corporate social responsibility and value for future generations (Bond, Morrison-Saunders and Pope, 2012; Ainger and Fenner, 2013; Cinelli *et al.*, 2014)

2.2.2 Approaches and tools

There is a plethora of approaches and tools to conduct sustainability assessment. This section comparatively reviews popular tools used in sustainability assessment and their applications in the context of wastewater management. The strengths and weaknesses of each tool are also discussed in terms of its suitability for this research. Table 5 summarises three families of sustainability assessment tools: biophysical, monetary, and indicators-based tools (Gasparatos *et al.*, 2008). Each family of tools are rooted in a specific valuation perspective and most of them are of a reductionist approach, which tends to quantify and aggregate impacts into a single unit of measurement. The following sub-sections briefly describe each tool and highlight their strengths and weaknesses.

Table 5. The three families of sustainability assessment tools (Gasparatos and Scolobig, 2012; Sala *et al.*, 2015)

Tool family	Tool examples	Valuation perspective	Stance on dimension reductionism
Biophysical tool	Life Cycle Assessment; Ecological footprint	Environmental	Reductionist approach
Monetary tool	Life Cycle Costing; Cost-Benefit Analysis;	Socio-economic	Reductionist approach
Indicators-based tools	Composite indicators; Multi-Criteria Decision Analysis	Depends on the methodology	Depends on the methodology

2.2.2.1 Life Cycle Assessment

Biophysical tools hold an eco-centric perspective and are designed to assess environmental impacts. Life Cycle Assessment (LCA) is a widely applied tool for assessing the environmental impact of a product or process over its whole life cycle. It is useful for making a holistic evaluation of impacts from "cradle-to-grave" and identifying environmental "hotspots" in the design and the operation of wastewater treatment works (Lundie *et al.*, 2004; Guest *et al.*, 2009; Corominas *et al.*, 2013; Rebello *et al.*, 2021). The standardised procedures to conduct a LCA include defining goal and scope, inventory analysis, impact assessment and interpretation (ISO, 2006). Corominas *et al.* (2020) provides a state-of-the-art review of the previous LCA applications on wastewater treatment and suggests that LCA has gained increasing interest and applications in the field. LCA can be flexible and diverse because it can either assess the impact of a wastewater treatment system in whole (from crude sewage to final effluent) or by stages (e.g. primary, secondary, tertiary or sludge treatment). For example, Garfí *et al.* (2017) used LCA to comparatively evaluate three biological treatment processes for small communities. Mills *et al.* (2014) assessed five different sludge treatment schemes to identify the optimal investment option using LCA combined with an economic assessment.

LCA has several strengths. Its flexibility allows it to be adapted to different assessment needs and scopes (Rønning and Brekke, 2013). Additionally, the results of a LCA can reveal valuable information about a product or process to support decision-making. The inclusion of the whole life cycle means that the comprehensive environmental impacts can be understood and avoids the transfer of problems (Rønning and Brekke, 2013; Zang *et al.*, 2015). The result also highlights 'hotpots' in the life cycle where the most environmental impact potentially occurs (Rønning and Brekke, 2013). However, LCA requires a significant amount of resources, time and data to perform (Balkema *et al.*, 2002; Niekamp *et al.*, 2015). As such, it can be difficult to meet the requirement on data quality (Teodosiu *et al.*, 2016) and data availability (Corominas *et al.*, 2013). One of the other concerns with data processing in LCA is that the results can be very sensitive to the definition of the functional unit and the data inventory used (Ainger and Fenner, 2013; Rønning and Brekke, 2013). This means that a sensitivity analysis should also be performed alongside the LCA results. Moreover, interpreting and communicating the results of LCA with decision-makers can also be challenging due to the complexity of information (Corominas *et al.*, 2013). Therefore, continuous stakeholder participation is important to achieve wide acceptance of LCA results (Guest *et al.*, 2009). Additionally, there is usually also a need for specialised software (e.g. GaBi®, and Simapro® etc) to perform LCA and it can be complicated to use for non-specialists.

2.2.2.2 Ecological footprint

Ecological footprint (EF) is another biophysical tool that measures the demand for natural capital of a population. Its measurement is defined by "…*the total area of productive land and water ecosystem required to produce the resources that the population consumes and assimilate the wastes that the population produced*." (Rees and Wackernagel, 1996). The analytical unit of EF is population (in a city, household, country, or region) and the measurement unit is the area of land and water (in global hectare). The analysis of the footprint includes two key accounts. The first account is the "*biocapacity*" (i.e. the ecological supply) for any population and the second is the footprint account (i.e. ecological demand) of that population (Gasparatos *et al.*, 2008). The comparison between these two accounts indicates whether there is a surplus or a deficit in the ecological supply. When the ecological footprint of a population exceeds its capacity, it implies the system is no longer sustainable. A good example can be by the work of the Global Footprint Network (Global Footprint Network, 2020), which compares the EF of most countries against the world biocapacity (i.e. one earth equivalent) alongside the Human Development Indices.

EF was not suitable for this research because it is mostly focused on the regional or national scope. Theoretically, the demand for natural capital of a process or service (such as a wastewater treatment plant) can also be expressed in EF, but it is not intuitive for providing decision support in an organisation. There has been no application of EF to the field of wastewater treatment so far. A popular alternative is the carbon footprint. Carbon footprint measures specifically the amount of greenhouse gas emission (in carbon dioxide equivalent) of an individual, organisation, product, or project (DEFRA, 2009; GHG PROTOCOL, 2015). However, carbon footprint tools intend to only measure GHGs emission in terms of carbon dioxide equivalent and does not evaluate other environmental impact categories.

2.2.2.3 Life Cycle Costing

Life Cycle Costing (LCC), or Whole-Life Costing (WLC), is the accounting of total costs over the whole life cycles of a project or an asset before making choices between alternatives (Woodward, 1997; Ainger and Fenner, 2013). This approach encourages decision-makers to develop the long-term outlook of the financial viability of a project rather than focusing on short-term savings. LCC follows the same principle of LCA, adopting a whole life cycle (i.e.

cradle-to-grave) perspective. LCC requires the following information to be provided for estimating the whole life cost: "*capital cost, asset life, the discount rate, operation and maintenance costs, disposal cost, uncertainty and sensitivity analysis*" (Woodward, 1997). And the data should be collected based on a clearly defined study scope. In the field of wastewater treatment, LCC has been widely used for comparative studies. For example, Koul and John (2015) applied LCC to evaluate the suitability of different wastewater treatment technologies alongside their treatment performances. The study suggests LCC can be a useful approach when used in conjunction with other metrics to determine the most cost-effective solution in a financially constrained condition. Tarpani and Azapagic (2018) used LCC to compare four advanced wastewater treatment technologies and five sludge treatment options. The results presented the contribution of different life cycle stages to the whole life cost of each treatment technology and provided strategic recommendations for its financial viability.

The key benefit of using LCC is to provide a comparative and predictive outlook over the long-term financial viability of assets and projects. The inclusion of whole life cycles prevents a narrow focus on the short-term cost-saving and offers strategic insights into cost optimisation. One of the challenges when using LCC is that the quality of results depends upon the "*supply of accurate, speedy and relevant information*" (Woodward, 1997). Similar to LCA, LCC requires a large amount of data and time to perform. The centralisation of various costing data necessitates an effective information-sharing process. Epistemologically, LCC only provides insights through the lens of a monetary perspective as it excludes environmental and social considerations.

2.2.2.4 Cost-Benefit Analysis

Cost-Benefit Analysis (CBA) is a widely applied appraisal tool to determine the overall feasibility of a project or several projects comparative at the early stage of planning (Gasparatos *et al.*, 2008; Djukic *et al.*, 2016). The feasibility is determined by two key concepts in CBA: the aggregated benefits and costs. When the benefit outweighs the cost (i.e. positive net profile), the project is considered economically feasible and often the greater the net profit the more desirable is the project (Chen and Wang, 2009; Molinos-Senante *et al.*, 2010). The quantification of benefits and costs is all converted into monetary units to eliminate the limitation of heterogeneity of measurement units. The project benefits include both monetary (such as profit of operating the service) as well as external environmental and social added value. In the context of wastewater treatment, the environmental benefits can be seen as the damage avoided by implementing a wastewater treatment technology. Molinos-Senante *et al.*, (2010) applied CBA to compare the net profit of 22 wastewater treatment plants. The environmental benefits in this study were estimated by calculating the "*shadow prices*" which represent the environmental damage if no action was taken. The shadow prices in this study were concerned with four main types of pollutants in wastewater: chemical oxygen demand (-131 €/g), suspended solids (5.1 €/g), nitrogen (-8.06 €/kg) and phosphorus (-30.9 €/kg). By adjusting to the proportionate treatment flow of wastewater treatment plants, the study concluded that the removal of phosphorus provided the highest environmental benefit on average. Another similar application by Djukic *et al.* (2016) used CBA to compare the 'benefits to costs' ratio to determine the economic desirability and feasibility of a wastewater treatment project.

The key strength of CBA is that it is easy to support decisions. The result of CBA in a single monetary value is intuitive to interpret and communicate to decision-makers (Pearce *et al.*, 2006; Lai *et al.*, 2008). The practice of CBA in the field of wastewater treatment has been popular and is relatively mature. However, Gasparatos *et al.* (2008) noted some limitations when using CBA in terms of the valuation and discounting of benefits and costs. First, the valuation of environmental and social benefits can be over-reliant on monetisation. Such monetisation tends to overlook "*emotions, beliefs, values*" (Sinden, 2004) and contributes to the "*commoditisation of everything*" (Bangser, 1982). The quality of results highly depends on the accuracy and availability of information. However, there is high uncertainty of converting environmental and social benefits and costs into monetary terms and it can be subjective when selecting reference to determine the shadow price or using contingent valuation methods (i.e. surveying people about their willingness to pay) (Bebbington *et al.*, 2007). Second, the selection of the discount rates for the economic value is highly assumption-based. The uncertainties from the subjective selection of valuation method and discount rates can reduce the objectivity of CBA results and such subjectivity is often not explicitly acknowledged (Bebbington *et al.*, 2007).

2.2.2.5 Indicator-based tools

Indicators are tools to provide manageable units of measurements and "*an operational representation of an attribute of a system.*" (Gallopín, 1996). The Composite Indicators/Indices (CI) are the aggregation of all indicators using a pre-determined methodology (Gasparatos *et al.*, 2008). The aggregation of all attributes greatly reduces the complexity of information and provide an overall outlook of the system. CI has attracted applications in a diverse range of fields because of its flexibility, ease to use and communicate in decision-making. A summary of applications of sustainability assessment using indicators in the field of wastewater treatment is given in Table 6. Most studies included indicators based on the three-pillar model of sustainability and applied a CI approach.

Table 6. Applications of indicators to assess sustainability of wastewater treatment systems/technologies

References	Environmental indicators	Social indicators	Economic indicators	Composite indices	Data source
Klapwijk and Eggels, 1997	✓	✓	✓		Egypt
Molinos-Senante *et al.*, 2014	✓	✓	✓	✓	
Mustapha *et al.*, 2018	✓			✓	
Sabia *et al.*, 2016	✓	✓		✓	Italy
Muga and Mihelcic, 2008	✓	✓	✓		US
Plakas *et al.*, 2016	✓	✓	✓	✓	Greek
Kalbar *et al.*, 2016	✓	✓		✓	India
Ren and Liang, 2017	✓	✓	✓	✓	China

A list of attractive features of CI has been summarised by Nardo *et al.* (2005). CI is attractive to use because it is easy to interpret and facilitates information processing (Zhou and Ang, 2009). It allows decision-makers to rank and identify the best alternatives using CI. However, the practice of aggregation and CI remains contentious due to the danger of oversimplification and subjectivity (Nardo *et al.*, 2005). Hence, sufficient justification of the selected aggregation method should be clearly documented. The selection of weights and indicators is also highly contestable because they can be influenced by personal and organisational biases. If weighting is applied during the aggregation of indicators, there are different weighting techniques available which can lead to different aggregated results.

An extension of the use of indicators and CI is the Multi-Criteria Decision Analysis (MCDA). MCDA is a collective term for methods that deal with multiple and often conflicting criteria and identify the most preferred option based on the preference systems of decision-makers (Belton and Stewart, 2002; Niekamp *et al.*, 2015; Ling *et al.*, 2021). Specifically, MCDA involves the construction of a criteria structure with indicators and the selection of a preference model to measure towards the overall objective. The use of MCDA is flexible as it can adapt to a great diversity of decision problems. MCDA has been widely applied in the field of environmental management (Kiker *et al.*, 2005; Huang *et al.*, 2011) and urban water management decisions (Lai *et al.*, 2008).

MCDA can provide several practical benefits for decision-makers. In a group-decision environment, MCDA has demonstrated potential to facilitate communication between stakeholders and analysts (Niekamp *et al.*, 2015). The process of problem and preference structuring in MCDA can highlight the similarity and conflict between decision-makers and stakeholders, promoting in-depth discussion between them (Saaty, 1980). MCDA also provides flexibility as there is a wide inventory of methods to choose from. However, great care is required to select a specific MCDA method. Different MCDA methods differ in their theoretical bases, and their preference and aggregation models (Belton and Stewart, 2002). Cinelli *et al.* (2014) argue that many MCDA analysts and practitioners tend to select a method directly by their affinity and familiarity with the method. In addition, Belton and Stewart, (2002) argued that many studies on MCDA are fragmented, focusing on favouring a single method rather than seeking an integrated and holistic perspective. Furthermore, MCDA has other potential limitations (Lai *et al.*, 2008). Firstly, preferential independence between criteria or attributes must be maintained. This refers to the condition that the preference for a criterion or attribute should not be dependent on the performance in other criterion or attribute. The lack of preferential independence may lead to invalid results. This condition is further explained in the theoretical foundation of MCDA in chapter 5. Secondly, there might be double counting when performing MCDA. Double counting occurs when redundant or repetitive criteria or attributes are included in results aggregation. This often leads to an inflated and overrated contribution to the result and therefore reducing its credibility. Thirdly, if a complex MCDA model is deployed, it may hinder the transparency and ease to communicate with decision-makers. Although a 'black-box' approach (embedding the method into a user-friendly system with hidden computation) can facilitate its operation, decision-makers may also prefer to review the process so the results can be understood and trusted (Hamouda *et al.*, 2009).

2.2.3 Proposing a suitable sustainability assessment methodology

Although a plethora of tools have been developed to perform sustainability assessment, there is a lack of standardised procedures and guidance on selecting an appropriate tool and conducting better quality sustainability assessment (Ashley *et al.*, 2008; Gasparatos, 2010; Gasparatos and Scolobig, 2012). To achieve a robust selection of tools to perform sustainability assessment, a list of considerations based on the literature reviewed is proposed for the research of this thesis:

- The theoretical basis of the tool: what concept of sustainability is relevant to the organisational context? (Gasparatos and Scolobig, 2012)
- Requirement of resources such as time, budget, data, specialised knowledge and software (Gasparatos and Scolobig, 2012; Niekamp *et al.*, 2015)
- Understanding of the needs and values of the stakeholders (Ashley *et al.*, 2008; Gasparatos, 2010; Gasparatos and Scolobig, 2012)

First, it is important to understand what sustainability means for the water company. The water company defines its sustainability policy as to "*strike a balance, do the right thing for people, for the performance of our business and for the natural environment*" (Thames Water Utilities Limited, 2018c). This policy places a strong emphasis on the integration of social, environmental value, and financial performance in one decision system, which closely resembles the Three-Pillar model of sustainability (such as in Figure 7a). The convergent space between those pillars represents a solution that can deliver sustainability and so, to operationalise the integration of this sustainability model, the selection of sustainability assessment tools needs to reflect and include all three dimensions of sustainability. The increasing complexity in the infrastructure project also calls for a paradigm shift from a single-perspective assessment to an interdisciplinary and integrated assessment approach (Ashley *et al.*, 2008; Gasparatos and Scolobig, 2012; Niekamp *et al.*, 2015).

Integration of the Three-Pillar model of sustainability can be achieved either by methodological integration or by indicators integration. Methodological integration refers to combining multiple single-perspective assessment tools (such as the tools in Table 5) into one assessment system. For example, some studies suggest LCA can be integrated with economic assessments such as LCC and Social LCA to acquire a broader perspective of sustainability (Corominas *et al.*, 2013; Molinos-Senante *et al.*, 2014; Padilla-Rivera *et al.*, 2019). However, the integration of LCA and LCC was deemed impractical for this research. Firstly, as mentioned earlier, a significant amount of cost, time and data are required to obtain reliable results in LCA and LCC (Balkema *et al.*, 2002; Niekamp *et al.*, 2015). Secondly, integration of results from LCA and LCC presents a practical challenge as the practice to integrate both methods systematically is less mature (Niekamp *et al.*, 2015). Previously, Mills *et al.* (2014) conducted a LCA and an economic assessment to evaluate the desirability of five sludge technological configurations for the water company. However, the method of aggregating LCA and LCC results was arbitrary. Specifically, scenarios were ranked separately in LCA and LCC and then two rankings were simply added to yield an overall rank score for each scenario. Although this aggregation approach seems simple to operate, no assumption was made to validate the addition of ordinal scale, which is often contestable due to uneven distances between rank numbers. Therefore, there are often practical complications when combining two assessment tools due to the challenge of theoretical and procedural incompatibility as well as resource intensity (Xue *et al.*, 2015). The complexity of combining methods can reduce the ease and practicality of an integrated sustainability assessment (Hacking and Guthrie, 2008; Tajima and Fischer, 2013).

The alternative to methodological integration is indicators integration, which provides measurements towards all dimensions of the indicators (examples in Table 6). The integration of indicators can be achieved readily and flexibly through a CI approach, with the possible extension of MCDA. Compared to methodological integration, integration through indicators is less resource-intensive to develop and perform. Therefore, MCDA, with a composite index, was proposed for this research. De Montis *et al.* (2004) and Cinelli *et al.* (2014) have suggested the potential usefulness of MCDA for decision-making in the context of sustainability. MCDA provides a flexible and transparent way to translate sustainability into manageable measures while integrating the preference and priorities of stakeholders (Tanzil and Beloff, 2006;

Gasparatos and Scolobig, 2012). It is also flexible in the way that it can be used for assessment based on either a weak or strong sustainability perspective, depending on the level of trade-off allowed (Gasparatos *et al.*, 2008). In terms of the position on a reductionist perspective, the growing criticism over the conceptual validity of a weak sustainability tool (i.e. a reductionist approach) is acknowledged, but that should not be at the expense of recognising its strength in practical decision-making. On this basis, it was decided to include CI as the main output of the assessment but also to retain and display the results of individual criteria and indicators before fully aggregated. Given that a range of MCDA models is available, a more in-depth literature review was conducted to support the selection of a detailed MCDA model and methods. This detailed MCDA literature review is presented in Chapter 5 after the initial case study to understand the decision context of the water company (Chapter 4). The development of the list of criteria and indicators for the sustainability assessment tool is then discussed in Chapter 6 and 7.

2.3 Linking sustainability assessment to decision making

2.3.1 Decision making and decision support

Based on the research objective, the sustainability assessment tool to be developed also needs to provide decision support in the organisation. This necessitates the understanding and discussion on how to build a decision support system that can accommodate the requirement of sustainability assessment and the needs of the organisational context. Decision science is fundamentally concerned with understanding and improving decision making. Broadly, a decision problem involves four key aspects: problem context, problem finding, problem-solving and a legitimation process (Kleindorfer *et al.*, 1993), as shown in (Figure 10). Kleindorfer *et al.*, (1993), p. 5 note three domains of decision science: descriptive, normative and prescriptive decision making. Descriptive decision science describes how people are making decisions. Normative decision science investigates how people should make decisions, and prescriptive decision science studies how to guide people to make normative decisions. Humans have limited memory, perceptual abilities, and information-processing abilities. Prescriptive intervention, such as decision aid, can alleviate task overloads on decision-makers' cognitive faculties and reduce decision complexity, especially when dealing with multiple conflicting decision objectives (von Winterfeldt and Edwards, 1986). A Decision Support System (DSS) or Decision Support Tool (DST) can be defined as "*…a* *management information system that also has some processing capacity designed to help the decision-maker use the information*." (von Winterfeldt and Edwards, 1986, p. 24). With a surge of computer-based technologies in the modern world, DSS can also be broadly defined as: "*interactive computer-based systems that help people use computer communications, data, documents, knowledge, and models to solve problems and make decisions*" (Power, 2002, p. 1)

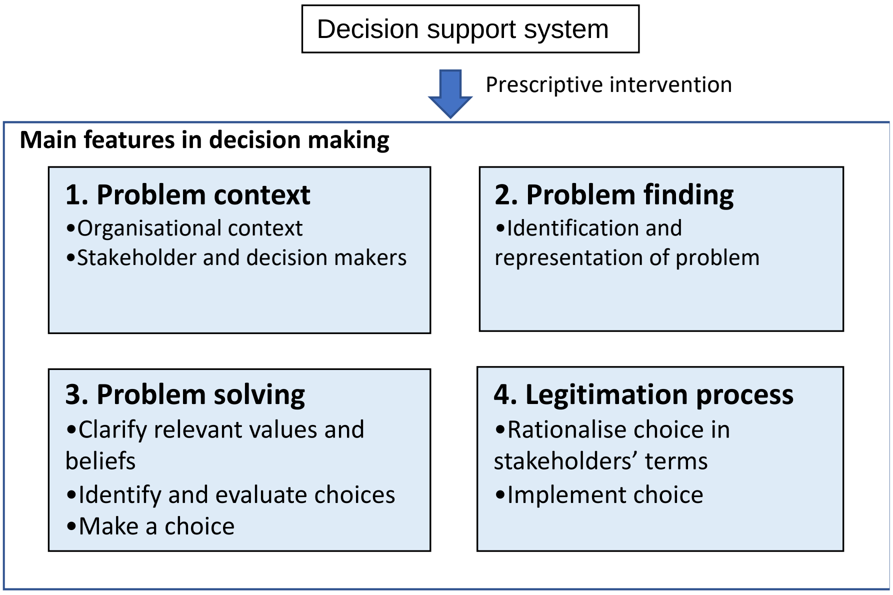


Figure 10. Key features in a typical decision-making process (Kleindorfer *et al.*, 1993)

The fundamental objective of a DST is to "*improve the performance of knowledge workers in organisations*" (Sprague, 1980). The knowledge workers refer to employees whose primary job involves data handling such as managers, analysts, and other professionals whereas organisations refer to the context where the challenges and opportunities present. The improvement of performance relates to the multiple events in problem solving such as problem finding and problem-solving (Figure 10). Based on this definition, it is important to understand who are involved in the decision-making process and potentially will use DST in the water company. This will be discussed along with the identification of user requirements in Chapter 4: Understanding the decision context.

The development of a DST consists of the following key components: the database, the analytical model or tool, and the user interface (Power, 2002). The database includes relevant data, knowledge and documents for the DST. The analytical model is the engine of the DST that manages and analyses information. The user interface is the communication device and platform that interacts with end-users. This is an important aspect of the development of a DST because it is closely linked to its usability (Hamouda *et al.*, 2009). Many DSTs appear to be too complex to use or untrustworthy because of a lack of transparency in operation. Therefore, it is critical to evaluate the usability of the DST as part of its validation and verification process.

2.3.2 Current trends of DST in the field of wastewater treatment

Given that a few DSTs have been developed in literature, most of them have two main roles in wastewater management and decision-making. First, DSTs often serve to assess the feasibility of wastewater treatment alternatives and select a suitable alternative as part of the design of STW. For example, Zeng *et al.* (2007) developed a decision optimisation model that incorporate MCDA and applied the model to compare 4 wastewater treatment alternatives based on 8 decision criteria. Loetscher and Keller (2002) developed a decision support system SANEX^{TM} to assess the feasibility of different sanitation alternatives. The system incorporates a Multi-Attribute Utility Theory that rates alternatives based on two major criteria (Implementability and Sustainability) aggregated from a list of 24 sub-criteria, with an additional cost. Castillo *et al.* (2016) developed an Environmental Decision Support System (EDSS) to evaluate and compare wastewater treatment alternatives for the pre-design of STWs. The system consists of a hierarchical rule-based approach for decision-makers to generate alternatives based on information including influent, effluent and decision priorities, followed by a multi-criteria analysis to rank those alternatives. Similarly, Sadr *et al.* (2018) built a DST that integrated two MCDA methods (AHP and TOPSIS) to compared and rank 10 wastewater treatment processes alternatives based on 10 selected decision criteria. The DST was also incorporated with a user-friendly interface and its importance in DST implementation was highlighted in the study. Second, DSTs have also been used to optimise the performance and operation of STWs. Pasqualino *et al.* (2009) utilised LCA as a DST to improve the environmental performance of an existing STW. Specifically, LCA was used to evaluate the environmental impacts of the process of wastewater and sludge treatment and identify impact ‘hotspots’ with the potential alternative to reduce those impacts. The study identified the highest impacts were located in the aerobic reactor for the liquid stream and the anaerobic digestor for the sludge stream, respectively, followed by recommendations to reduce the energy consumption and increase the energy efficiency of the operation. Kim *et* *al.* (2016) developed a decision support system to aid the operational decisions of STWs. The system integrates diagnosis, prediction and scenarios based modules to account for the variations in the influent and effluent qualities and suggest appropriate decisions at the operational level. Overall, those studies demonstrated the use of DST can aid complex decision-making and provide better information for decision-makers (Lai *et al.*, 2008).

Several trends of DSTs in wastewater management were highlighted (Hamouda *et al.*, 2009). First, although technical considerations (such as performance targets for pollutants removal) of the wastewater treatment system were the common priorities in decisions, recent studies (such as the examples abovementioned) also incorporate non-technical aspects into their decision support systems. Second, most DSTs use heuristic knowledge to drive decision making, especially for the non-quantitative aspects in the wastewater treatment design; Third, there is a need to develop integrated DSTs that are usable. Many studies highlighted the need to develop an integrated approach (Niekamp *et al.*, 2015) driven by the increasing complexity in wastewater decision-making (Castillo *et al.*, 2016; Ullah *et al.*, 2020). The 'integration' entails the inclusion of the Three Pillars of sustainability into the decision as well as stakeholder participation (Ashley *et al.*, 2008; Lai *et al.*, 2008; Ullah *et al.*, 2020). Lastly, the usability aspect of DSTs is also important but often overlooked in its design (Hamouda *et al.*, 2009). The DSTs with a sophisticated and complex design may fail to incorporate the usability aspect and lead to low adoption rates by the users. This combination of drivers has accelerated the research of using MCDA for sustainability assessments in water and wastewater management (Lai *et al.*, 2008). The discussion above supports the choice of MCDA as a suitable sustainability assessment methodology and validates the need for a decision aid approach that address those trends above.

Chapter 3 Research design

This chapter first presents an overview of different elements in a research design relevant to the present study. Section 3.4 then presents the epistemological position adopted for this research and summarises the broad research design applied. In this thesis, a research design includes the selection of a research approach as well as methods underpinned by a research paradigm (Figure 11). Research approaches are systems of plans, procedures and specific research methods with underlying assumptions (Creswell, 2009). Common approaches include those that are qualitative, quantitative and mixed methods. Research methods are specific methods or techniques to conduct data collection, analysis and interpretation.

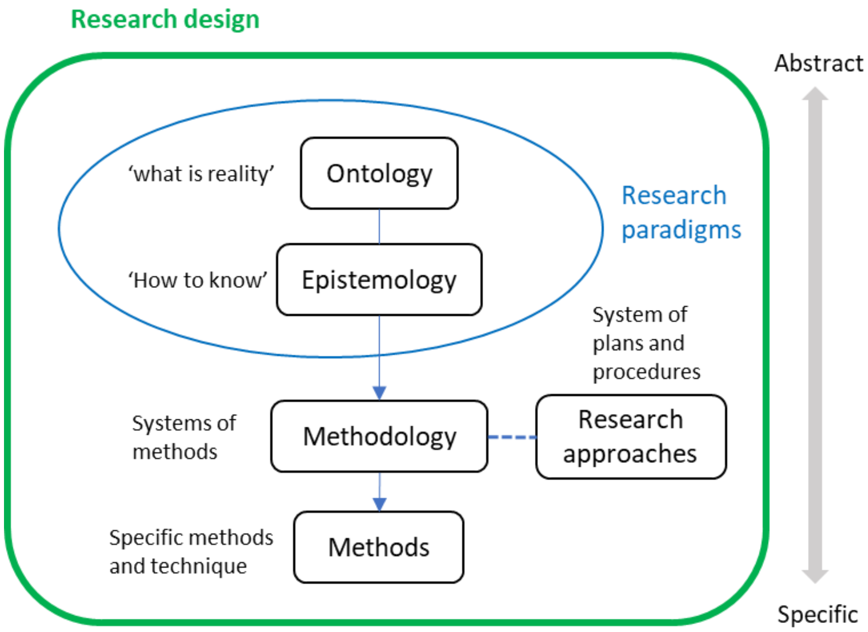


Figure 11. The definition of a research design in this thesis involves a research paradigm, methodology/approaches and methods.

3.1 Research ontology and epistemology

The methodological design is often determined by the ontological and epistemological assumptions made and underpinned by the overall research question. Ontology refers to beliefs about our reality and existence (Currivan and Gilbert, 2008; Scotland, 2012). There are two contrasting schools of ontology: realism and relativism (Levers, 2013). Realist implies that only one truth or reality exists and it does not change. Similar to realism, objectivism states that reality is objective and is external to the social actors. In contrast, relativist suggests that multiple realities exist and the meaning depends on the context. In this domain, realities are constructed from human minds, consciousness and experience (Levers, 2013). Therefore there can be multiple realities and they consist of finite subjective experience (Denzin and Lincoln, 2000).

If ontology is the study of being and the nature of reality, epistemology looks into how new knowledge about reality is acquired (Levers, 2013). In a deeper sense, epistemology is more about the relationship between the 'knower' and the 'knowledge'. A positivist epistemology implies that reality is objective and independent of human minds and positivists believe knowledge is obtained by taking measurements and observations (Mackenzie and Knipe, 2006). Such an epistemological stance sees knowledge objectively and excludes any personal value from the researcher (Bryman, 2016). Research with a positivist epistemology often adopts a quantitative approach to establish causality and predictability through hypothesis testing (Currivan and Gilbert, 2008, p. 138). This type of logic is also known as deductive reasoning, which is commonly applied to subjects in natural science (Bryman, 2016). On the other hand, interpretivists disbelieve that reality is objective because knowledge is relative to a particular context and circumstance (Levers, 2013). Interpretivists use qualitative approaches to construct understanding and stories to explain social behaviour and phenomena of interest (Currivan and Gilbert, 2008), which prevail in social science research. A summary of the classification of main ontologies and epistemologies is shown in Table 7.

Table 7 Summary of major research paradigms. Adapted from (Mackenzie and Knipe, 2006)

Paradigms	Ontology	Epistemology	Research approaches	Example of methods
Positivist paradigm	Realist/ Objectivist	Positivist	Quantitative; deductive	Experiment
Interpretivist paradigm	Relativist/ Subjectivist	Interpretivist	Qualitative; inductive	Interviews; Observations
Pragmatic paradigm	-	Pragmatist	Mixed methods; abductive	Mix of both quantitative and qualitative methods

3.2 Research paradigms

A research paradigm consists of a system of beliefs and assumptions made by researchers and is a reflection of the underlying "*epistemological understanding*" of the world (Feilzer, 2010). Morgan (2007) discussed different interpretations of a research paradigm. Philosopher Thomas Kuhn (1970) used the term 'paradigm' to describe a set of beliefs shared among scientists on how a problem should be studied and understood. Paradigm can also be defined as a 'worldview' that describes a "*basic set of beliefs that guide actions*" (Guba, 1990, p.17).

Worldviews are also seen as a "*general philosophical orientation*" about the world, reality and the nature of research (Creswell, 2009, p.6). Alternatively, paradigms can also be regarded as a system of specific ontological and epistemological stances. This definition puts a stronger emphasis on the nature of knowledge and the acquisition of it and it is the most widely applied definition in social science research (Morgan, 2007). As such, the selection of a paradigm is fundamental to the subsequent choices of methodology, methods and research design (Mackenzie and Knipe, 2006).

3.3 Mixed methods research

As mentioned above, there are two research paradigms reflecting the dichotomy of epistemological stances: the positivist paradigm and the interpretivist paradigm. The positivist paradigm arises from a positivist epistemology that focuses on the objective reality and it can only be learned through objective measurements. This often leads to a research design that is quantitatively orientated (Feilzer, 2010). In contrast, the interpretivist paradigm (or constructivist paradigm) operates on the notion that multiple realities exist and they are constructed by subjective and contextual understanding. Such a paradigm often leads to a qualitatively orientated methodology (Feilzer, 2010). However, there have been debates between purists (who advocate for a single paradigm) over which paradigm is ideal. This ongoing conflict was known as the 'Paradigm Wars' (Johnson and Onwuegbuzie, 2004). Purists believe that paradigms are incompatible implying that quantitative and qualitative methods should not be mixed because of their divergent epistemologies. Although Kuhn (1970) argued that paradigms are incommensurable, Bryman (2016) suggested that this should not imply incommensurability between quantitative and qualitative research as they share areas of commonality. The dichotomy has been challenged with the advent of a new paradigm known as 'mixed methods research'. The mixed methods here usually refer to the mix of quantitative and qualitative methods (Creswell, 2009). This paradigm sees qualitative and quantitative research as a continuum rather than two competing choices. Johnson and Onwuegbuzie (2004) argued that epistemological and methodological pluralism can ultimately lead to more effective research given that modern research problems are increasingly more interdisciplinary and complex. Bryman (2016) mentioned a few common reasons to use mixed methods such as to provide a greater validity of findings (known as 'triangulation'), completeness of the understanding and offset the limitation of one method.

It is also worth mentioning that mixed methods research is not defined by the system of research epistemologies but rather focuses on the methodology itself.

Mixed methods research is often associated with or indeed represented as, a ‘pragmatist paradigm’ (Tashakkori and Teddlie, 2003). The essence of pragmatism focuses on “*what works*” and making empirical inquiries to solve practical problems in the real world (Creswell, J., & Plano Clark, 2009; Feilzer, 2010). This does not require the researcher to commit to one ontological and epistemological stance (Creswell, 2009). Pragmatists often involve ‘abductive reasoning’ to address the relationship between theory and data (Morgan, 2007). Abductive reasoning refers to a logical connection that allows researchers to move “*back and forth between induction and deduction*” (Morgan, 2007; Feilzer, 2010). Researchers are free to use different research approaches to acquire knowledge to address the research purpose. A central question needs to be asked by the researcher: ‘what are the best ways to derive knowledge to solve the research’? Therefore, mixed methods with a pragmatic approach are useful to understand and study complex social phenomena (Mason, 2006; Currivan and Gilbert, 2008; Cronin *et al.*, 2012).

3.4 Overall research design for this research

3.4.1 Pragmatic paradigm

This section discusses what research design is suitable for the aim and context of this research. The overarching research aim is to develop a sustainability assessment tool that informs wastewater asset decisions in the water company. A pragmatic paradigm was chosen for this research because:

1) The research aim is fundamentally about developing a solution, which is strongly aligned with the practical emphasis of pragmatists (i.e. what works?).
2) Research in sustainability assessment is inherently inter-disciplinary and epistemologically pluralistic (Vildåsen et al*.*, 2017). A single school of epistemology may be insufficient to address all topics in this research. For example, a positivist perspective applies to use quantitative indicators to measure environmental impacts of a STW. However, an interpretivist perspective is more suitable to develop an understanding of how decisions are made in the company as this is a socially constructed process.

Following the adoption of a pragmatic paradigm, a multiphase mixed methods design (Creswell, 2009, p. 228) was considered for the methodological foundation of this research. In order to develop a sustainability assessment tool to inform decisions, the whole research process was broadly divided into three sequential phases: the exploratory phase, the development phase and the implementation phase (Figure 12). The development process is an adaptation from the general rapid prototyping framework for developing a DST (Power, 2002, p. 63), including 5 steps:

1. Identifying user requirement
2. Develop and test the first prototype
3. Revise and iterate the process
4. Repeat step 3 if needed
5. Pilot testing or full-scale implementation

As action research, the development process in this research involved iterative cycles of action, evaluation and reflection in order to progress to the solution (i.e. the decision support tool). The researcher conducted the research, not for the sole purpose of "*discovering new factors or revising existing theories*", but to acquire information for the "*practical application to the solution*" (Stringer, 2010).

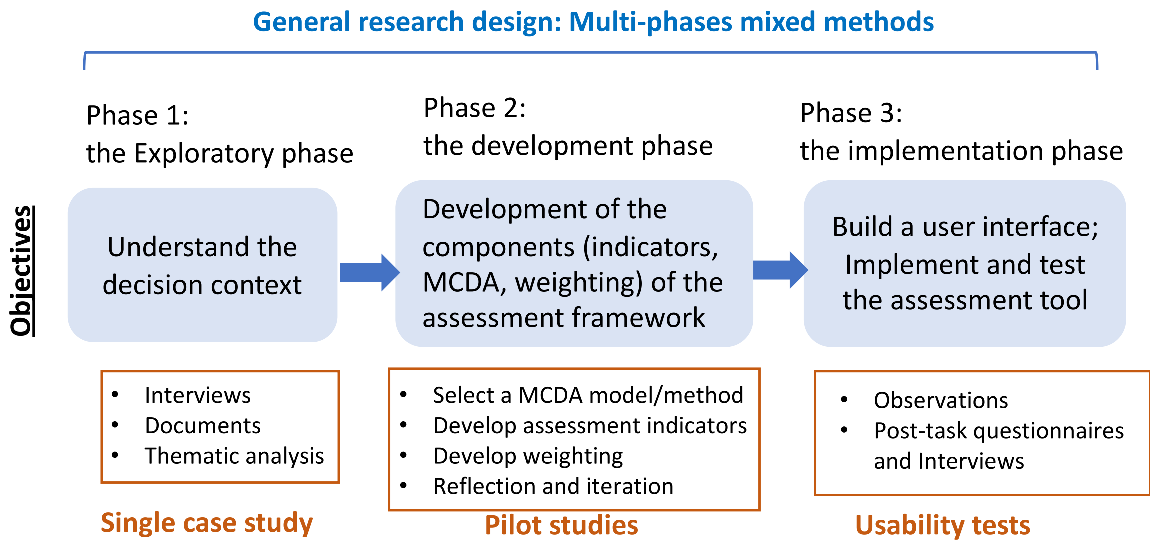


Figure 12. The illustration of the multi-phases mixed methods design for this research.

3.4.2 The exploratory phase

The explorative phase aimed to explore the decision and organisational context and establish a preliminary understanding of the wastewater asset decision making in the water company. The understanding of the decision-making process in an organisation can be implicit and hidden. Making this knowledge explicit using qualitative research helped inform design decisions when developing the assessment tool. The exploration of information in this study was more focused on the decision-making for wastewater technological selection, as the scope and type of solution was pre-determined by the water company. The research approach for this phase can be characterised as:

- *Inductive*: This research aimed develop knowledge from the 'ground-up' because the researcher had little information about the decision-making for wastewater technological selection at the beginning of the research.
- *Contextual*: the query is relevant to a specific organisational context (i.e. decision-making at the water company).
- *Interpretivist*: this predominantly involves qualitative research methods to collect, analyse and interpret the data from the perspective of decision-makers.

3.4.2.1 Case study approach

A case study approach was selected for understanding the decision context of the water company for the reasons above. A case study provides in-depth analysis for a specific case (an individual, a programme, a group, a community etc.) to understand a social phenomenon in a context (Yin, 1994). A case study approach is particularly useful for answering the 'how' or 'why' questions research and also the 'what' question in exploratory studies (Yin, 2014). The case study design consists of five components according to Yin, (2014):

1. The questions of study: the underlying question should reflect the relevance, interest and importance of the topic and help define the appropriate unit of analysis (i.e. the case) and suitable methods of inquiry. The fundamental research question for this phase was "*how to select a wastewater treatment process and what are the challenges in making that asset decision*?".
2. the propositions of study: propositions give meaning and purpose to the study questions and drive theory development. Based on the choice of an inductive

approach, no theoretical proposition was initially placed regarding the decision-making process in the water company.

3. The definition and unit of the 'case': a case can be an individual, a small group, a programme, a location, a specific event etc. (Burton, 2012). The definition of the case in this study was 'the current decision-making process' (to select wastewater treatment processes and technologies) in the water company. A group of people such as asset decision-makers could have been the unit for the case study. However, this was rejected because the focus of this case study was on the decision-making process itself, rather than on the behaviour of the social actors in the process. An additional consideration was that decision makers in the company are likely to change over time.
4. Linking data to the study question and proposition: this entails that the method of data collection and analysis should be selected in accordance with the central question of the inquiry.
5. Criteria for results interpretation: this refers to the construct validity of the case study. Construct validity refers to the "*extent to which a study investigates what it sets out to investigate*" (Gibbert *et al.*, 2008) and "*whether a procedure leads to an accurate observation of the reality*" (Crozier *et al.*, 1994). This study used multiple sources of evidence to enhance the construct validity, which is referred to as data triangulation (Yin, 2014; Mishra and Rasundram, 2017)

3.4.2.2 Data collections and analysis

Specifically, semi-structured interviews were conducted to collect information from company stakeholders. Semi-structured interviews were preferred to other types of interviews because they allow flexibility in the questioning while retaining the overall themes of questions. The use of interviews was also considered suitable as an opportunity to build rapport with company stakeholders. Thematic analysis was used to analyse the transcriptions of interviews by creating codes and identifying overarching themes within the data. Thematic analysis is flexible to conduct as it can adapt to different needs of studies (King, 2004; Braun and Clarke, 2006) and research questions (Nowell *et al.*, 2017). It is suitable for any study that "*seeks to discover using interpretations*" (Alhojailan and Ibrahim, 2012). It offers a structured approach to provide a rich and useful account of the data. (King, 2004). Additionally, it is accessible to those who are new in qualitative research or had little experience with other qualitative methods (Braun and Clarke, 2006; Nowell *et al.*, 2017).

This case study also applied data triangulation. Data triangulation refers to the convergence of multiple sources of evidence in a case study (Yin, 2014). Documents (including both company reports and notes from personal communications) were used as additional sources of evidence to corroborate the findings of the interviews. Analysis of documents was done to provide interpretations on the same research question and discussed in conjunction with the results of the analysis of interviews. Document analysis was considered suitable because it is efficient, cost-effective, and "*unobtrusive*", which means they are less affected by the influence or biases from the researchers (Bowen, 2009). Overall, the findings of this case study established a basic understanding of the wastewater decision-making process in the water company and highlight the potential challenges in that decision. The detailed procedures of data collection and analysis of this case study are presented in Chapter 4.

3.4.2.3 Other potential approaches for the exploratory phase

Another alternative qualitative research approach for this study is ethnography which intends to collect rich data from observations and interviews to understand how a group of people live or work in a culture. However, it was considered unsuitable for the nature of the question in this study. First, ethnography is generally applied for studying the culture of a group as it derives from an anthropological interest. In practice, ethnographic studies require a long time for the researcher to participate in the 'field' to make participatory observations. However, the decision-making process often operates at a higher management level which makes it difficult for the researcher (i.e. a non-employee) to participate. Additionally, some types of information (e.g. videos, photos, private documents) may not be accessible to the researcher due to business confidentiality and data privacy in the company. Narrative research was also not unsuitable for this study because it focuses on the story-telling of the lives and personal experiences of individuals (Currivan and Gilbert, 2008) and therefore it deviates from the present study's focus on the decision-making process in the organisation.

3.4.3 The development phase

The primary objective of this phase was to develop the key components of the assessment tool. The tool was based on the MCDA approach with CI as suggested by the literature review in Chapter 2. There were three components in a MCDA based framework: the selection of indicators, the development of weighting and the preference model (Figure 13). These components were eventually integrated into a user interface as a tool. It is worth noting that there is a clear distinction between a MCDA 'framework' and 'tool'. Frameworks refer to the structural integration of concepts and procedures whereas tools are the specific software and applications that incorporate the framework and its methodology (Sala *et al.*, 2015). Although the development of the assessment tool was underpinned by a structural framework, the emphasis of the research deliverable was placed on a practical tool that can be adopted by the water company to support decision-making.

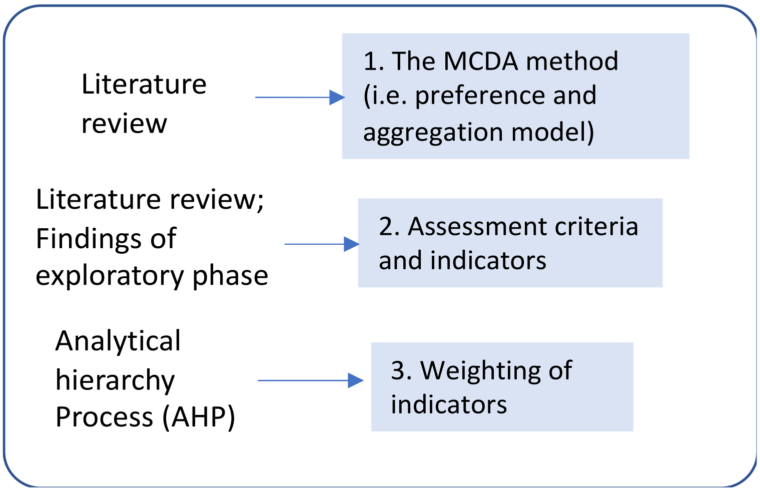


Figure 13. The components of the assessment framework and methods in the development process.

A range of methods was used to identify and select the most appropriate designs for individual components of the MCDA. First, additional literature on different MCDA methods was reviewed in Chapter 5. A comparative review of the theory and strengths and weaknesses of individual MCDA models and methods was undertaken to provisionally propose a suitable MCDA method for the assessment framework. Literature on relevant assessment indicators for evaluating wastewater treatment systems was also reviewed. This information was assessed and discussed with the findings from the exploratory phrase before developing a list of assessment criteria and indicators. Lastly, the weightings for these indicators were developed by the Analytical Hierarchy Process (AHP) with the stakeholders in the water company.

3.4.3.1. Pilot studies

Pilot studies were used in the development phase to pre-test the proposed assessment framework before implementation. Pilot studies are often conducted to test and validate the feasibility of the instrument, methods or protocols to be used in the main study (In, 2017). Pilot studies are also known as 'feasibility tests' or 'vanguard trials' in the context of clinical experiments (Thabane et al., 2010). Conducting pilot studies can increase the quality and also the likelihood of success of the main study (Thabane et al., 2010; Malmqvist et al., 2019). Pilot studies can be conducted for either quantitative (e.g. clinical treatments) or qualitative studies (e.g. social interventions). For qualitative studies, pilot studies can be seen as the 'pre-testing' or 'trying-out' of a research instrument (Thabane et al., 2010).

Pilot studies were conducted in the development phase to examine the reliability and feasibility of proposed methods and components in the MCDA framework. The end goal of the development phase was to identify a suitable design of the assessment tool before assembling the tool with a user interface. Two pilot studies were conducted in the timeframe of this research project :

- **Pilot study 1** (Section 6.3): After the components in the assessment tool was developed, it was applied to a 'real' business case. The case involved the retrospective comparisons of potential wastewater treatment investment options at a STW. Performance ratings of assessment indicators were collected as secondary data from a previous internal company report. Pairwise comparisons between indicators were collected through online questionnaires with internal stakeholders to develop indicator weights using AHP. A Simple Additive Weighted (SAW) model was used to aggregate the performance ratings to calculate the composite score of each wastewater treatment option. The ranking of investment options was compared to the previous decision made by stakeholders to discuss the consistency of the results in addition to sensitivity analysis for AHP.
- **Pilot study 2** (section 7.2): Based on the insights gained from the first pilot study, revisions were made to some components of the assessment tool. This included an updated list of assessment indicators and an alternative weighting method. The revised

assessment tool was then applied to a second business case (comparison of different phosphorus removal options at another STW). This pilot study was used to comparatively analyse the assessment results from applying different MCDA models regarding their feasibility and consistency of the option rankings. The different normalisation techniques for MCDA were also compared. Insights from both pilot studies were used to confirm the a suitable MCDA model and weighting method for the assessment tool before its implementation.

3.4.4 The implementation phase

3.4.4.1 Building the user interface

The third phase of the research was to build a prototype of the assessment tool with a user interface. The implementation of this tool involved testing its usability with the end users in the water company. The interface was based on a user-centred design, which focuses on the 'users' rather than the 'system' itself. Rubin and Chisnell (2008, p. 6) note that, in a user-centred design, the purpose of developing the tool should always be attempting to improve aspects of the human performance (e.g. rationality and coherence of decision-making). The user-friendliness should be integrated into the design of the tool for its practical success (Hamouda *et al.*, 2009; Huysegoms and Cappuyns, 2017; Wong-Parodi *et al.*, 2020). Although advancing computer science and technologies have opened up more opportunities for technical implementation, making decision support systems communicate with users remains as a challenge. Therefore, a usable design of the tool interface should accommodate both 'how the system works' and 'how the system communicates with users'.

Microsoft Excel® was selected as the design environment for the tool interface. Excel® was considered the most suitable design platform for following reasons:

1) Most of the company employees are familiar with using it
2) It is easy to navigate and make changes for the developer
3) The file format is compatible to the company file depository and easily accessible to users

Heuristic principles was applied to guide the design of the interface. This is a type of evaluation of a user interface by visually inspecting the design of the interface and examining its strengths and flaws (Nielsen and Molich, 1990). Based on years of repeated research, Nielsen (1994) proposed a comprehensive list of guidelines for a good design of a user interface, which is known as “Usability Heuristic principles”. Usability heuristics were applied to the early design of this assessment tool as it is easy to use and costs little time (Nielsen and Molich, 1990; Simeral and Branaghan, 1997). However, heuristic evaluations do not provide any simulation of a real task environment with intended users (Simeral and Branaghan, 1997). As such, they may not be able to reveal major design flaws that can occur in real tasks. More importantly, heuristics evaluations do not provide information about the magnitude of a design flaw nor offer potential solutions (Nielsen and Molich, 1990). Therefore, this study applied heuristic evaluations as ‘rules of thumb’ in the early design of the tool interface before ‘the full-scale’ usability testing with users.

3.4.4.2 Usability testing

Usability tests were used as the main approach for testing the assessment tool. Usability tests are a type of user interface evaluation approach that involves the direct participation of intended users (Bastien, 2010). It is widely used in the validation and implementation process of a user-centred design process. The goals of usability tests or evaluations are to examine the interactive system (of a product or tool) in aspects such as usefulness (i.e. how well the system completes the task which it was designed for), efficiency (resources required to complete the task) and the satisfaction of users (International Organization For Standardization, 1998; Bastien, 2010). Rubin and Chisnell (2008, p. 5) characterise usability by the following attributes:

- **Usefulness** is the degree to which a tool or a product achieve its designed purpose and this directly drives users’ desire to use it in the first place.
- **Efficiency** refers to how quickly the users achieve the goal using the tool, which is often measured by the time to complete the task.
- **Effectiveness** refers to the degree to which the tool operates as expected by the users and its ease of operation. This is often measured quantitatively by the rates of success or errors that occurred in completing a task
- **Learnability** can refer to either how quickly and easily a user can become comfortable and competent with using the tool or the ability for the user to pick up the tool after a period of inactivity.

- **User satisfaction** includes the '*perceptions, feelings, and opinions*' of users of the tool. This is an indication of the success of the tool based on the user experience. User satisfaction can be measured both quantitatively and qualitatively. Whilst quantitative information indicates the extent to which the tool works or not, qualitative information can reveal 'why there is a problem' and 'how to fix the problem'. Therefore, high user satisfaction is an integral element of a sound user-centred design.

Table 8 The four types of usability tests and their intended uses (after Rubin and Chisnell, 2008).

Type	1. Exploratory	2. Assessment	3. Validation or verification	4. Comparison
When	Early stage of development	Early stage and mid-stage	Late stage of development and close to release of the product	In any development stage
Purpose	To establish preliminary information and assumption about users and their needs. Testing basic product functions	To evaluate the effectiveness of the basics of the design concept. The evaluation requires users to perform tasks. Quantitative measures of performance will be collected	To measure usability against benchmarks or confirm the design changes that seek to remedy problems from earlier tests. Full data on all performance criteria will be collected.	To compare the functionality and performance of very different designs.
Why it is useful?	Provide soundness and foundation for the main design decision	Provide a straightforward indication of how the 'product' performs in reality and expose major design flaws	Provide a usability benchmark; Identify critical risk before the product finalisation	Identify creative and alternative design

There are four types of usability tests depending on the purpose of the test and the design stage of the product (Table 8). In this research, the 'assessment' type was used to examine the prototype of the assessment tool involving end-users to complete tasks in a simulated environment. This type of test is especially useful for providing a direct indication of the performance of the tool and revealing design flaws. A potential second round of usability testing (validation and verification) may also be included if major flaws were identified from the first round of testing. The detailed design and procedures of the usability testing are presented in Chapter 8.

3.4.4.3 Data collection and analysis

Two types of data can be collected from usability tests: 1) performance data and 2) preference data (Rubin and Chisnell, 2008, p. 88). Performance data are measurements of participant behaviour when completing the task. Observations were made to record the performance of users such as time for completion and number of errors. Preference data provide information about participants' opinions and experiences from using the tool. This was collected through post-task questionnaires and debriefings asking the users about their perceptions and satisfaction after the completion of the task. The alternative way to collect qualitative data is the think-aloud approach which asks participants to verbalise their thoughts and actions while performing the task. The strength of the think-aloud approach is that it can uncover a great amount of information processing in participants and expose a problem promptly. Compared with post-tasks interviews, think-aloud captures 'live' data so participants do not need to recall their thought process after the task is completed (Baauw and Markopoulos, 2004). However, the main drawback is that it diverts cognitive resources when completing a task, which may compromise the participant's performance. Given that this assessment tool was completely new to users in the water company before the usability testing, post-task interviews were considered a more suitable approach for users to retain maximum cognitive capacities when testing the tool.

The data analysis of the usability test aimed to evaluate different aspects of usability. The performance data from usability testing was analysed to provide an indication of the ease of use. This was complemented by the analysis of preference data from the questionnaires and post-task debriefing to examine the usefulness and user satisfaction. Quantitative data (e.g. questionnaire scales) indicated 'what' and 'where' were the problems and the qualitative data (e.g. transcription of post-task debriefing) revealed 'why' problems occurred and offered insights into 'how' it may be resolved. The specific data collection and analysis methods for the usability tests are described in Chapter 8.

Chapter 4 Case study: exploring the wastewater decision context

This chapter describes the detailed designs, methods, data collection and analysis of the case study which was selected to understand the decision context and challenges for selecting wastewater treatment options in the water company. The central research question of the case study was "*how to select a wastewater treatment process and what are challenges in making that asset decision*?" There were two sources of evidence in this case study: semi-structured interviews with company stakeholders and documents published by the water company.

4.1 Methods and materials

4.1.1 Design of interviews

It is crucial to acquire an understanding of the current decision-making practice in the water company so it can inform the development of the sustainability assessment framework. Semi-structured interviews were conducted to collect qualitative information on the current decision drivers and challenges when selecting a wastewater treatment process. The main questions in the interviews were:

1) What are your role and responsibilities in the water company?
2) What are the current decision aspects and drivers to select wastewater treatment processes and technologies?
3) What are the challenges when making that decision?

The first question allows participants to introduce themselves and understand the professional background of the participants. The second question aims to elicit descriptions of how the current decisions are made to select a wastewater treatment process. This includes, but is not limited to, the aspects and drivers in that decision-making process. The final question builds on the information from the second question to identify the challenges and suggestions for improvement based on the experience and perception of the participant. Although the original scope of interview questions focused on the whole 'decision-making process', it was further refined and narrowed down to the 'decision aspects and drivers'. The focus on the whole decision making process was very broad and in the trial run with some interviewees, the information collected was very scattered, and sometimes, irrelevant because they might not know how to response to the question.

Although the order of questioning was intended to be the same, the style of questioning varied slightly across different participants to ensure questions were understandable. Semi-structured interviews were preferred to other types of research interviews because it allows a certain degree of flexibility while maintaining a fixed set of topics. This also allows the researcher to expand upon the responses of participants and explore novel topics of interest. Each interview was designed to last around 30 mins and was conducted in person at the premises of the water company.

4.1.2 Interview recruitment

A stakeholder analysis was conducted jointly with the company representative to identify the target sample groups for the interview recruitments. Strategic Planning and Investment, Delivery Office and Operations were identified as the three key company departments for recruiting participants (Figure 14). Particularly, the department of Strategic Planning and Investment has the highest interest and influence on this research given that its primary function is to manage existing asset and plan for potential investment needs. External stakeholders (e.g. Ofwat, Environmental Agency) were not recruited in the interviews because it was considered very difficult to recruit a representative sample from each external stakeholder as well as the concern over the potential conflict of interest. Thereafter, a list of potential participants in these three departments was initially provided by the company representative and this list has been expanded by the snowball sampling technique. This sampling technique involves a chain of referrals made by participants (Browne, 2005). The technique provides practical advantages for the researcher to conduct explorative and qualitative research especially when the researcher has little prior knowledge on the topic and limited access to the target population (Hendricks and Blanken, 1992). Invites for interviews were sent out by email and appointments were directly arranged through the email calendar. The recruitment processes stopped once the target sample size (10-15 participants) has been reached or the information received had almost saturated[3]. Before the start of recruitment, an ethics review[4] was submitted and approved by the Ethics Committee

[3] Information saturation is reached when there is little new information acquired in the data collection process.

[4] The reference of the ethics review is UEC 2018 081 FEPS. A favourable ethical opinion was granted on 19th September 2018.

of the University of Surrey, which a list of documents included a Contingency plan (**Appendix 1**), Participant information sheet (**Appendix 2**), Consent form (**Appendix 3**), Recruitment email (
Appendix 4), Interview schedule (**Appendix 5**), Risk Assessment form (**Appendix 6**).

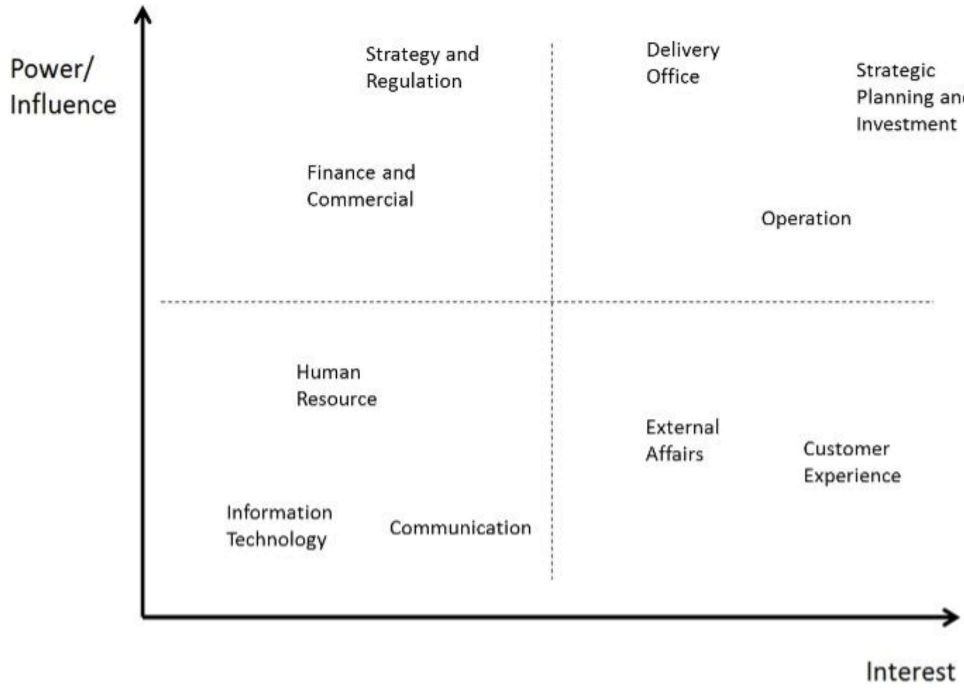


Figure 14. A stakeholder analysis was done to identify the target sample groups for interview recruitments. All departments in the water company were mapped based on the interest and influence on this research project.

4.1.3 Data analysis

Each interview was audio-recorded (upon consent received from the participant) and the recording was manually transcribed into text. The transcriptions were compiled and coded in NVivo 12® for thematic analysis. Thematic analysis is a data analysis method that "*systematically identifies, organises and offers insight into patterns of meaning across a data set*" (Braun *et al.*, 2019). The general procedures of thematic analysis were applied (Braun and Clarke, 2006; Nowell *et al.*, 2017):

1. Familiarising yourself with the data
2. Generating initial codes
3. Searching for themes
4. Reviewing potential themes

5. Defining and naming themes

The analytical process revolves between data reduction, data display and data conclusion (Miles and Huberman, 1994). Particularly, the process of developing and refining codes was to reduce the dimension and complexity of the information. Themes were developed to capture interesting patterns within codes and express the underlying meanings of the data (Erlingsson and Brysiewicz, 2017). As the data analysis progresses and iterates, codes can be updated and added to reflect new insights. There were three rounds of coding in the coding process (Figure 15). Preliminary codes were generated along with data collection to provide a detailed description of the information present in transcription at the early stage of data analysis. Specifically, an inductive coding approach was adopted to identify keywords and phrases within texts and use them for naming codes. This type of code that describes the literal content is known as semantic code (Braun *et al.*, 2019). The coding process was then iterated to produce secondary codes as the understanding of data evolved. The secondary codes were then further refined and renamed to final codes, which were condensed into themes for interpretation. Finally, themes were compiled and discussed with respect to the research question of this study. Thematic maps were created to visualise the development and results of each stage of coding.

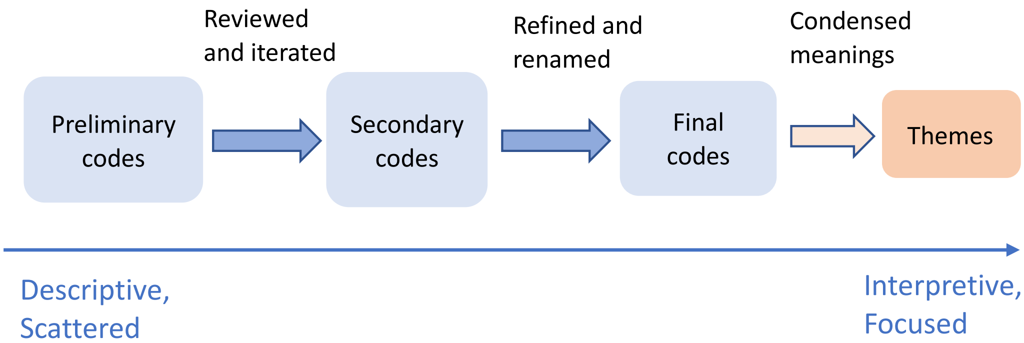


Figure 15. The coding process in the data analysis of interviews in this study.

Another round of data analysis was performed on a different source of evidence. After the collection and analysis of interviews, document including company reports (in the public domain) and notes of personal communications were coded and analysed. Document analysis is a "*systematic procedure for reviewing and evaluating documents*" which combines elements of content analysis and thematic analysis (Bowen, 2009). The criteria for the selection of documents were relevance, authenticity (Bowen, 2009; Bryman, 2016) and accessibility. There were eight reports published by the water company (mostly published in AMP 6[5]) selected for the document analysis:

- 2015-2040 Long Term Strategy[6]
- Annual Performance Report 2018-2019 (Thames Water Utilities Limited, 2019b)
- Corporate, Responsibility and Sustainability report 2016-2017 (Thames Water Utilities Limited, 2017a)
- Asset Management Policy 2016[6]
- Asset Management Policy 2020 (Thames Water Utilities Limited, 2020a)
- Reporting and Materiality Statement[6]
- 5-year Business Plan for 2015-2020 (Thames Water Utilities Limited, 2014)
- 5-year Business Plan for 2020-2025 (Thames Water Utilities Limited, 2019c)

The reasons for selecting those documents were: first, they were considered relevant to the topic of study which may contain information about the broad decision-making process and drivers in the company. Second, these documents were of authenticity as they have been reviewed and audited. Last, they were accessible in the public domain and exempt from potential restrictions of information sharing due to confidentiality. The limitation of using company reports in the public domain was that the information in these reports might mainly focus on corporate reporting rather than internal decision-making for wastewater asset investment. In light of this limitation, additional notes from informal conversations or consultations with managers in the department of Strategic Planning and Investment were also included in the document analysis to enhance the completeness of the dataset. There were four notes[7] of personal communications included in the document analysis:

- Meeting notes with a wastewater process modeller (March 2018)
- Meeting notes with a waste asset manager (March 2018)
- Meeting notes with a wastewater process engineer (April 2018)
- Meeting notes on the existing decision support tools for wastewater asset planning (April 2018)

[5] AMP 6 (Asset Management Period) covers the 5-year interval of 2015 to 2020.
[6] These documents are no longer available to access in the public domain.
[7] The original files of notes of personal communications will not be published and shared due to confidential and business sensitive information.

Company reports and notes of personal communications were compiled into one database for analysis in NVivo 12®. The codes derived from the thematic analysis of interviews were used as guidance for coding these documents. Identical or similar information was coded under the same name. However, new codes were also developed to capture novel information. As some company reports are lengthy, only the sections that are relevant to the research question and context (i.e. wastewater asset investment decision) were reviewed for coding. Keywords and phrases (e.g. "Strategy", "Decision", "Wastewater") were searched in NVivo12® to assist information navigation and coding process. Codes were then developed and grouped into categories. The findings from the document analysis were used to corroborate the results of interviews as part of the data triangulation of this case study.

4.2 Results

4.2.1 Thematic analysis of interviews

There were 14 participants accepted the interview invitations over the course of three months. As shown in Figure 16, most of the participants were from the department of Strategic Investment and Planning. Comparatively, the numbers of participants recruited from other departments were smaller due to a lower acceptance rate to recruitment invitations.

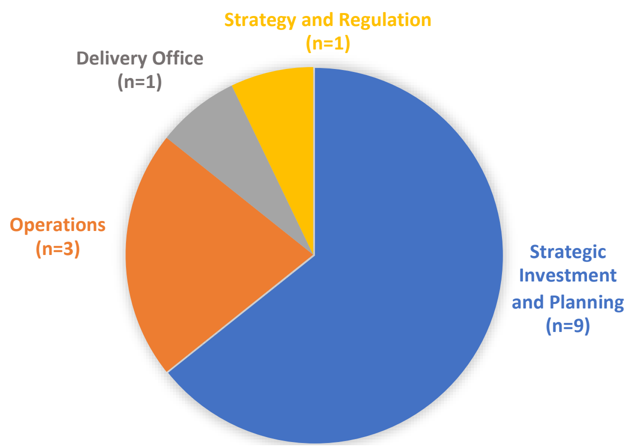


Figure 16. The number of participants by company departments

Audio recordings of interviews were transcribed into text files and compiled into a single database in NVivo12®. Transcriptions were reviewed individually and then collectively to gain familiarity with the data. During this process, preliminary codes were developed inductively by identifying common and interesting words or phrases. There were 52 preliminary codes and 6 codes clusters from the first round of coding (Figure 17). These codes were a descriptive and comprehensive reflection of the content of the interviews. The lines between codes in Figure 17 represent connections between information. However, it was realised that these codes provide limited meaning and interpretation of the research data because they were too many of them and some were irrelevant to the research question. Secondary codes were developed after reviewing the transcriptions and iterating the coding process. The second round of coding removed irrelevant codes and merged overlapping codes to reduce the level of "noisy" data and complexity. Figure 18 shows the thematic map that contains 37 secondary codes and 6 code clusters (Ling *et al.*, 2021). The secondary codes were then further refined and renamed to interpret data with respect to the research question and context. The final result of thematic analysis included 21 codes and 6 themes (Figure 19). The summary of descriptions of each code can be found in Table 9.

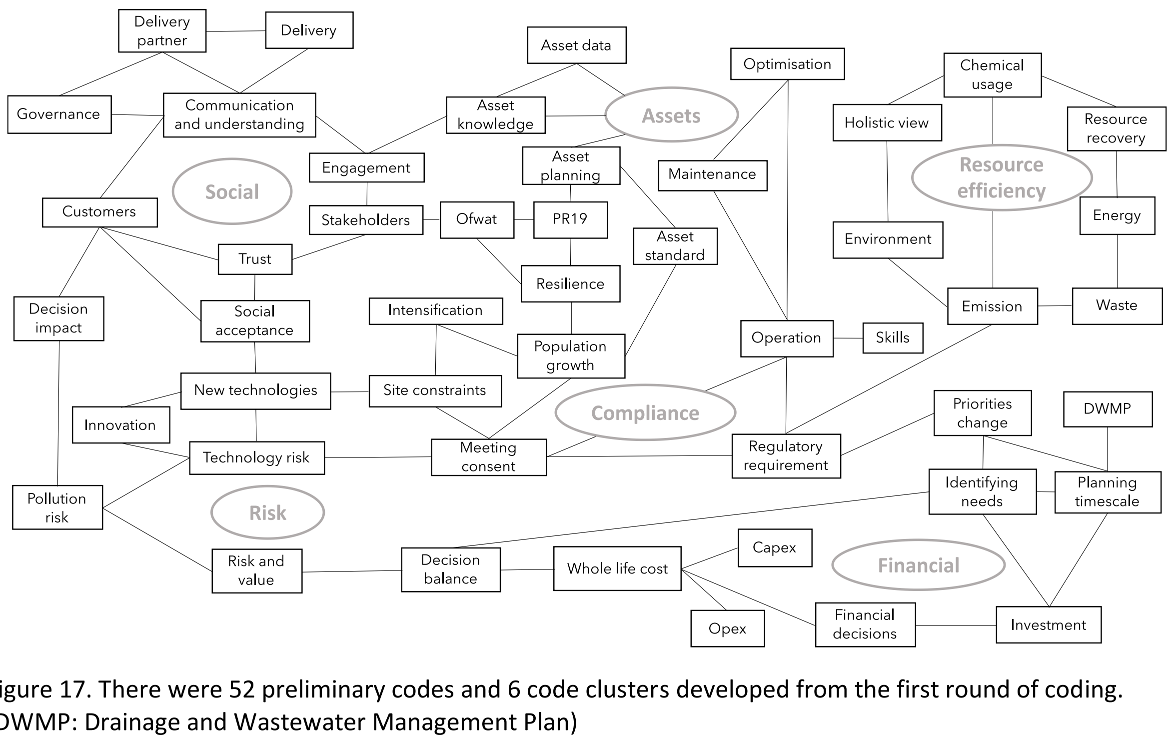


Figure 17. There were 52 preliminary codes and 6 code clusters developed from the first round of coding. (DWMP: Drainage and Wastewater Management Plan)

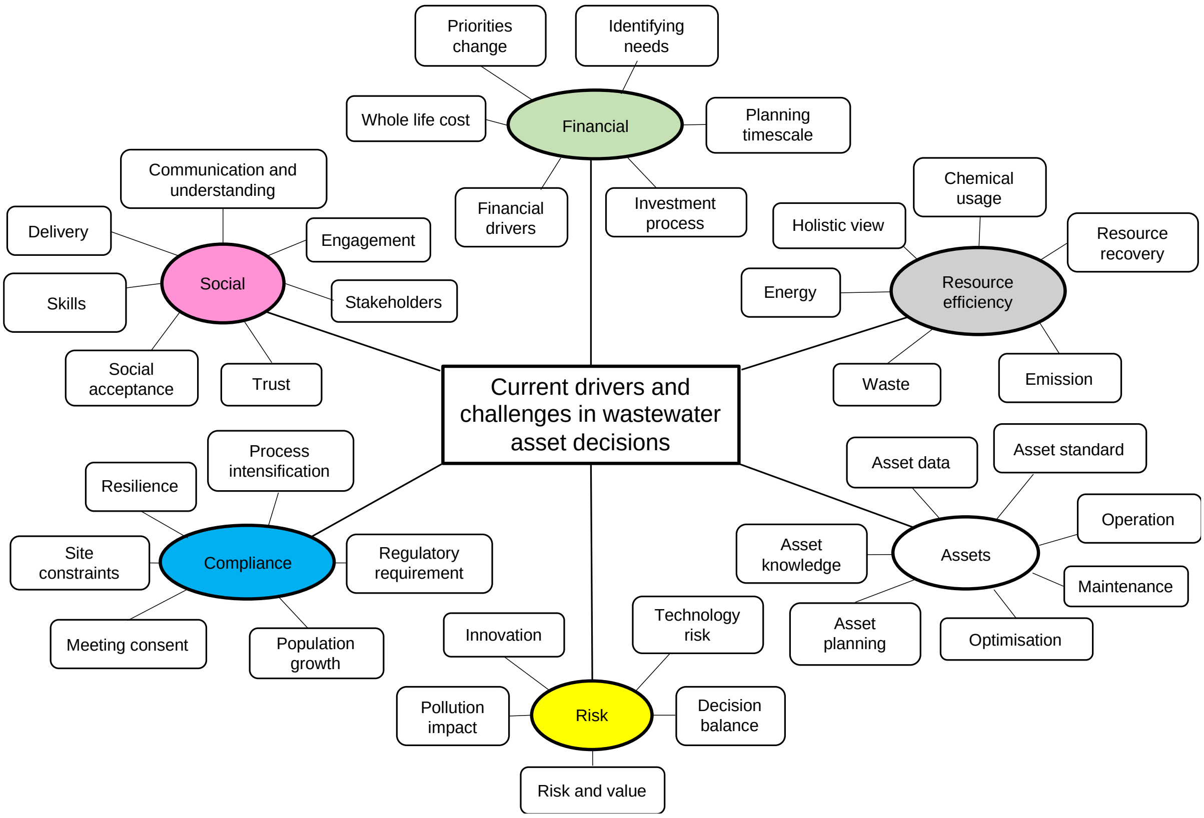


Figure 18. There were 37 secondary codes developed from the second round of coding (Ling et al., 2021)

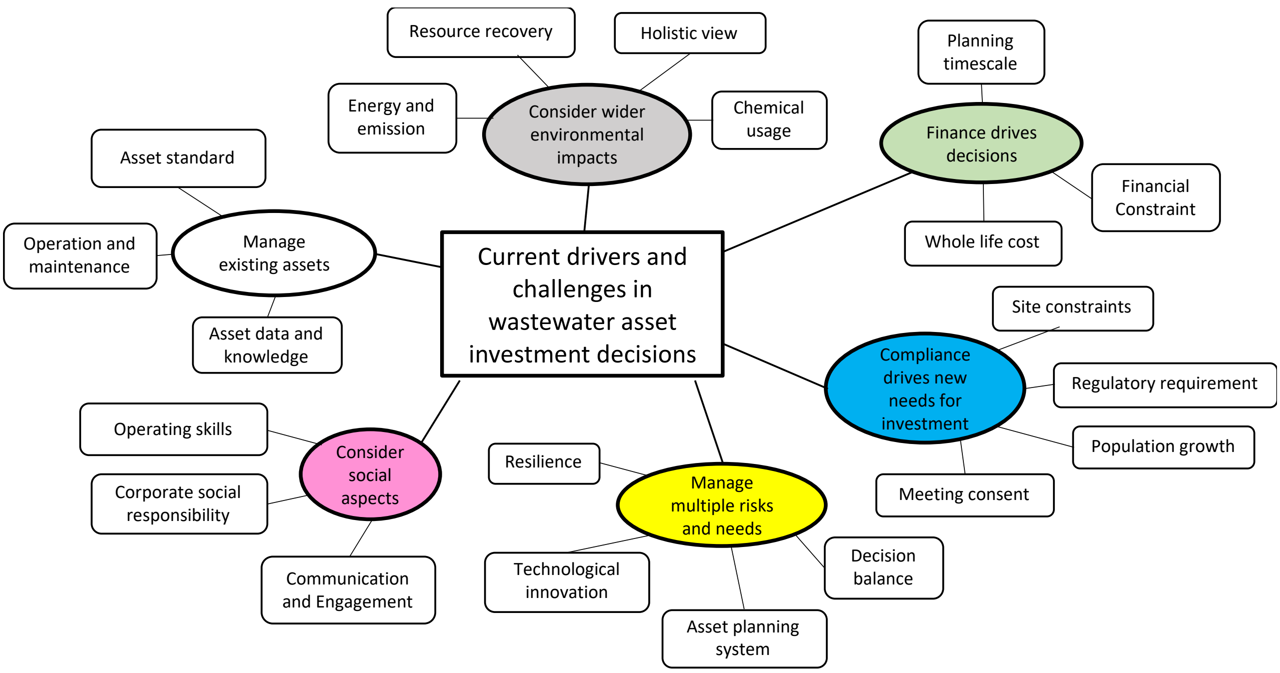


Figure 19. The final 21 codes after the secondary codes were refined and renamed. 6 themes were developed to condense and interpret the meanings of codes

Table 9. The names of descriptions of codes developed from the final round of coding

Codes	Descriptions	Code	Descriptions
Asset data and knowledge	Having access to accurate, complete and quality data is important to identify asset problems and inform investment decisions	Resource recovery	The current sludge strategy is to "recover more, produce less and do it more efficiently". Need to think about wider resource recovery opportunities.
Asset standard	There is an asset standard that includes known technologies to guide option selection based on the needs of sites.	Site constraints	Every site is different in its investment priority. Land shortage drives the selection of alternative solutions.
Chemical usage	Chemical use and its associated carbon emission should be considered as a wider environmental impact	Operating skills	There are implications of adopting new technologies on the requirement of operating skills.
Decision balance	There are balances to be achieved between decision criteria such as Opex VS Capex, and between cost, performance and risks of solutions	Corporate social responsibility	The corporate strategy is about "*doing the right things for customer, environment, employees, wider stakeholders and shareholders*”. We need to demonstrate value for money for investments.
Holistic view	It is suggested to look at wider impacts of decisions and incorporate a system and catchment-based approach.	Whole life cost	Whole life cost is a key consideration in selecting asset investment solutions
Technological Innovation	We need to trial new technologies at a small scale and understand their operation, benefits and risks as part of innovation.	Asset planning system	We have an asset planning system that looks at the needs and risks of investment.
Meeting consents	Meeting sites consent and compliance drives the design and selection of wastewater treatment process.	Communication and engagement	It is important to communicate and engage with our stakeholders to achieve a good process design and its delivery. There is the difficulty of communicating technical information in decision making.
Planning timescale	Decision priorities change and expand over time. There is a challenge in the designing process because of regulation and population changes.	Energy and emission	Energy neutrality and carbon efficiency will become more important in investment decisions. We also want to generate more energy.
Population growth	Population growth drives the design and selection of a treatment process, alongside the quality consent of the site.	Financial constraint	The investment decision can be financially constrained depending on how much budget is left for the current asset management period.
Regulatory requirement	New regulation drives the need for new solutions.	Operation and maintenance	It is important to know how well the assets are operated and maintained. It can be more difficult for new treatment technologies.
Resilience	Resilience is part of the corporate strategy and long-term direction for improving assets.		

The meaning of each theme is summarised below:

1. ***Consider wider environmental impacts***: Net energy and carbon emission are increasingly important environmental drivers when selecting a solution. However, the challenge is to incorporate wider environmental considerations into the decision system.
2. ***Finance drives decisions***: Asset decisions are often driven by financial drivers because it is always constrained in an AMP. Whole life cost is a key consideration when selecting solutions for asset investment. The challenge is that priorities and needs for investment change and expand over time and how to adapt financial resources to address that.
3. ***Compliance***[8] ***drives new needs for investment***: Investments are driven by a few needs such as to meet capacity demand (as a result of population growth and land shortage) and to comply with legal consent at sites (such as discharge qualities).
4. ***Manage multiple risks and needs***: The water company use its asset planning system to evaluate the value, risks and performance of potential solutions. The challenge is to achieve a good balance between them and de-risk the asset base.
5. ***Consider social aspects***: The current corporate strategy is "*doing the right things for customer, environment, employees, wider stakeholders and shareholders*". It is important to communicate the vision and engage with different stakeholders and teams in the decision-making process.
6. ***Manage existing assets***: The water company has an asset standard that provides baseline asset solutions to guide the selection of wastewater treatment processes and technologies for STWs. The challenge is to collect sufficient data and acquire a good knowledge of how well assets are operated, maintained and optimised.

4.2.2 Document analysis

There were 25 codes developed from reviewing all documents (Figure 20). They were allocated one of the six codes categories similar to the thematic data of interviews: *Asset, Environmental aspect, Social aspect, Financial aspect, Compliance and Risk management*.

[8] Here it mainly refers to environmental compliance.

Novel information and codes derived from document analysis were highlighted in blue whereas reoccurring codes were not coloured.

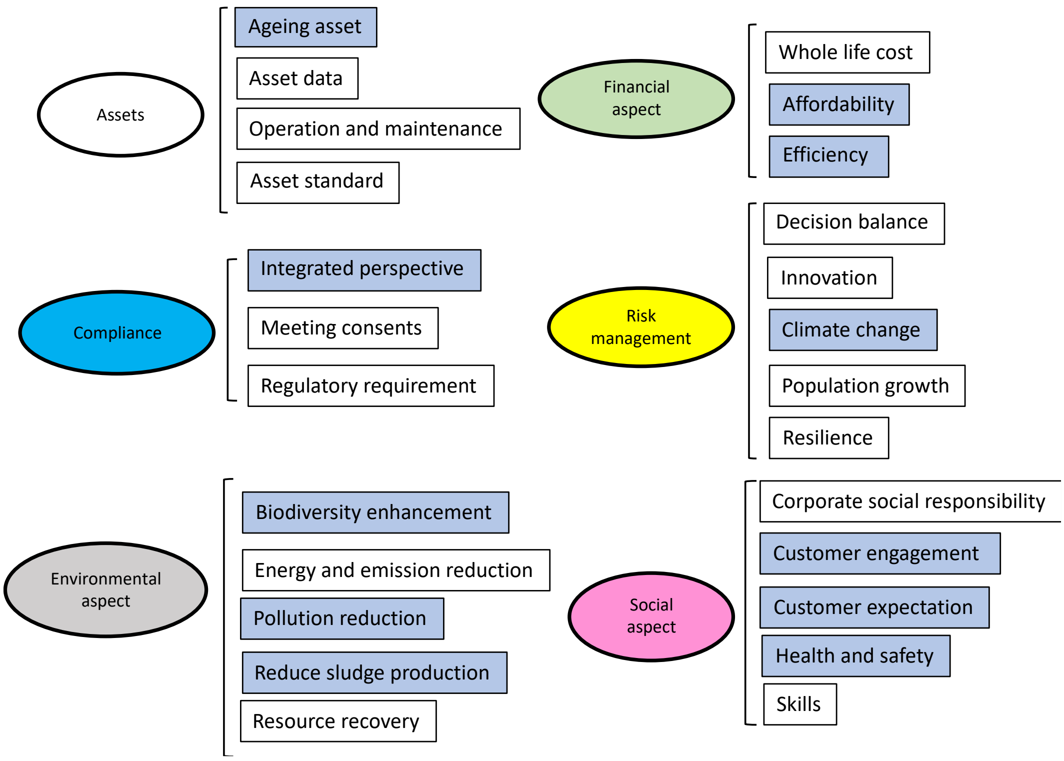


Figure 20. In total 25 codes and 6 categories were developed from the coding process of documents. Novel information was highlighted in blue codes

The descriptions of new codes are summarised below:

- ***Ageing asset***: it was highlighted in some company reports that asset ageing is also a driver for making new investments, especially when the asset is almost at the end of its asset life.
- ***Integrated perspective***: This was a new trend highlighted in the sections of planning strategy to move away from a "*single perspective*". This refers to the integration of multiple team functions and the adoption of a "*catchment-scale*" and systematic planning approach.
- ***Biodiversity enhancement***: this was mentioned as a new environmental commitment proposed in the strategy documents. For example, a target of 5% net biodiversity net gain was proposed for more than 250 operational sites of biodiversity interest (Thames Water Utilities Limited, 2019c). This implies that the environmental strategy will focus on both environmental protection and enhancement.

- ***Climate change***: the impacts of climate change on asset planning and operational risks were highlighted in the documents. It compounds the challenge for long-term asset resilience along with population growth and ageing assets.
- ***Pollution reduction***: many documents also mentioned reduction of pollution is a key environmental performance indicator. For example, the 5-year Business Plan proposed a target of an 18% reduction in serious pollution incidents (Thames Water Utilities Limited, 2019c). This reiterates the significance of environmental protection as part of corporate responsibility.
- ***Reduce sludge production***: some documents mentioned that, as part of sludge management, reducing sludge production is also a key strategy in addition to increasing resource recovery.
- ***Affordability***: the affordability of water and wastewater service is a key consideration in making the business plan. The water company has the responsibility to demonstrate value-for-money for asset investments.
- ***Efficiency***: It is important to identify investments and solutions that are cost-effective and brings the best value for customers, which also ties into the consideration of affordability
- ***Customer expectation****:* The delivery of (water and wastewater) services need to meet customer's expectation. And the level of expectation is increasing.
- ***Health and Safety****:* The health, safety and the wellbeing of employees are also top priorities of corporate strategies.

4.2.3 Convergence of findings

The data analysis of interviews and documents highlighted a few key insights into the drivers and challenges when making wastewater asset decisions. First, wastewater asset investments are driven by a variety of needs and risks. Compliance is one of the top priorities of the water company. This refers to meeting the local effluent quality consents at STWs, complying with regulatory requirements and other environmental commitments. Other risks also trigger investment needs. For example, population growth, as mentioned in multiple sources of evidence, is a major driver for asset investment or upgrade to address the risk of meeting capacity demand. Additionally, climate change and ageing assets (highlighted in documents) are compounding risks in long-term asset planning. And thus, resilience is an important part of investment strategy to mitigate compliance and operational risks, which has been highlighted as a key theme by recent Price Review by the industry regulator (Ofwat, 2017).

Second, making wastewater asset investment decisions are inherently complex and include multiple criteria and stakeholders. A range of environmental, social, financial criteria need to be considered when selecting a wastewater treatment process, as reflected by different code categories in the thematic map (Figure 19). Particularly, finance is a primary driver for that decision because it can be constrained in an AMP. Whole life cost is a key indicator when comparing treatment options. This finding is consistent with the results of decision-mapping by Ashley *et al.* (2003) which suggests that currently a narrow set of economic and technical criteria are included in decision making in the water service providers. Similarly, the review by Hamouda *et al.* (2009) suggests that financial factors such as "*financial viability*" and "*cost minimisation*" are the second major objective after the technical objectives. However, the result of this case study also highlights a trend to include wider environmental impacts such as chemical consumption and carbon emission. In terms of the sludge strategy, there is also a strategic shift from "*reduction*" of waste to "*resource recovery*", which encourages greater recovery of energy and nutrients from the sludge treatment. With respect to the social aspect, corporate social responsibility plays a critical role in that decision making because "*customers are at the heart of the business*" as suggested in both findings of the interviews and documents. This implies that investment decisions are also driven by the needs to demonstrate value-for-money and affordability for customers.

Third, the findings suggest that it is challenging to achieve a balance between different decision drivers and criteria. Although there is a realisation of integrating multiple criteria into the decision system, some interviewees highlight the difficulty to deal with them in the investment decision. Such difficulty can be interpreted in two ways. Firstly, the integration is difficult due to a great number of drivers and criteria in that decision. The selection of an optimal wastewater treatment option not only needs to incorporate the major objective of achieving compliance and fulfilling the investment need but also demonstrate financial viability with wider social and environmental values. Given the trend in current environmental regulations and policies, the number of drivers will continue to increase such as the new carbon reduction target and biodiversity net gain. The second element of that challenge is the integration of different stakeholders and teams in the company. Stakeholders often have different interests and priorities which may give rise to conflicts. Many interviewees indicated that the decision-making process and delivery of solutions should be supported by stakeholder communication and engagement. The inclusion and accommodation of different stakeholders in the decision-making process is key to understanding and resolving this conflict. This recommendation is widely supported in literature on the topic of water management (Ashley *et al.*, 2003; Starkl and Brunner, 2004).

The limitation of this case study is that the findings are bound by the organisational context and the time when this study was conducted. Decision priorities can change and expand over time. It was considered to repeat this study in the later stage of the project to update on the findings. However, it was deemed impractical due to the time required for preparation, recruitment and data collection. There is also limited transferability of the findings to other water companies. The information analysed and generated is unique to the setting and requirement of this water company, although the findings reflect some shared priorities driven by regulations. However, the transferability of the findings is not the priority as the insights derived from this case study needs to inform the development of an assessment tool tailored to wastewater asset decisions in the water company. Another limitation is that there were only 14 participants recruited for the interview and most of them were based in one department of the company. Although the intention was to recruit a wide range of stakeholders across the company, it was observed that only a limited number of stakeholders responded and accepted the interview invitation due to time availability and limited access to contacts. This may affect the comprehensiveness of the themes derived from the interviews and potentially the number of priorities in the subsequent assessment tool. However, based on the existing 14 participants in this study, the information provided from the interviews was relatively consistent and it can be argued that new themes were unlikely to be developed if more stakeholders participated. This was also why additional documents were analysed to corroborate the results of the interviews.

Overall, the findings of this case study suggest that the challenge of selecting a suitable wastewater treatment option lies in the integration of numerous and complex decision criteria and the involvement of stakeholders. This further supports and validates the suitability of MCDA as the assessment methodology as proposed in the literature review (Chapter 2). This case study condensed a basic understanding of the decision and organisational context of the water company and inform the development of the assessment tool in the following stages. It also enabled the researcher to learn more about the water company whilst building rapport with stakeholders in the research process.

Chapter 5 Selection of a Multi-Criteria Decision Analysis model

5.1 Introduction

This chapter provides a comprehensive review of different Multi-Criteria Decision Analysis (MCDA) methods and models in terms of the theories, procedures, potential strengths and weaknesses. A suitable model was then proposed for the underlying operational model in the sustainability assessment tool. This is one of the major steps in the development of the assessment framework for this research, as shown in Figure 13 in Chapter 3.

MCDA is a collective term that includes a range of methods and models that provide decision-makers with procedures to analyse their preference and identify the best alternative according to a preference model (Belton and Stewart, 2002; Niekamp et al., 2015). The limitation of human judgement is that decision making can be influenced by biases and "*recent, memorable or successful experiences*". (Communities and Local Government, 2009). The development in MCDA is stemmed from a shift from a heuristic and intuitive thinking to a more logical and coherent approach when facing multiple and conflicting decision criteria (Keeney and Raiffa, 1976; De Montis *et al.*, 2004; Communities and Local Government, 2009). MCDA explicitly deal with different criteria (and attributes[9]) of a decision problem by applying a specific mathematical model for aggregating them (Belton and Stewart, 2002; Communities and Local Government, 2009; Cinelli *et al.*, 2014). However, MCDA does not ultimately yield an ideal solution because, arguably, it barely exists in reality (Zanakis *et al.*, 1998; Belton and Stewart, 2002; Niekamp *et al.*, 2015). Instead, it provides a coherent and logical process for decision-makers to deal with complexity. And the purpose of MCDA is to aid the decision process rather than replace or automate it (Communities and Local Government, 2009). Moreover, another advantage of MCDA is creating dialogues between decision-makers and enhancing understanding when exploring a complex decision problem (Saaty, 1980; Zanakis *et al.*, 1998; Communities and Local Government, 2009; Cinelli *et al.*, 2014).

[9] Attributes generally refers to specific and quantitative measurements of "*performance associated with a criterion*" (Belton and Stewart, 2002). However, attributes are sometimes used interchangeably with the term '*criteria*' in MCDA. Thus some scholars would also use Multi-Attribute Decision Making (MADM) in literature. (Yoon and Hwang, 1995).

MCDA can solve different types of decision problems such as choice-making, ranking or sorting (French and Roy, 1997). The general process in solving a multi-criteria choice problem include, but not exhaustively (Communities and Local Government, 2009; Jansen, 2011):

- Establishing the decision context, objective and scope
- Identifying options or alternatives for achieving the objectives
- Identifying the criteria or attributes to assess the options. This often involves building a criteria hierarchy or value tree in some MCDA methods.
- Collecting performance of criteria or attributes of all options. Performance data is often compiled and organised into a performance matrix.
- Data analysis: assigning weighting and score aggregation.
- Interpreting results and informing decision

There are dozens of MCDA methods or models available to use. Each method was developed based on a distinctive set of procedures, theoretical bases and axioms. Although the classification of MCDA has been diverse across studies, there are generally three schools of models and methods, namely: Utility/Value theory-based models, Goal Programming, and Outranking methods (Belton and Stewart, 2002; Lai *et al.*, 2008; Ishizaka and Nemery, 2013). The fundamental difference between various MCDA methods is "*how to model and articulate preferences of decision-makers*" when facing multiple decision criteria (Guitouni and Martel, 1998). This difference is manifested in two ways (Belton and Stewart, 2002). First, the model determines how would the decision-maker evaluate the importance or desirability based on different performance levels of each criterion or attribute. Second, it also refers to how to aggregate preference of multiple criteria for each alternative in order to make comparisons between alternatives. Although it has been recognised that the selection of a MCDA model is often motivated by familiarity and affinity (Guitouni and Martel, 1998; Cinelli *et al.*, 2014), an in-depth review was considered necessary for proposing a suitable MCDA model or method for this research (section 5.2 to 5.5). The detailed technical procedures of each model are not included in this chapter but are presented in later chapters where it is applied to the assessment tool.

5.2 Utility and Value theory models

5.2.1 Multi-Attribute Value Theory (MAVT)

One of the fundamental MCDA models is based on the Multi-Attribute Value Theory (MAVT). It is derived from the Expected Utility hypothesis which was initiated by Daniel Bernoulli in 1738. It was observed when people making decisions with uncertainty, the option associated with the largest utility was often chosen rather than the one with maximum monetary gain. The utility of an outcome can be defined as its potential "*value, benefits and desirability*" (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). The maximisation of Expected Utility (*EU*) is the principle for selecting a desirable option and it can be determined by a probability-weighted additive model as:

$$EU = \sum_{i=1}^{n} p_n u_n \qquad \text{Equation 1}$$

Namely, the expected utility U of an option i equals the sum of products of partial utility u and probability p of occurrence. This concept of expected utility has been further expanded by Neumann *et al.*, (1947) and Keeney and Raiffa (1976) to riskless[10] outcomes. They highlighted the use of value function to capture decision-maker's preferences and homogenise the units of different attributes for aggregation. Similar to the principle of utility maximisation, the objective of MAVT is to select an alternative that provides the maximum aggregated value. Keeney and Raiffa proposed a linear additive model in MAVT which the probability (p) is replaced by weights (w) as known as

$$v_i = \sum_{i=1}^{n} w_n v_n = w_1 v_1 + w_2 v_2 + \cdots + w_n v_n \qquad \text{Equation 2}$$

where n is the number of attributes and v is the partial value function of individual attributes for option i. Weights in value aggregation have two layers of meaning. It can indicate the relative importance (or referred to as importance coefficient) of all criteria. Mathematically, it can serve as the scaling factors for trade-off (Yoon and Hwang, 1995; Pöyhönen and Hämäläinen, 2001; Belton and Stewart, 2002). This concept of scaling factor can be expressed mathematically:

---

[10] While the utility theory is concerned with outcomes with risks and probability, Multi-Attribute Value theory (MAVT) is associated with riskless options (von Winterfeldt and Edwards, 1986; De Montis *et al.*, 2004). It does not entail options do not have risks but rather excluding them from the mathematical model.

For any two alternatives a and b with two attributes x and y (Figure 21). Supposed these two alternatives are perceived indifferent (symbol '$\sim$') with the same aggregated values, i.e. $a \sim b$ as $V(a) = V(b)$, then:

$$w_x v_x(a) + w_y v_y(a) = w_x v_x(b) + w_y v_y(b) ,$$

So $w_x \cdot \Delta v_x = w_y \cdot \Delta v_y$, Then $\frac{w_x}{w_y} = \frac{\Delta v_y}{\Delta v_x}$

(given that $\Delta v_x = |v_x(a) - v_x(b)|$ and $\Delta v_x = |v_y(b) - v_y(a)|$)

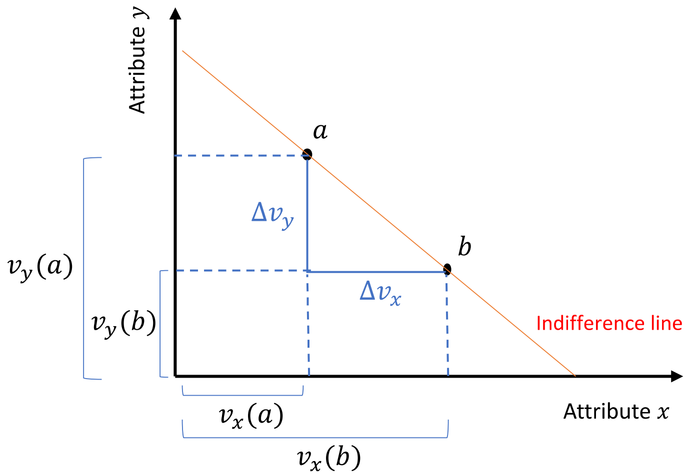


Figure 21. Illustration of the weights of attributes between two alternatives that are indifferent to each other. The indifference line represents the set of options that are indifferent to decision-makers.

In order to validate the concept of preference in MAVT, it must satisfy the underlying axioms in terms of preferential relations (these also apply to other MCDA models based on value theory):

- *Asymmetricity:* for any pair of alternatives *a* and *b* that belongs to set *A* (*a, b* ∈ *A*), the preference order cannot be *a* ≻ *b* (i.e. *a* is preferred to *b*) and *b* ≻ *a* (i.e. *b* is preferred to *a*) simultaneously.
- *Transitivity*: for any alternative *a*, *b* and *c (a, b, c* ∈ *A*), if *a* ≻ *b*, *b* ≻ *c*, then *a* ≻ *c*
- *Completeness*: for any pair of alternatives *a* and *b* (*a, b* ∈ *A*), the preference relation can only be *a* ≻ *b*, *a* ≺ *b* or *a* ~ *b* (i.e. *a* is indifferent to *b*).

In addition to the axioms above, a few conditions also need to be met to validate a linear additive value function:

- **Preferential independence** (mutual/pairwise): The preference and trade-off between attributes x and y are irrespective of the level of other attributes. For example, the trade-off between criterion environmental benefit and social benefit does not take the level of cost into account. For a detailed explanation, please refer to Belton and Stewart (2002), p. 88. Such preference independence should also be mutual if there are more than two attributes, as suggested in Keeney and Raiffa (1976), p. 105 (Theorem 3.3)
- **Corresponding trade-offs condition**: Given any two attributes x and y, the amount of trade-off that the decision-maker is willing to make in attribute x is corresponding with the level of x only, irrespective of the level of y, and vice versa. Detailed illustrations of this condition can be found in Belton and Stewart (2002), p. 89, and Keeney and Raiffa (1976), p. 90.

One of the major strengths of MAVT is it has a rigorous theoretical and axiomatic foundation. The use of the linear additive model to aggregate attributes is also relatively simple to use. However, there are practical limitations when using MAVT. First, it provides limited decision support as it does not indicate how to determine the value functions of individual attributes (Communities and Local Government, 2009). For example, if the global value function of option i can be determined by only two attributes x and y, there lies the question of what are the partial value function of x and y respectively? Such an additive model can be expressed as:

$$v_i = w_x v_x(x) + w_y v_y(y) \qquad \text{Equation 3}$$

There are many types of value functions (examples in Figure 22) and it is important that they should reflect the decision-maker's value judgement accurately. However, there is not a universal approach to develop partial value functions $v_x(x)$ and $v_y(y)$. The construction of partial functions often involves asking the decision-maker a series of questions on how much value is perceived based on a given magnitude in an attribute. Thus, the process of constructing a rigorous mathematical function can be repetitive and time-consuming (Cinelli *et al.*, 2014). Second, a linear additive model may not suffice if the condition of preferential independence is not satisfied. In a real-life scenario where attributes are not preferentially independent of each other, other aggregation models such as a multiplicative model may be more appropriate. However, without specialised knowledge and experience, decision-makers could face the difficulty of constructing such a multi-attribute model and operationalising it confidently.

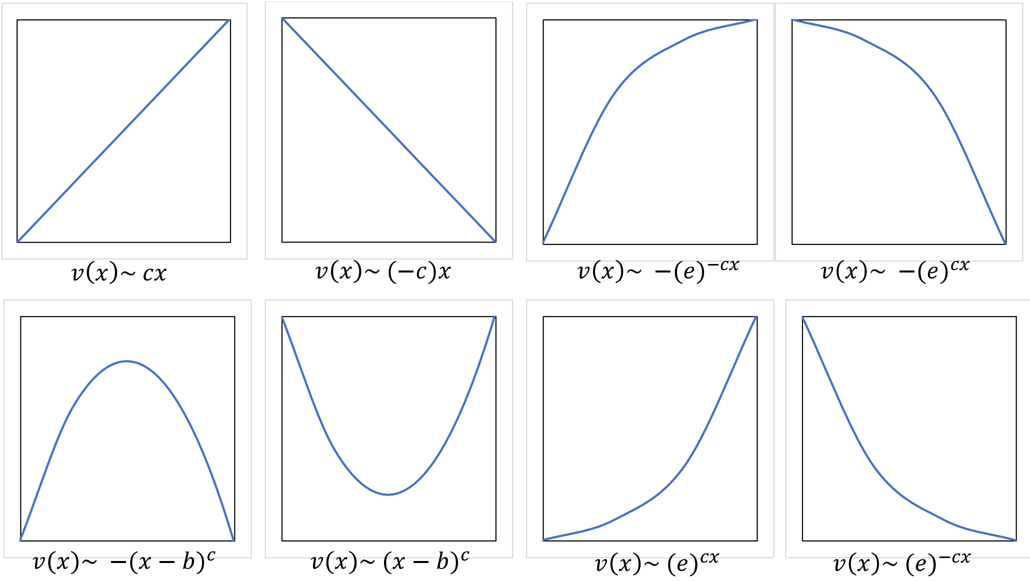


Figure 22. Possible shapes of different partial value functions. The symbol ~ means the function is "proportional to". Adapted from von Winterfeldt and Edwards (1986), p.238

5.2.2 Simple Additive Weighted (SAW) model

The Simple Additive Weighted (SAW) model is one of the most popular models in MCDA (Yoon and Hwang, 1995). It shares similar theoretical origins and mathematical operations to MAVT. SAW also uses a linear additive model to aggregate partial values (see Equation 2) with weightings from all attributes. The only difference of SAW is that it uses a linear normalisation technique to derive partial value function [11] and achieve homogenisation of attribute measurements (Yoon and Hwang, 1995; Prasetiyo and Baroroh, 2016). Before applying a linear normalisation, decision-makers need to identify the type of each attribute based on the direction of value increment. Attributes that provide increasing value with an increasing level of performance are referred to as 'positive' or 'benefit' attributes, whereas those that provide decreasing values are referred to as 'negative' or cost' attributes. For example, in the context of wastewater treatment, positive attributes can be biogas production or reliability of

[11] The partial value functions in SAW are considered as "pseudo-value functions" (Yoon and Hwang, 1995) because linear normalisation only produces a normalised value rather than a function of value.

wastewater treatment technology, as they are to be maximised; Negative attributes can be operational expenditure or energy consumption. The goal of normalisation is to convert the attribute scales to a value range from 0 to 1 for comparisons and further aggregation. Common linear normalisation techniques are the Linear Sum method, the Linear Max method and the Linear Max-Min method (Cinelli *et al.*, 2014; Miranda and Prasanna, 2014). The normalised performances in all attributes are then multiplied by their respective weights for aggregation. As SAW is based on the value theory, the same axioms mentioned above are also required to validate this model.

5.2.3 Simple Multi-Attribute Rating Technique (SMART)

Following the development of MAVT by Keeney and Raiffa in the 1970s, it was realised that constructing partial value functions in MAVT can be difficult. There was a motivation that "*simpler tools are easier to use and more likely to be useful*" (Edwards and Barron, 1994). Simple Multi-Attribute Rating Technique (SMART) was, thereafter, developed by Edwards (1977) using a linear approximation to derive partial value contribution of each attribute. It allows users to estimate the value of an attribute directly on a scale of 0-100 proportional to the range of attribute performance by using one of four pre-defined linear functions (Figure 23). The use of them is dependent on the type of attribute: Type (a) applies to positive attributes; Type (b) applies to negative attributes; Type (c) is suitable when there is an interior maximum in the value system; Finally, type (d) applies to intangible or unquantifiable attributes where direct rating can be applied (Edwards and Barron, 1994). The approximation of value measurement entails that the alternative that has the worst performance on an attribute x should be allocated with a value of 0 whereas 100 for the best performance. As such, the performance of different attributes can be converted to a homogenised scale for further aggregation.

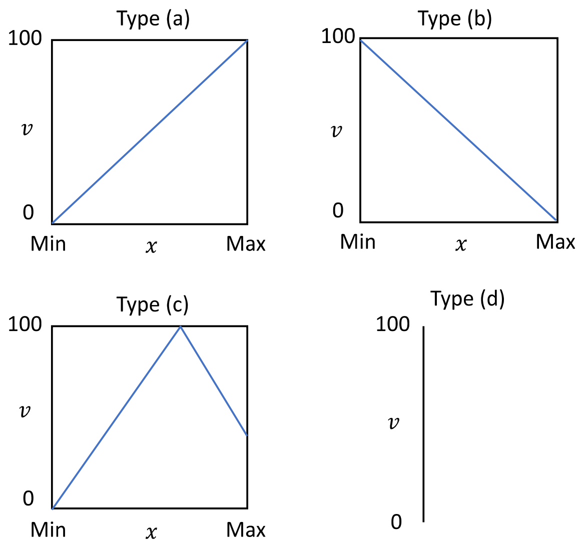


Figure 23. Four types of linear approximation functions to provide value approximation (Edwards and Barron, 1994).

Edwards believed that SMART has the potential of providing simplicity and robustness. SMART requires no judgements of preference from decision-makers (Olson, 1996, p.34). However, it is important to ensure that simplification does not lead to erroneous value estimation. There are two conditions that validate the use of linear approximation in SMART (Edwards and Barron, 1994). First, the real value function of the attribute should be monotonic[12]. If the value of an attribute can both increase and decrease depending on the attribute performance, then linear approximation may not be appropriate (unless it is a type c function in Figure 23). Second, the greater the curvature of a value function, the less suitable will be for a linear approximation. One way of checking the assumption of linearity is to ask decision-makers: "if the performance of an attribute is increased by a fixed level, what is the marginal gain or loss in the value?" After some repetitions, if the marginal gains or losses are considerably different, then the condition of linearity may not be satisfied.

5.3 Analytical Hierarchy Process

Developed by Thomas Saaty (Saaty, 1980), Analytical hierarchy process (AHP) is the most widely applied MCDA method (Sipahi and Timor, 2010; Ossadnik *et al.*, 2016; Dos Santos *et al.*, 2019). The objective of AHP is to assist decision-makers to solve a complex decision problem by translating complexity into a manageable and understandable preference structure which is called a criteria hierarchy (Yoon and Hwang, 1995; De Montis *et al.*, 2004).

[12] A monotonically increasing function must not decrease, or vice versa.

The criteria hierarchy often consists of layers of decision elements including the overall objective, criteria, sub-criteria, indicators and alternatives (Figure 24). The principal task for decision-makers in AHP is making pairwise comparisons between criteria as well as between alternatives. The fundamental question to be asked is "*how important is item a compared to item b*?". A 9-point scale (often referred to as the AHP judgement scale) is used to provide preference judgement (Table 10). For example, if the decision-maker thinks item *a* is moderately more important or better than item *b*, a scale number 3 is selected. If the direction of the preference is the opposite (i.e., item *b* is moderately more important than *a*), then the reciprocal value (i.e., 1/3) is selected. Responses from pairwise comparisons are then compiled into a reciprocal matrix to determine preference scores (which is called "priority" in AHP) of alternatives and criteria weights (Saaty, 1980). Finally, the performance scores are aggregated by multiplying corresponding criteria weights to calculate a global priority score for each alternative using the same linear additive model in MAVT.

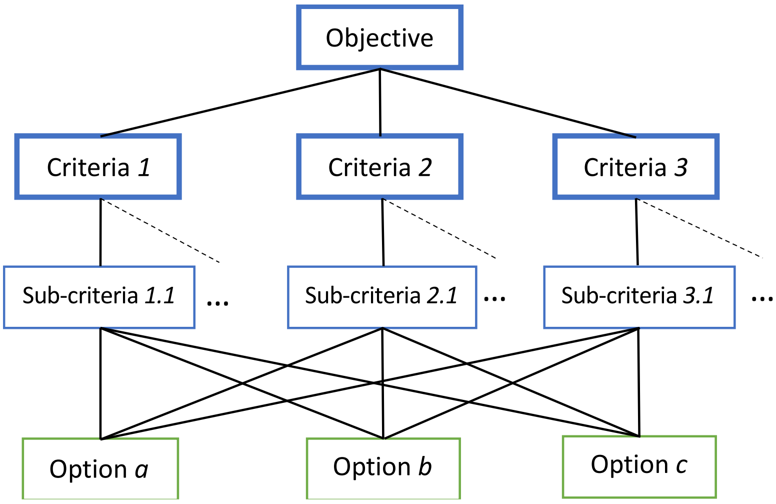


Figure 24. A visual example of the hierarchical structure to be developed in problem solving using AHP.

Table 10. AHP judgement scale for making pairwise comparisons (Saaty, 1987, 2002)

Scale intensity	Definition
1	Equally important
3	Moderately more important
5	Strongly more important
7	Very strongly more important
9	Extremely more important
2,4,6,8	Intermediate values between two scale points
Reciprocals	The preference order is inversed

AHP has demonstrated its simplicity and thus gained popularity in practical decision-making. According to Huang *et al.* (2011), It is the most widely used MCDA approach in the environmental field. It offers a user-friendly process for decision-makers to perform preference judgement with a standardised process. Pairwise comparison using the ratio scale is simple to perform and it does not rely on any specialised knowledge or experience of MCDA. However, it has also attracted debates from MCDA theoreticians and practitioners in terms of its theoretical and technical validity (Bernasconi et al., 2010). Some literature argued that AHP is inconsistent with the concept of utility theory and the normative goal of utility maximisation (Dyer, 2008). The defending statements pointed out that AHP was developed independently of utility theory and decision science, and it is more based on the theory of measurement (Harker and Vargas, 1987; Saaty, 1990; Forman and Gass, 2001). The argument on the technical aspect focused on the robustness of using scale ratio and eigenvector for preference modelling and weight calculation (Belton and Gear, 1983). Specifically, the preferential judgement based on a ratio scale can be internally inconsistent (Davison and French, 1987; Communities and Local Government, 2009). Later Saaty and his colleagues argued that scale ratio can sufficiently perform preference estimation although it is not perfectly accurate (Saaty, 2003). A consistency ratio has been developed in AHP to check the consistency of the preference judgement that feeds into the reciprocal matrix.

In terms of practical concern, users should be aware of the time factor when dealing with a large number of items in the criteria hierarchy (Communities and Local Government, 2009). For example, 45 pairwise comparisons are needed if a hierarchical layer contains 10 criteria or alternatives. This can be time-consuming and repetitive for decision-makers. Another concern of using pairwise comparisons is the rank reversal phenomenon (Belton and Gear, 1983; Velasquez and Hester, 2013). This implies that the ranks of the desirability of alternatives may change if any option is removed or added. This can be checked by conducting additional sensitivity analysis (Saaty, 1994).

5.4 Goal Programming: TOPSIS

The rationale behind Goal Programming is to set up a desirable goal or a satisfaction level and identify options that are closest to achieving that goal using a mathematical programming algorithm (Belton and Stewart, 2002, p. 9). One of the popular MCDA methods utilising the

Goal Programming principle is the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Developed by Hwang and Yoon (1981), TOPSIS is based on the principle that the best alternative should have the shortest distance to the ideal solution[13] and the longest distance to the worst ideal solution[12]. Therefore, one of the key steps is to model the ideal and worst ideal solution. Mathematically, the ideal solution is a set consisting of the best performance of all attributes whereas the worst ideal set consists of the worst performance. Then the geometric distances (Euclidean distance) between each attribute of each alternative and the best/worst ideal solution are determined (Illustrated in Figure 25). A similarity index is then calculated to reflect the relative 'closeness' of each alternative to the ideal solution. The alternative with the highest similarity index is considered the most desirable. Before the modelling of the best and the worst solution, the performance of attributes needs to be normalised. Thus, the key assumption in TOPSIS is the monotonicity of the underlying utilities of each attribute.

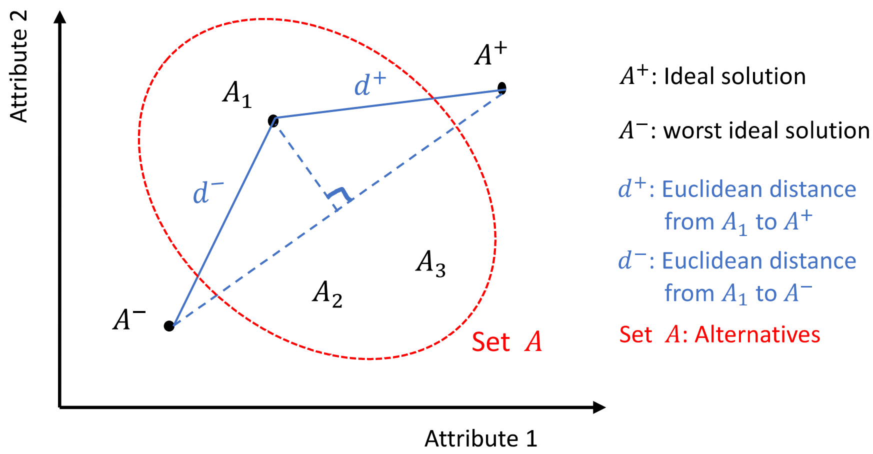


Figure 25. Illustration of the Euclidean distances for alternative A_1 in a two-dimensional space in TOPSIS. Adapted from Yoon and Hwang (1995).

The advantage of TOPSIS is that its rationale is relatively simple to understand and intuitive for decision making (Kim *et al.*, 1997; Roszkowska, 2011). It is also relatively easy to compute in spreadsheets and results can be easily visualised (Roszkowska, 2011). The major weakness

[13] In TOPSIS terminology, the ideal solution refers to Positive Ideal Solution (PIS) or A^+ whereas the worst ideal solution refers to Negative Ideal Solution (NIS) or A^-.

is that Its procedure does not include weight elicitation (Shih *et al.*, 2007; Roszkowska, 2011). TOPSIS may also be subject to potential rank reversals because the distances to the best and worst ideal solution may change if the number of alternatives changes (García-Cascales and Lamata, 2012). However, TOPSIS appears to have the fewest rank reversals among other MCDA methods according to the simulation by Zanakis *et al.* (1998).

5.5 Outranking method: ELECTRE

ELECTRE (ELimination Et Choix Traduisant la REalité[14]) is an outranking method that fundamentally applies the concept of dominance when comparing alternatives. The most common outranking method is the ELECTRE developed by Roy (1971). In contrast to the value theory-based MCDA models, ELECTRE establishes preference models through determining outranking relations based on pairwise comparisons of every two alternatives. Alternative *a* is said to outrank *b* when there are strong arguments that *a* is considered at least as good as *b* without strong evidence to refute it (Belton and Stewart, 2002). ELECTRE[15] uses two parameters, concordance and discordance index, to build those arguments (Figueira *et al.*, 2010, p. 59). Figure 26 illustrates a simplified process of building an outranking argument. Specifically, the concordance index measures the strength of evidence to support the argument that one alternative outranks another. The discordance index measures the degree of disagreement in this outranking relation. Threshold values are then applied by decision-makers to both indices to help determine whether there is an outranking relation between the two alternatives. Generally, an option is considered more desirable when it has a concordance index higher than the chosen threshold value and its discordance index lower than the threshold value (Communities and Local Government, 2009).

[14] The English translation of the term is ‘ELimination and Choice Expressing REality’

[15] There are many versions of ELECTRE available. This chapter describes ELECTRE I which deals with choice type of decision problem. For a comprehensive review of other versions, please see Govindan and Jepsen (2016) and Figueira *et al.* (2016)

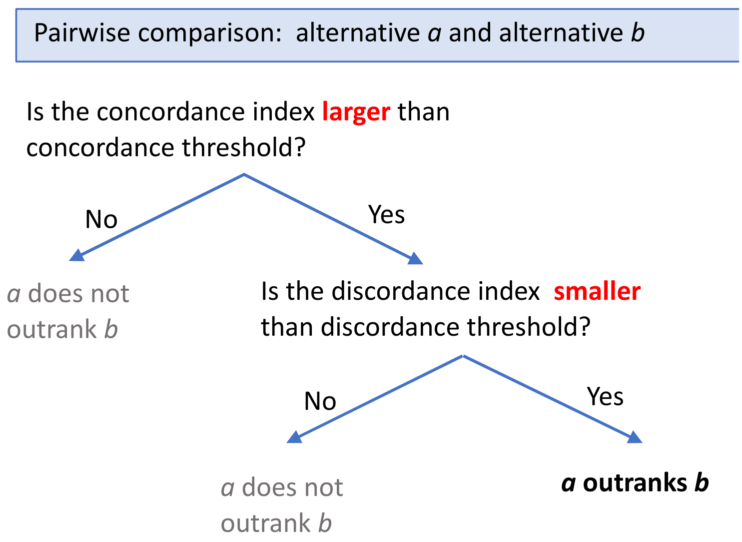


Figure 26. An illustration of the reasoning behind the determination of an outranking relation between two alternatives. Adapted from Belton and Stewart (2002)

The unique feature and also a potential strength of ELECTRE is its mode of elimination. As a non-compensatory[16] approach, a unique feature of outranking methods is that it seeks to eliminate alternatives that have a poor performance score on any one criterion. Therefore, it does not allow trade-offs between different criteria and attributes (Figueira *et al.*, 2010). The strength of this is that to a certain extent it reflects the reality where an alternative will not be considered acceptable if it performed poorly on a single criterion. Additionally, as this method was developed separately from the utility and value theory, it does not need to satisfy the same axioms. For example, the preference relations in ELECTRE do not have to be logically transitive (i.e. if $a \succ b$ and $b \succ c$, $c \succ a$) (Yoon and Hwang, 1995).

However, there are a few concerns before using an outranking method. First, the concept of outranking can be difficult to understand and the mathematical operations can be complicated and even counter-intuitive for new users (Belton and Stewart, 2002). In addition, the result of outranking is strongly sensitive and subjective to the parameter input of the threshold values for the concordance and discordance indices (Yoon and Hwang, 1995). Due to the operation of pairwise comparisons, rank reversals are also possible (Wang and Triantaphyllou, 2008). Moreover, the ELECTRE I version can only be used for selecting the best

[16] Non-compensatory refers to the condition of trade-offs between performance of criteria is not entirely or partially allowed.

alternative but not ranking alternatives because it does not assign an aggregated score to them (Figueira *et al.*, 2010).

5.6 Comparative review

To move beyond familiarity and affinity when selecting an appropriate MCDA method, a holistic list of aspects need to be considered. It can be said that selecting an appropriate MCDA model is a multi-criteria problem in itself (Yoon and Hwang, 1995; Govindan and Jepsen, 2016) as each one presents strengths and weaknesses. Table 11 summarises the six models and methods given a list of common criteria for selection (Roszkowska, 2011; Cinelli *et al.*, 2014). Discussion of each suitability category (theoretical foundation, robustness, feasibility) was provided below and a tentative MCDA model was thereafter proposed.

Table 11. A comparative summary of six different MCDA models or methods. Comparison dimensions are adapted from Cinelli *et al.*, (2014) and Roszkowska (2011)

Dimension	Comparison criteria	**MAVT**	**SAW**	**SMART**	**AHP**	**TOPSIS**	**ELECTRE**
Theoretical foundation	Preference model	Value function	Pseudo-Value function (Linear normalisation)	Value function (Linear approximation)	Pairwise comparisons (Ratio scales)	Goal programming by geometric distance	Pairwise comparison outranking relation
	Aggregation model	Weighted additive	Weighted additive	Weighted additive	Weighted additive	Similarity index	Concordance; Discordance
Robustness	Weighting included?	No	No	Yes	Yes[2]	No[2]	No
	Compensation allowed	Yes[1]	Yes[1,2]	Yes[1,2]	Yes[1]	Yes[2,4]	No[1,2]
	Rank reversal	Not possible[1]	Not possible[1]	Not possible[1]	Possible[1,2,6]	Possible[2]	Possible[1]
Feasibility/ Applicability	Ease to learn and use	Difficult[1,2,5]	Easy[2]	Easy[2]	Easy[2,6]- Medium[5]	Easy[2,5]- Medium	Medium[5]- Difficult[1,2,3]
	Software dependency	High	Low	Low[6]	Medium	Medium	High

1. (Cinelli *et al.*, 2014); 2.(Velasquez and Hester, 2013); 3.(Belton and Stewart, 2002); 4.(Communities and Local Government, 2009); 5. (Ishizaka and Nemery, 2013); 6. (Németh *et al.*, 2019);

5.6.1 Theoretical foundation

MAVT, SAW and SMART share the same preference and aggregation model as they all utilise value theory and the principle of value maximisation. Although AHP elicits preference judgements by determining ratio scales through pairwise comparisons rather than building value functions, it also adopts a weighted sum aggregation model to aggregate multiple criteria to calculate a 'global' score for each alternative. TOPSIS operates from Goal Programming and determines preference by calculating the distances to the reference ideal scenarios. ELECTRE focuses on building outranking relations by calculating the concordance and discordance index between every two alternatives. Overall, value theory-based models (MAVT, SAW, SMART) and AHP are quite similar whereas the other two are distinctive in their underlying theories and how they operate.

5.6.2 Robustness

5.6.2.1. Weighting

Weights are commonly used to reflect ratios for value trade-offs when aggregating criteria and attributes. Alternatively, they are considered as of importance coefficients (Cinelli *et al.*, 2014). Most MCDA models do not include a guideline to calculate weights except for AHP and SMART. As mentioned above, pairwise comparisons in AHP can be either used to determine the preference of alternatives or weights of criteria. SMART has been further developed to SMARTS (Simple Multi-Attribute Rating Technique with Swing weights) and SMARTER (SMART Exploiting Rank) by Edwards and Barron (1994) to improve the functionality of SMART. Swing weights are used in SMARTS to also reflect the range or spread of performance of an attribute as Edwards and Barron (1994) argued that the importance of an attribute also depends on the value range. SMARTER further simplified the elicitation of swing weights by directly converting ranks of attributes into weights using Rank Order Centroid (ROC) weights (please see Barron and Barrett, 1996). Besides SMART and TOPSIS, other MCDA model requires combining a separate weighting technique to derive attribute or criteria weights.

5.6.2.2. Compensation

Most MCDA models are based on a compensatory approach, which implies that a better performance in one attribute or criteria can compensate for an inferior performance in another. And the degree of trade-offs is reflected by the criteria weights. ELECTRE does not allow compensations between criteria as it includes threshold values which eliminate alternative that has a poor performance in any criteria (Belton and Stewart, 2002; Cinelli *et al.*, 2014). The strength of a non-compensatory model is that it complies with a strong sustainability perspective (i.e., non-reductionist approach) so that trade-offs between the environmental dimension and other dimensions are not allowed. However, it is not realistic to completely exclude trade-offs in complex engineering projects (El Amine *et al.*, 2014). Additionally, from the results of the previous interviews, wastewater asset decisions aim to reach a "*balance*" between different criteria. This implies that trade-offs are inevitable and therefore a compensatory MCDA model would be suitable.

5.6.2.3. Rank reversals

Rank reversal is another technical concern in MCDA because it affects the consistency of the decision results if assessment parameters change. This issue potentially exists in any model that incorporates pairwise comparisons as part of the preference modelling such as AHP and ELECTRE. Due to the nature of pairwise comparisons, removal or addition of an attribute or criterion may lead to changes in the ranking of alternatives and misleading results. Rank reversal may also occur in TOPSIS. Although it does not use pairwise comparison, the preference score is dependent on the geometric distances to the ideal and worst ideal solution. By the mathematical definitions of ideal and worst ideal solution, the removal and addition of alternatives can alter the distances to these two reference points. One way of addressing rank reversal is to include a sensitivity analysis to examine the sensitivity of results to changes in the parameters. In comparison, the preference order in value theory related models (MAVT, SAW, SMARTS) does not usually reverse because the calculation of the preference score of each alternative is independent of that of another alternative.

5.6.3 Feasibility

Feasibility is particularly important for the practical success and implementation of MCDA as a decision support tool because the end-users in the water company may not have the specialised knowledge or resources to use it. This is aligned with the trend identified in Lai *et al.* (2008) that less complex MCDA methods are more preferred in water management decisions. Hobbs *et al.* (1992) Also stated that ease of use should be considered when selecting a MCDA method. MAVT and ELECTRE are less preferred to other models because they are difficult to learn and use. Specifically, building partial value functions in MAVT can be complex and time-consuming (Cinelli *et al.*, 2014). And it requires decision-makers to be equipped with some existing knowledge and great cognitive capabilities if a large number of attributes are considered. Similarly, ELECTRE has been recognised as a difficult method to learn and use, mainly because the concept of outranking is complex and less intuitive (Cinelli *et al.*, 2014). SAW is well known for its ease to use and popularity in MCDA applications. There are various comments about the simplicity of AHP and SMART. AHP is easy to understand in general. However, it can be resource-intensive depending on the complexity of the criteria structure, especially when there are many items to be compared. The development of SMART is to alleviate the requirement of building partial value functions, and on this note, it is easier to use than MAVT. However, the use of SMART is more limited due to the requirement of linearity of value functions.

Although specialised software is increasingly available to simplify the procedures for users (such as those in Ishizaka and Nemery, 2013, p. 5), the dependency on them implies extra resources such as the costs for purchasing user licences and time to learn and use them as well as potential data privacy concerns in a business environment. In light of this consideration, MAVT and ELECTRE were provisionally excluded from this research due to a high requirement of specialised software and difficulty to use.

5.7 Proposing a suitable MCDA model

The theory and practice of MCDA science are closely intertwined and mutually complementary (Fishburn and Lavalle, 1999). Therefore, it is difficult to pinpoint the most suitable MCDA model without the specific context of a case study and practical experience of using them. Nevertheless, based on the comparative discussions above, it seems that SAW, SMART, AHP and TOPSIS are more feasible than MAVT and ELECTRE. Particularly, SMART and AHP are the most preferred ones because they also include weighting calculations in their methodology. Rank reversal is less likely to occur in SMART than AHP, and therefore, SMART was provisionally proposed as the underlying MCDA model to build into the sustainability assessment tool. However, SMART was later considered unsuitable because it requires the range of performance data of indicators to develop weights using swing weights, which was not available at the early development stage. Therefore, AHP was selected as the alternative MCDA methodology for building the assessment tool.

In this chapter, only the fundamental schools of models and methods were reviewed. There are many new MCDA models developed in recent years to improve on an existing method or to extend its application in wider fields. For extensive reviews of all MCDA models or methods, please refer to Guitouni and Martel (1998) and Wątróbski *et al.* (2019). In our opinion, the list of models reviewed in this chapter should be sufficient to identify a suitable model for this research. Additionally, Yoon and Hwang (1995) argue that the most important key to the success of performing MCDA is the development of the attributes and criteria structure because the results of different MCDA models might not differ significantly (for example see Hobbs *et al.*, 1992).

Chapter 6 Development of assessment indicators and weighting

This chapter describes the process and results of the development of key components in MCDA: assessment indicators and weightings. Once the assessment indicators and weightings were proposed, they were applied to a business case as a pilot study to discuss their reliability and feasibility. The content of this chapter is primarily based on the publication Ling *et al.* (2021) with some extensions[17].

6.1 Selection of assessment indicators

6.1.1 Literature review

The development of relevant assessment criteria and indicators is a critical step to provide measurements towards the decision objective in MCDA. For this study, a suite of relevant sustainability criteria and indicators were proposed based on information from the literature review and the results of the case study in chapter 4. Balkema et al. (2002) present a comprehensive review of all sustainability indicators used to assess wastewater treatment systems. However, the review by Balkema et al. (2002) only covers studies published before 2000. Therefore, indicators from recent studies were reviewed in this research and summarised in Table 12. As those studies have different scopes and goals, the following sub-section critically discussed the relevance and significance of each indicator based on the requirement of this research and the context of the water company. The findings from the previous case study in Chapter 4 were also used to support this discussion. A list of indicators was selected for the sustainability assessment as a result of the discussion is presented in section 6.1.2.

6.1.1.1 Environmental and technical indicators

Environmental and technical criteria have a greater number of indicators compared to the social or economic indicators, reflecting the growing importance of environmental impacts of wastewater treatment plants (Mustapha et al., 2018). Molinos-Senante *et al.* (2014) define some major aspects of environmental impacts in wastewater treatment as "*quality of*

[17] The content of publication is an inseparable part of research development in this thesis. Specifically, the same data, methods and results are presented in this chapter.

effluent", "*resources used*", "*emission*" and "*the potential for the technologies to be updated to implement additional processes"*.

Table 12. Common indicators to assess wastewater treatment processes and systems from literature reviews. Units vary across different references, and some normalised the units by land or population equivalent. Adapted from Ling et al. (2021).

Dimension	Indicator	Unit	References
Environmental and Technical	Pollutant removal potentials	%	a, b, c, d, e, h
	Energy consumption	kWh/yr.	a, b, c, d, e
	Greenhouse gases emission	$kgCO_2$ eq/yr.	e, h
	Land required	m^2/p.e.	a, b, c, d, f, g
	Resource recovery potential	(depends)	b, c
	Reliability	Qualitative	b, c, d, f, g
Social	Public acceptance	Qualitative	b, c, d, f, g
	Odour impact	Qualitative	a, b, c, d
	Noise impact	Qualitative	b, c, d
	Visual impact	Qualitative	b, c, d
	Complexity	Qualitative	a, b, c, d, g
Economic	Capital cost (Capex)	£k	a, b, c, d, g
	Operational cost (Opex)	£k/yr.	a, b, c, d, g

a. (Muga and Mihelcic, 2008); b. (Ahmed *et al.*, 2017); c. (Molinos-Senante *et al.*, 2014); d. (Plakas *et al.*, 2016); e. (Mustapha *et al.*, 2018); f. (Kalbar *et al.*, 2016); g. (Ren and Liang, 2017); h. (Sabia *et al.*, 2016)

Pollutant removal potentials

Pollutant removal potentials of the wastewater treatment process are an important indicator used by all sewerage companies. Removal rates of pollutants directly affect the effluent quality and environmental compliance of the site. Water companies are legally required to comply with effluent standards at all STWs according to the European Commission Urban Waste Water Treatment Directive (European Commission, 1991). Particularly, the effluent standard includes Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Nitrogen (TN) and Total Phosphorus (TP). In the UK, the specific effluent standards vary across different catchments and this standard is agreed among Environment Agency and the water company. In recent years, the effluent standards for TN and TP are becoming more stringent as excessive nitrogen and phosphorus in the effluent can lead to eutrophication, especially in ecologically sensitive regions (Bunce *et al.*, 2018). Breach of environmental consent can lead to reputational and financial penalties. Therefore, achieving satisfactory effluent standards and environmental consents is a paramount objective of a STW. This has been highlighted in the findings of the previous case study (under the theme "*Compliance*"). Many studies reviewed in Table 12 have included pollutant removal pollutants when comparing different wastewater treatment processes or technologies (Muga and Mihelcic, 2008; Molinos-Senante *et al.*, 2014; Plakas *et al.*, 2016; Sabia *et al.*, 2016; Ahmed *et al.*, 2017; Mustapha *et al.*, 2018), despite the naming of this indicator slightly varied. Thus, this indicator was considered important and relevant and was included in the sustainability assessment.

Energy consumption

Many studies have considered energy consumption when evaluating different wastewater treatment systems (Muga and Mihelcic, 2008; Molinos-Senante *et al.*, 2014; Plakas *et al.*, 2016; Ahmed *et al.*, 2017; Mustapha *et al.*, 2018). In the treatment process, aeration and operating pumps are energy-intensive and contribute to the majority of operational energy footprint (Muga and Mihelcic, 2008). The reduction of energy consumption has been an increasingly important objective when designing a STW because it can potentially lead to a reduction in operational cost and associated carbon emission (if the power is sourced from fossil fuels). The importance of energy reduction has also been highlighted from the results of the case study (under the code "Energy and emission"). Therefore, energy consumption was also included in this research when assessing different wastewater treatment processes.

Greenhouse gases (GHG) emission

This indicator measures the emission of a range of greenhouse gases (e.g. carbon dioxide, methane, nitrous oxide) associated with the operation of the wastewater treatment process. Recent climate regulations and policies have accelerated the reporting of GHGs emissions in many industrial activities. UK government published the Climate Act in 2008 (HM Government, 2008), which set a target of 80% reduction of GHGs emissions by 2050 compared to the 1990 baseline level. This target has been stretched further to a "Net Zero" by 2050 to accelerate the decarbonisation of the economy as part of the Industrial Strategy (HM Government, 2017). At the water company level, there is a new target of Net Zero Carbon by 2030 to be achieved (Thames Water Utilities Limited, 2020c). Therefore, it is relevant and important to include Greenhouse gases emission when evaluating different wastewater treatment processes. This was also mentioned in the previous case study under the code “Energy and emission” and the theme “Consider wider environmental impacts”.

Land required

Land requirement has also been included in many studies reviewed (Muga and Mihelcic, 2008; Molinos-Senante *et al.*, 2014; Kalbar *et al.*, 2016; Plakas *et al.*, 2016; Ahmed *et al.*, 2017; Ren and Liang, 2017). It is a critical factor to be considered in urban areas where land is more limited and expensive. As suggested in the previous case study (under code “Site constraints”), land can be an investment constraint on upgrading assets and it is compounded by other drivers such as meeting higher treatment capacity and quality. Therefore, this indicator directly affects the potential feasibility of wastewater treatment processes. However, this was not included as a separate assessment indicator in this research because its relevance depends on the cases of specific sites. It was decided to use this indicator as a screening factor to eliminate any unfeasible treatment options prior to the sustainability assessment. For sites that have no land shortages, this indicator can be used as a denominator for normalising the performance of other indicators (e.g., carbon emission per area as CO_2eq/m^2).

Resource recover potential

Resource recovery potential refers to how much energy and nutrient can be potentially generated or recovered from the sludge produced in the wastewater treatment process (Molinos-Senante *et al.*, 2016). Power generation has been a key sludge strategy of the water company to offset the energy use for operating the treatment process and achieve the Net Zero Carbon, as suggested in the finding of the case study (under code “Resource recovery”). Current technologies in the water company include a pre-digestion treatment called Thermal Hydrolysis Process (THP) to improve the biogas yield and Anaerobic Digestion (AD) (Mills *et al.*, 2014). Then the biogas produced from AD is used to generate power at a Combined Heat and Power (CHP) unit. A greater power generation from sludge can reduce the dependency of the water company on power imported from the grid and reduce associated carbon emissions. Therefore, this indicator was considered relevant and important to this research.

Reliability

Reliability is a popular indicator with various definitions. Balkema *et al.*, (2002) define reliability as "*the sensitivity of the system to malfunctioning of equipment and instrumentation*". According to Kalbar *et al.* (2016), the reliability of the wastewater system refers to "*the possibility of achieving adequate performance for a specific period of time under specific conditions*". Similarly, Molinos-Senante *et al.* (2014) define reliability as "*the probability of mechanical failures and the impact of failures upon effluent quality*". These definitions place a strong emphasis on whether the treatment process or technology can achieve the effluent standard and how likely it would fail. Therefore, reliability should be to be considered when selecting a wastewater treatment option because it may affect the compliance of effluent qualities.

6.1.1.2 Social indicators

In contrast, social indicators are used less often in previous studies. This is because they can be difficult to quantify and the selection of them vary considerably across different contexts of studies (Balkema et al., 2002; Muga and Mihelcic, 2008; Popovic et al., 2013; Ahmed et al., 2017). For social indicators that are not directly quantifiable, Molinos-Senante et al. (2014) and Popovic et al. (2013) have used a scale to convert qualitative information into quantitative data. Common social indicators include public acceptance and nuisance including noise, odour and visual impact.

Public acceptance

Public acceptance or perception has been included in some reviewed studies (Molinos-Senante *et al.*, 2014; Kalbar *et al.*, 2016; Plakas *et al.*, 2016; Ahmed *et al.*, 2017; Ren and Liang, 2017). Balkema *et al.*, (2002) provide a brief definition of acceptance as "*the social-cultural perception of new sanitation concept and associated behavioural change*". However, this indicator is more relevant in projects to implement a new sanitation facility in rural areas or decentralised treatment plants where customers are more likely to be impacted. In the context of this research, most of the STWs owned by the water company are centralised and distant from residential areas so customers have little interaction with treatment facilities. In addition, customers have limited technical and operational knowledge of wastewater treatment technologies. It was initially considered not useful to be included in the indicators because the assessment tool was built for internal decision-making at the site level. However, this does not entail public acceptance is not relevant and important. It has been included in other larger strategic planning programme of the company such as PR19 and the Water Resources Management Plan where different investment options were consulted with the public.

Nuisance (Odour, noise, visual impact)

Nuisance in the context of social impacts generally refers to odour, noise, traffic and visual impact (Bradley *et al.*, 2002). Odour can arise from the operation of a STW due to the anaerobic decomposition of organic compounds that contain nitrogen and sulphur. The major contributors of odour in a STW are hydrogen sulphide, volatile fatty acids (VFAs) and other fermentation products during the storage of sludge and sludge liquor (Karageorgos *et al.*, 2010). Excessive levels of odour can be aesthetically unpleasant and cause public complaints. However, odour can be managed and mitigated by investing in additional odour control units and improving digestion quality. Nevertheless, it is important to understand the potential odour level of different wastewater treatment processes and the impact on the public, and therefore, odour was selected as a social indicator in the sustainability assessment.

Comparatively, noise and visual impacts are not as problematic as odour issues. Most STWs managed by the water company are built away from towns and populated neighbourhoods. Wastewater facilities are usually located in an enclosed operated site with almost no access for the public, and the potential noise and visual impact is smaller compared to odour treatment. This trend is also reflected in studies by Molinos-Senante *et al.*, (2014) and Plakas *et al.* (2016) where the importance weights assigned for noise and visual impact were much lower than the weights for odour. Therefore, noise and visual impacts were excluded from this research.

Complexity/Operability

Many studies have included complexity as an assessment indicator but its definition varies. Molinos-Senante *et al.*, (2014) refer to it as the complexity of the construction and operation of the wastewater treatment process, whereas Ren and Liang (2017) define it as the "*operability*" or "*simplicity*" of the treatment process. Srdjevic *et al.* (2012) translate the complexity of the process into the level of education (i.e. training) required for operators. This was also highlighted in the finding of the case study (under the code “Operating skills). Greater complexity to operate implies extra training, costs and operational risks. Therefore, this should be considered when assessing different wastewater treatment options. And this research defined operability as an indicator to reflect the technical complexity and the ease to operate the treatment process or technology.

6.1.1.3 Economic indicators

The major economic/financial indicators used in most review studies are capital expenditure (Capex) and operational expenditure (Opex) (Muga and Mihelcic, 2008; Molinos-Senante *et al.*, 2014; Plakas *et al.*, 2016; Ahmed *et al.*, 2017; Ren and Liang, 2017). Capex refers to the capital expenditures such as the construction of wastewater assets and obtaining land permits whereas Opex includes the costs associated with resources (staff, energy, chemicals and other contracted services) used for operation and maintenance of the facilities. As in the last regulatory period (AMP 6), water companies were strongly advised to use a ‘Totex’ (total expenditure) approach which both Capex and Opex should be considered together to evaluate the financial viability of the asset investment (Thames Water Utilities Limited, 2019b). This is also supported by the findings of the previous case study that whole life cost is a prime financial consideration (the codes “Whole life cost”). However, from the results of interviews in chapter 4 (under the code “Decision balance”), the balance between Opex and Capex can change over time and vary across different projects. Therefore, it was considered more suitable to include both Opex and Capex as two separate indicators in the sustainability assessment rather than combining them into one indicator.

6.1.2 Selection of assessment indicators

Based on the discussion above, a list of assessment criteria and indicators was selected and organised into a criteria hierarchy. The development of a criteria hierarchy is a major step in MCDA to translate the overall objective into measurable units (Belton and Stewart, 2002; Communities and Local Government, 2009). Figure 27 shows the first version of the criteria structure proposed for this sustainability assessment tool. The first layer of the hierarchy is the overall decision objective, which is to assess the sustainability of different wastewater treatment processes and to identify the more preferred option. The second layer includes the criteria based on the Three-Pillar model of sustainability and the layer below include the assessment indicators with respect to each criterion. Criteria can be defined as the requirement or standards to select a sustainable wastewater treatment process and indicators are the specific measurements or assignments of value to reflect the fulfilment of assessment criteria (Pavlovskaia, 2014; Ling *et al.*, 2021). It is worth noting that the fourth pillar 'Resilience' was also added to the criteria structure to reflect the significance of the long-term operational compliance and flexibility in the water company (Ling *et al.*, 2021).

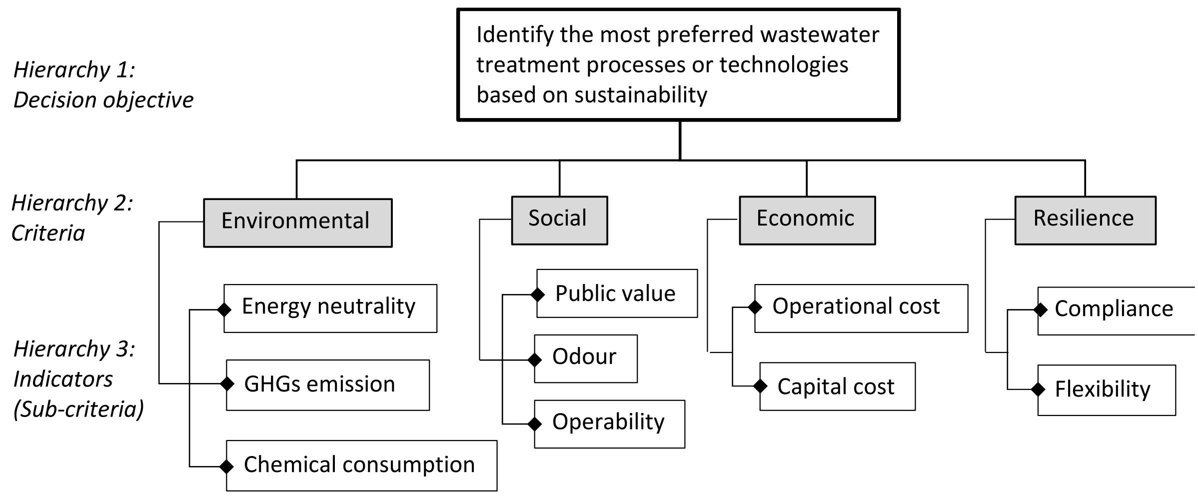


Figure 27. the criteria hierarchy proposed for the multi-criteria sustainability assessment tool in this research. (Ling *et al.*, 2021)

There were 10 indicators selected for the criteria hierarchy. Most of them were derived from the previous literature review and discussion. The definition of each indicator was revised and summarised in Table 13. Under the environmental criterion, Energy neutrality was added to the list to combine energy consumption and energy recovery into a 'net' indicator. Although Chemical consumption was not mentioned among reviewed studies in Table 12, it was highlighted by the results of interviews that chemical use in the wastewater treatment process is also an important factor (under the code "Chemical use"). A higher amount of chemical use is less desirable because of extra cost, embodied carbon (manufacturing and transport of chemicals) and excessive levels of metals in the process. Therefore, it was added to the criteria structure. Public value was also added in the social criterion to reflect the added social values to the local communities by implementing the wastewater treatment technologies. The inclusion of public value in decision-making was reflected by the latest corporate responsibility strategy of the water company (Thames Water Utilities Limited, 2021). Within the resilience criterion, Compliance was added to combine Pollutant removal potentials and Reliability of the treatment technologies. Flexibility was also added to this list to assess the readiness of treatment processes or technologies to adapt to future changes. Being flexible is particularly important for current assets to adapt to future changes such as climate change, population growth and evolving regulations (Balkema *et al.*, 2002; Ling *et al.*, 2021). The flexibility and adaptability of technologies were mentioned in the notes of personal communications where asset planners highlighted its importance when dealing with planning uncertainties. Overall, the proposed list of indicators and criteria hierarchy was not definitive as they will be revised as part of the continuous development and testing.

Table 13. The basic definitions of all indicators used for the assessment and the type of indication. Positive indication refers to the preference or desirability increases with the value of the indicator (e.g., operability) whereas negative indicators refer to those with diminishing desirability with higher value (e.g., costs). Adapted from (Ling *et al.*, 2021).

Indicators	Criteria	Definitions	Type of Value Indication
Energy neutrality	Environmental	Net energy consumption of the wastewater treatment process (Consumption minus recovery from sludge)	Negative
GHGs emission	Environmental	Total of direct and indirect carbon emission associated with the wastewater treatment process	Negative
Chemical consumption	Environmental	The total amount of chemical use in the operation of wastewater treatment (e.g., chemical dosing and polymer)	Negative
Public value	Social	Public value refers to the potential benefit provided to our community such as biodiversity gain or other amenities.	Positive
Odour	Social	The odour impact of the treatment process and sludge storage on the community	Negative
Operability	Social	The ease to operate the process, which is associated with the manpower resource as well as the level of skills and training required for operators.	Positive
Opex	Economic	Cost related to materials (consumables), staff cost (operators), power consumption, hired and contracted services (e.g., transport; service contract for specific treatment process)	Negative
Capex	Economic	Capital cost related to the construction and commissioning of the treatment process or technology.	Negative
Compliance	Resilience	The ability and the overall confidence of technology/process to meet the site compliance such as flow and quality consents and risks of failure.	Positive
Flexibility	Resilience	The ability of technology/process to adjust or upgrade to adapt to climate change, population growth and regulatory changes.	Positive

6.2 Weighting development using AHP

6.2.1 Individual weightings

As mentioned in chapter 5, AHP was selected for the MCDA model and it can be used for weighting calculation. The operation of AHP is based on three components: anatomy of the problem as a hierarchical structure, pairwise comparisons and calculation of criteria priorities (i.e., weightings) (Bottero *et al.*, 2011). Pairwise comparison is the primary task of AHP. The fundamental question to be asked is 'how important is criterion *a* compared to criterion *b*?' Each comparison determines the direction and degree of importance between two criteria or indicators using a semantic scale (see Table 10 in Chapter 5). For example, a scale number 3 refers to "*criterion a is moderately more important than criterion b*" whereas 1/3 refers to the opposite preference direction (criterion *b* is moderately more important than criterion *a*).

$$A = \begin{bmatrix} 1 & a_{1,2} & \dots & a_{1,m} \\ 1/a_{1,2} & 1 & \dots & a_{2,m} \\ \dots & \dots & 1 & \dots \\ 1/a_{1,m} & 1/a_{2,m} & \dots & 1 \end{bmatrix} \quad \text{Equation 4}$$

Each pairwise comparison generates a ratio number that feeds into a reciprocal matrix. A reciprocal matrix A of $m \times m$ is constructed based on m number of criteria (or indicators) to be compared at each hierarchical level (Equation 4). $a_{1,m}$ indicates the judgement scale between the first criterion and the m-th criterion, etc. In total, a number of $\frac{1}{2}m(m-1)$ comparisons are required per matrix given the property of reciprocity in AHP. A total of 5 matrices and 14 pairwise comparisons were required to calculate weights based on the criteria hierarchy as shown in Figure 27, including:

- One matrix with 6 pairwise comparisons between 4 top criteria (environmental, social, economic and resilience);
- One matrix for the indicators of the environmental criterion with 3 pairwise comparisons (between Energy neutrality, GHGs emission, and Chemical consumption);
- One matrix for the indicators of the social criterion with 3 pairwise comparisons between (Public value, Odour and Operability);
- One matrix for the indicators of the economic criterion with 1 pairwise comparison (between Capex and Opex)

- One matrix for the indicators of resilience criterion with 1 pairwise comparison (between Compliance and Flexibility)

Once the judgements on pairwise comparisons were collected, the weights can be acquired by either calculating the eigenvectors of the matrix or using the geometric mean method, which usually produce similar weight values (Ivanco *et al.*, 2017). The geometric mean method was used for this study due to its simplicity and compatibility with Microsoft Excel®. First, the geometric mean ($\bar{A}_r$) in r-th row of a $m \times m$ matrix was calculated as Equation 5,

$$\bar{A}_r = \sqrt[m]{\left(\prod_{i=1}^{m} a_{r,m}\right)} = \sqrt[m]{a_{r,1}\, a_{r,2} \dots a_{r,m}} \qquad \text{Equation 5}$$

and then $\bar{A}_r$ was normalised by the sum of geometric means of all rows to derive the weight of r-th item so that all weights sum up to 1, as shown in Equation 6 below

$$w_r = \frac{\bar{A}_r}{\sum_{i=1}^{m} \bar{A}} \qquad \text{Equation 6}$$

By Using these operations, weights were assigned to both indicators (Sub-criteria level) and criteria (top-criteria level) according to the criteria hierarchy in Figure 27. The final indicator weights were calculated by multiplying the corresponding criterion weight as a rescaling factor. For example, the weight of Energy neutrality was derived by multiplying its weights from the AHP matrix and the weight of environmental criterion as its global criterion.

A consistency check was also conducted to determine the consistency of pairwise comparisons from participants. In AHP, a Consistency Ratio (C.R.) is calculated between the Consistency Index (C.I.) and the average Random Consistency Index (R.I.) for each pairwise comparison matrix:

$$\text{C. R.} = \frac{\text{C. I.}}{\text{R. I.}} \qquad \text{Equation 7}$$

The Consistency Index (C.I.) is first determined by:

$$\text{C. I.} = \frac{(\lambda_{max} - n)}{(n - 1)} \qquad \text{Equation 8}$$

whereas λ_{max} is the principal eigenvalue of the matrix and n is the number of items compared in the reciprocal matrix. For a perfectly consistent matrix, C.I. should be 0 as $\lambda_{max} = n$ (Saaty, 1987). The Random Consistency Index (R.I.) can be determined by referring to R.I. derived by Saaty in Table 14. A value of 0.1 for C.R. was used as a guideline for inconsistency threshold (Saaty, 1987). Responses with C.R. much greater than 0.1 were considered inconsistent and were excluded from the further aggregation into group weightings.

Table 14. The random Consistency Indices (R.I.) used for consistency calculation in AHP. There are no values of R.I. when n=1 and n=2 because of the condition of pairwise comparisons and reciprocity

n	1	2	3	4	5	6	7	8	9	10
R.I.	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

6.2.2 Group weightings

To develop a weighting profile that involves more than one decision-maker, there are three approaches to aggregate individual weightings, namely “*sharing*”, “*comparing*” and “*aggregating*” (Belton and Pictet, 1997).

1) *Sharing*: this refers to reaching a consensus by understanding and reducing differences in preference through discussion and negotiation. Eventually, only one set of pairwise comparisons input will be used for calculating weights.
2) *Comparing*: this retains the individual inputs of pairwise comparisons but their results of weights are compared and negotiated among decision-makers to select the most representative weighting profile.
3) *Aggregating*: this allows individual decision-makers to provide their own inputs of pairwise comparisons and a representative value will be derived using mathematical aggregation. This reduces the differences among decision-makers without explicitly addressing the underlying reasons for differences in their preferences.

Due to the prevailing COVID-19 social distancing conditions when this research was conducted, mathematical aggregation was considered the most feasible approach and therefore selected for developing the group weighting profile (Ling *et al.*, 2021). Two mathematical aggregation approaches are available: Aggregation of Individual Judgements[18] (AIJ) and Aggregation of Individual Priorities[19] (AIP) (Basak and Saaty, 1993). AIJ is recommended for groups of individuals who are willing to act and behave as one collective 'individual' whereas AIP is more appropriate for individuals with different underlying interests and value systems (Forman and Peniwati, 1998; Ossadnik *et al.*, 2016). It can be argued that for this study AIJ would be a better choice by assuming that all participants are willing to act as an 'individual' because they collectively represent the interest of the water company. However, AIP is also reasonable because each participant may hold personal preference, which is strongly influenced by the role in the company. Additionally, it was not possible for participants to share their inputs and results in person (such as in a focus group) due to the social distancing condition. Although online forum could be used as a platform to share results, it would be difficult to monitor and manage the communications between different participations, in addition to the technical challenge of designing such a forum. Considering this dilemma, both aggregation approaches were included and compared in this study. For the AIJ approach, aggregation by GM (Geometric Mean) should be used (Basak and Saaty, 1993) whereas both GM and AM (Arithmetic Mean) can be used in the AIP method (Forman and Peniwati, 1998; Ossadnik *et al.*, 2016). As a result of this combination, there were three specific aggregation methods used:

1. AIJ-GM: Aggregation of Individual Judgements using the geometric mean

$$\bar{a}_k^g = \sqrt[q]{\left(\prod_{i=1}^{q} a_k\right)} = \sqrt[q]{a_k^1\, a_k^2 \ldots a_k^q} \qquad \text{Equation 9}$$

   whereas $\bar{a}_k^g$ is the group judgement with respect to the k-th criterion or indicator aggregated from individual judgements $(a_k^1, a_k^2, \ldots, a_k^q)$ based on q number of decision-makers.

2. AIP-GM: Aggregation of Individual Priorities using the geometric mean

[18] Judgements here refer to input of pairwise comparisons, which feed into the AHP reciprocal matrix, such as $a_{1,m}$ in Equation 4.

[19] Priorities here refer to the weights derived from the AHP reciprocal matrix.

$$\overline{w}_k^g = \sqrt[q]{\left(\prod_{i=1}^{q} w_k\right)} = \sqrt[q]{w_k^1\, w_k^2 \dots w_k^q} \qquad \text{Equation 10}$$

whereas $\overline{w}_k^g$ is the group weight of k-th criterion or indicator aggregated from individual weights $(w_k^1, w_k^2, \dots , w_k^q)$ based on q number of decision-makers.

3. AIP-AM: Aggregation of Individual Priorities using the arithmetic mean

$$\overline{w}_k^g = \left(\sum_{i=1}^{q} w_k\right)/q = \frac{(w_k^1 + w_k^2 + \cdots + w_k^q)}{q} \qquad \text{Equation 11}$$

Equal weights between participants were applied when aggreging individual weightings. This was based on the assumption of equal importance between individuals (Forman and Peniwati, 1998). The group weights of indicators were then normalised (similar to Equation 6) so that all weights sum up to 1.

6.2.3 Collection of AHP responses

Although focus groups were originally preferred to collect responses for AHP, this was not possible due to COVID-19 social distancing restrictions. Therefore, the remote collection method was used as an alternative. For the collection of responses of pairwise comparisons, stakeholders in the Department of Asset, Planning and Investment[20] were invited via an online questionnaire to perform pairwise comparisons between criteria and indicators. The selection of participants was based on the stakeholder analysis conducted previously for the interview recruitment (Figure 14 in Chapter 4). Another criterion used for selecting participants was that participants should have previous experience working with wastewater asset planning and management. Microsoft Forms® was used as the platform for conducting questionnaires due to its user-friendly features, flexibility and compatibility.

The invitations for questionnaires were distributed to a pre-defined list of contacts by email with the attached link to the questionnaire. In the invitation email, participants were also invited to share this questionnaire with wider contacts who within the key stakeholder groups as part of the snowball sampling. Once each questionnaire was completed, the responses

[20] The name of the department has been changed from Strategic Planning and Investment (used in chapter 4) due to organisational re-structuring. The function of the department remains the same.

were recorded by Microsoft Forms® and a notification was forwarded to the data owner (i.e., the researcher).

6.2.4 Design of questionnaires

In terms of the design of the questionnaire, the ‘rating’ question type was selected as it provides the most compatible functionality with the AHP judgement scale. An example of the pairwise question in the questionnaire is shown in Figure 28. However, one drawback of using the built-in rating question style is that it only allows a maximum number of 9 scale points per question which does not fulfil the AHP judgement full scale (i.e., 8+1+8=17 points for both directions of preference). There were potentially two ways to implement the AHP scale to the rating scale in Microsoft Forms®. First, a reduced AHP judgement scale of 5 points (reduced from 9) with a reduced granularity can be used. Alternatively, two separate questions can be used per pairwise comparison to first ask the direction of the preference and then to what extent on a full AHP scale of 9 points. After trialling both scale designs, it was considered easier and quicker to complete pairwise comparisons using a reduced scale. There has been literature supporting the use of a reduced scale. Previous studies have shown that the C.R. threshold might be too restrictive when using the full AHP scale (Lane and Verdini, 1989; Murphy, 1993) and C.R. threshold also depends on the granularity of the scale used (Salo and Hämäläinen, 1997). Additionally, Pauer *et al.* (2016) suggest that a smaller scale could lead to more consistent responses.

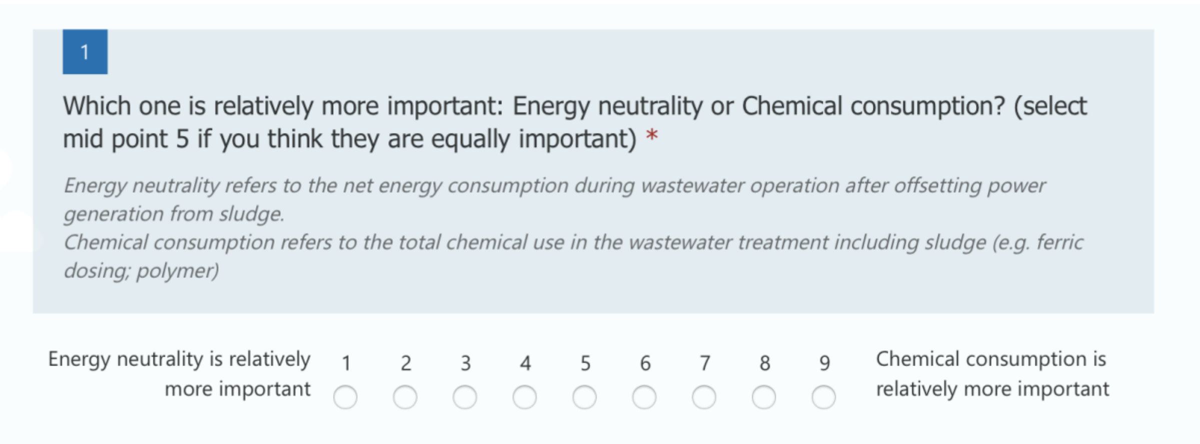


Figure 28. An example of the rating question to make a pairwise comparison in questionnaire on Microsoft Forms®.

Another observation during the pilot trial of AHP was that the extreme scale values (1 and 9) were rarely selected. Given this consideration, a revised AHP scale of 1 to 5 (i.e., 1,2,3,4,5) per preference direction was attributed to the scale. Alternatively, a 5-point scale consisting of 1,3,5,7,9 could also be allocated. However, this tended to produce a much higher inconsistency ratio (C.R.) in the trial. Therefore, the 5-point scale from 1 to 5 was implemented in the design of each question as shown in Figure 29. As such, the 9 scale points were divided into 5 points for each direction of preference (with a shared scale point for equal importance).

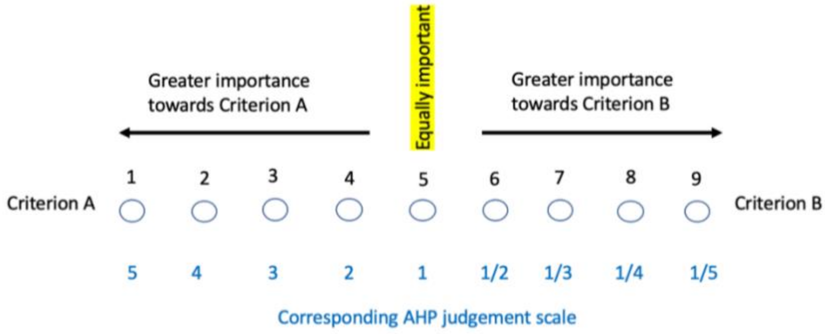


Figure 29. A reduced AHP scale was implemented to the rating question in the questionnaire.

In total, there were 14 questions of pairwise comparisons (based on 5 matrices) in each questionnaire. The comparisons were made first at the lower level (i.e., between indicators in each criterion) of the hierarchy and then at the top level of the hierarchy (i.e., between criteria) based on the hierarchy structure in Figure 27. The set and the order of questions were the same for all participants. This order was used because feedback from the trial run suggested that it was more difficult to start comparisons at the top criteria level because their definitions were broad. In addition to the 14 pairwise comparisons, an extra question was included to ask the participant to directly rank all indicators and criteria by their importance. This ranking was used to cross-check the results of weighting derived from AHP. Specifically, Spearman's Rank Correlation coefficients were calculated in SPSS® to examine their similarities.

The responses of all questionnaires were then processed for the calculation of weights. A calculation spreadsheet was built in Microsoft Excel® (Figure 30) to streamline the computation process by embedding equations mentioned in section 6.2.1 (before group aggregation). The responses in each questionnaire were extracted and converted to the AHP scale numbers. By putting the scale numbers in the spreadsheet, the indicator and criteria weights and consistency ratios are automatically calculated. The weightings derived from each participant (i.e., individual weightings) were aggregated into the group weightings using the methods mentioned in section 6.2.2.

2. Input sheet			Comparisons at sub-criteria level		which one is more	To what extent?		
Criteria	Sub-criteria level	Global Weights	Criterion A	Criterion B	A or B?	Select a scale	AHP input	Consistency Raio (CR)
Environmental	Energy neutrality	0.125	Energy neutrality	Chemical consumption	A	2	2	-3.828E-16
	Chemical consumption	0.063	Energy neutrality	Total emission	A	2	2	
	Total emission	0.063	Chemical consumption	Total emission	A = B	1	1	
Social	Public value	0.023	Public value	Odour	A	3	3	0.0633737 3
	Odour	0.010	Public value	Operability	B	3	1/3	
	Operability	0.053	Odour	Operability	B	4	1/4	
Financial	Opex	0.098	Opex	Capex	A = B	1	1	n/a
	Capex	0.098						
Resilience	Flexibility	0.117	Flexibility	Compliance	B	3	1/3	n/a
	Compliance	0.350						
	Criteria		Comparisons at the criteria level					
			Criterion A	Criterion B	A or B?	Select a scale	AHP input	Consistency Raio (CR)
	Environmental	0.2508122	Environmental	Social	A	3	3	0.0589219 6
	Social	0.0861025	Environmental	Financial	A	2	2	
	Financial	0.1962712	Environmental	Resilience	B	3	1/3	
	Resilience	0.4668141	Social	Financial	B	3	1/3	
			Social	Resilience	B	4	1/4	
			Financial	Resilience	B	2	1/2	

Figure 30. The spreadsheet interface for weight calculations using the an example input from questionnaires. Input sections are highlighted in yellow cells. Inputs are converted to the AHP scale numbers (green cells).

6.2.5 Results and discussion

6.2.5.1 Individual weightings

In total, twelve questionnaires were completed and the responses were collected on Microsoft Forms®. However, three inputs of responses were inconsistent (C.R. much higher than 0.1) and they were excluded from further data analysis. The indicators and criteria weights derived from the other nine responses are shown in Table 15. The rankings of the weights derived by AHP are also shown in Table 16 with direct ranking given by each participant.

In terms of the ranking across all profiles, the results derived from AHP suggest that Compliance was often assigned the highest weights (6 out 9 participants). Most participants (7 out 9 participants) also indicated that Compliance is the most important indicator based on their direct ranking. This closely reflects the finding in the previous case study on the decision context where compliance was highlighted as a top priority. In contrast, indicators in the social criteria (Public value, Odour and Operability) tended to have the lowest weights derived from AHP, except for Input 8 where Energy neutrality and Chemical consumption had the lowest weights. A weight value of less than 0.1, and sometimes less than 0.05, was frequently attributed to these indicators in the social criterion (Table 15).

Table 15. Individual weightings of indicators and criteria derived from AHP questionnaires.

		Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8	Input 9
Indicators	Energy neutrality	0.179	0.049	0.065	0.089	0.077	0.057	0.128	0.028	0.051
	Chemical consumption	0.043	0.097	0.082	0.056	0.253	0.057	0.044	0.028	0.051
	Total emission	0.062	0.097	0.103	0.211	0.139	0.170	0.031	0.084	0.101
	Public value	0.018	0.085	0.028	0.076	0.204	0.023	0.009	0.029	0.066
	Odour	0.041	0.053	0.111	0.026	0.089	0.021	0.023	0.093	0.076
	Operability	0.047	0.034	0.111	0.109	0.039	0.098	0.057	0.037	0.029
	Opex	0.082	0.121	0.125	0.083	0.041	0.253	0.092	0.080	0.170
	Capex	0.082	0.121	0.125	0.028	0.041	0.084	0.275	0.080	0.170
	Flexibility	0.111	0.172	0.042	0.054	0.039	0.048	0.068	0.136	0.072
	Compliance	0.334	0.172	0.208	0.268	0.078	0.191	0.273	0.407	0.215
Criteria	Environmental	0.285	0.243	0.250	0.356	0.469	0.283	0.203	0.140	0.203
	Social	0.105	0.172	0.250	0.212	0.331	0.142	0.089	0.159	0.170
	Financial	0.164	0.243	0.250	0.110	0.083	0.337	0.367	0.159	0.341
	Resilience	0.446	0.343	0.250	0.322	0.117	0.238	0.341	0.542	0.286

Table 16. The ranking of indicators weights derived from AHP (**A**) and direct ranking (**D**) by each respondent. Top indicators (with the highest importance) are highlighted in green whereas the worst ones are highlighted in red. Spearman's Rank Coefficients are shown to indicate the similarity between the rankings (α = 0.05).

	Input 1		**Input 2**		**Input 3**		**Input 4**		**Input 5**		**Input 6**		**Input 7**		**Input 8**		**Input 9**	
	A	**D**	**A**	**D**	**A**	**D**	**A**	**D**	**A**	**D**	**A**	**D**	**A**	**D**	**A**	**D**	**A**	**D**
Energy neutrality	2	2	9	9	8	8	4	4	6	3	6	5	3	5	9	6	8	5
Chemical consumption	8	6	5	5	7	9	7	7	1	2	6	6	7	6	9	9	8	7
Total emission	6	7	5	4	6	10	2	3	3	4	3	4	8	9	4	4	4	6
Public value	10	10	7	7	10	6	6	2	2	5	9	8	10	10	8	10	7	10
Odour	9	9	8	8	4	3	10	9	4	10	10	10	9	7	3	5	5	8
Operability	7	8	10	10	4	2	3	5	10	9	4	7	6	2	7	2	10	4
Opex	4	4	3	2	2	5	5	6	7	7	1	1	4	4	5	8	2	3
Capex	4	3	3	1	2	4	9	8	7	6	5	3	1	3	5	7	2	2
Flexibility	3	5	1	6	9	7	8	10	9	8	8	9	5	8	2	3	6	9
Compliance	1	1	1	3	1	1	1	1	5	1	2	2	2	1	1	1	1	1
Spearman's Coefficient ρ	0.924		0.795		0.652		0.83		0.535		0.888		0.758		0.665		0.5	
Significance (p)	<0.001		0.006		0.041		0.003		0.111		0.001		0.011		0.036		0.141	

Based on the significance threshold of 0.05, 7 out of 9 ranking profiles had a significant Spearman's Rank correlation (i.e., $p < 0.05$) between the AHP ranking and direct ranking, indicating a strong similarity between them. This result suggests that AHP has the potential to reliably elicit decision-maker's preferences on the order of importance of these indicators, although there were also occasions where AHP produced rather different rankings of weights such as Input 5 and Input 9. Although it would be statistically more robust to collect more responses, the pool for selecting eligible participants was limited by the size of the department. However, on this occasion, the comparisons of individual weighting profiles indicated a high level of similarity between them and additional responses (especially from the same department) may not alter the overall weighting profile significantly.

6.2.5.2 Group weightings

The group weightings were derived using three different aggregation methods (mentioned in section 6.2.2.) and are shown in Table 17. Overall, the three weight aggregation methods produce very similar results of weights. There were small variations in the weight values but they did not differ considerably. The results of using geometric mean (AIJ-GM and AIP-GM) were more similar to each other than that of arithmetic mean (AIP-AM). The rankings between AJI-GM and AIP-GM were almost identical whereas the ranking of AIP AM was slightly different. Additionally, the weights of the 4 top criteria (environmental, social, financial and resilience) were identical between AJI-GM and AIP-GM.

In terms of the weights of individual indicators, Compliance was assigned the highest weight values (0.242, 0.254, 0.238 respectively) in all three aggregation methods, followed by Opex and Total emission (Figure 31). Public value was assigned the lowest weight in AIJ-GM and AIP-GM but Odour was the lowest one in AIP-AM. However, in the weighting profile of AIP-AM, the weight of Public value was only 0.001 higher than that of Odour and this difference is almost negligible. It is worth noting that the weightings developed based on pairwise comparisons only indicate the relative importance between all indicators selected in this study. Although Public Value was assigned with the lowest weight, it is still an important indicator but it provides less potential value if trade-offs with the performance of other indicators.

Table 17. Group weighting profiles of indicators and criteria using three combinations of aggregation and normalisation techniques. The items with the highest weight were highlighted in green whilst the lowest ones were highlighted in red.

	Methods / Indicator	AIJ-GM		AIP-GM		AIP-AM	
		Weights	Rank	Weights	Rank	Weights	Rank
Indicators	Energy neutrality	0.083	5	0.081	6	0.080	6
	Chemical consumption	0.076	7	0.074	7	0.079	7
	Total emission	0.117	3	0.113	3	0.111	4
	Public value	0.050	10	0.046	10	0.060	9
	Odour	0.063	9	0.058	9	0.059	10
	Operability	0.069	8	0.063	8	0.062	8
	Opex	0.117	2	0.120	2	0.116	2
	Capex	0.103	4	0.106	4	0.112	3
	Flexibility	0.080	6	0.084	5	0.082	5
	Compliance	0.242	1	0.254	1	0.238	1
Top criteria	Environmental	0.276	2	0.276	2	0.270	2
	Social	0.182	4	0.182	4	0.181	4
	Financial	0.220	3	0.220	3	0.228	3
	Resilience	0.321	1	0.321	1	0.321	1

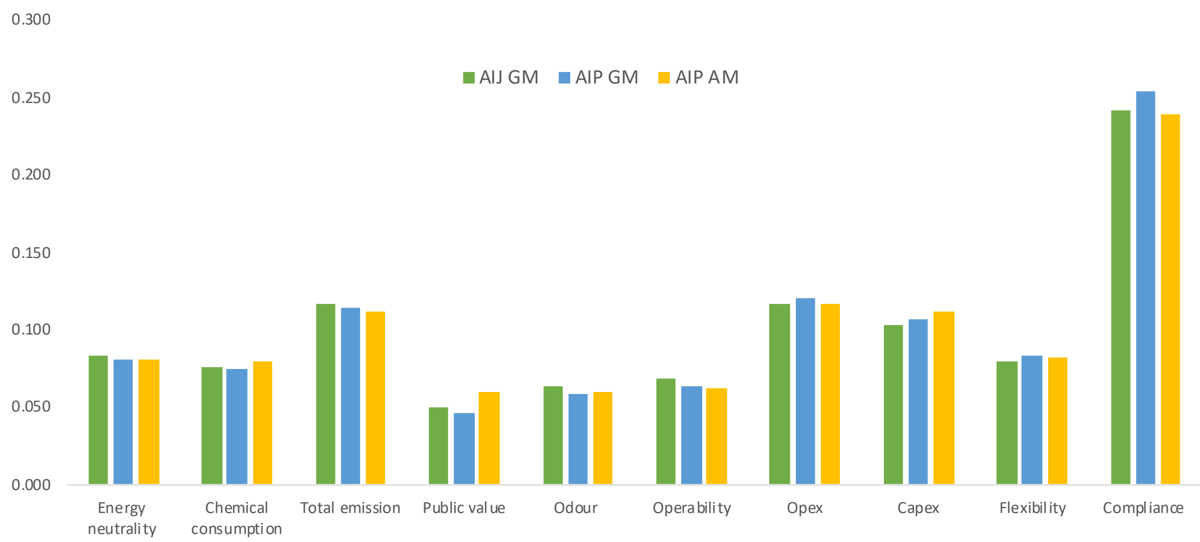


Figure 31. The group weighting profiles in bar charts derived from three aggregation methods.

Although three aggregation methods produced very similar results, the group weighting profile derived from AIP-GM was selected for the sustainability assessment. This is because: firstly, AIP requires the underlying condition that participants act as individuals with their own preferences rather than a single group. Due to the nature of online questionnaires, there was no opportunity for participants to collaborate and act as one ‘individual’. Secondly, despite very similar results, AIP was considered more feasible than AIJ as it requires fewer procedures and time to compute. Yap *et al.* (2019) also highlight that AIP is the best aggregation method because it provided the most accurate ranking. Additionally, AIP-AM was excluded as its results are more different from that of the two methods using GM. Overall, the decision on selecting a group weighting profile was relatively trivial as all aggregation methods produced very similar weights and rankings of importance.

The results of individual and group weightings were presented to each participant in informal meetings. The majority of participants indicated they agreed with the individual weighting results and group weighting results. In terms of the experience with AHP, most participants found pairwise comparisons “*straightforward*” and “*easy to use*”. Some participants found it “*interesting*” to review weightings of other participants with different perspectives and suggested it would be useful comparing weighting profiles across different departments in the water company. However, two participants were unsure of the results of pairwise comparisons and expressed difficulty of understanding and using AHP. AHP can be challenging for participants to understand due to the unfamiliar rating scale and the requirement for internal consistency when making comparisons (Schmidt *et al.*, 2016).

6.3 Pilot study 1: Application on selecting a secondary treatment process scheme for a new STWs

A sewage treatment works (denoted as STW A) was selected as a pilot study for applying the criteria hierarchy and group weightings developed from previous stages. STW A has a designed capacity of approximately 100,000 population equivalent and the following seven treatment technologies were considered (pre-determined by the company stakeholders) for the implementation of the treatment process, namely:

1) Activated Sludge Process (solo, as a baseline) (ASP)
2) Dissolved Air Flotation (DAF)
3) Chemically Assisted Primary Settlement (CAPS)
4) Ballasted Activated Sludge Process (B-ASP)
5) Sequencing Batch Reactor (SBR)
6) Granular Activated Sludge Process (G-ASP)

7) Mainstream De-ammonification process (De-ammo)

As those technologies or processes were to be implemented at various stages of wastewater treatment, basic process designs are provided in (Figure 32) to make the scope comparable. Technical and operational details were not included as they were not the focus of this study and also due to confidentiality reasons.

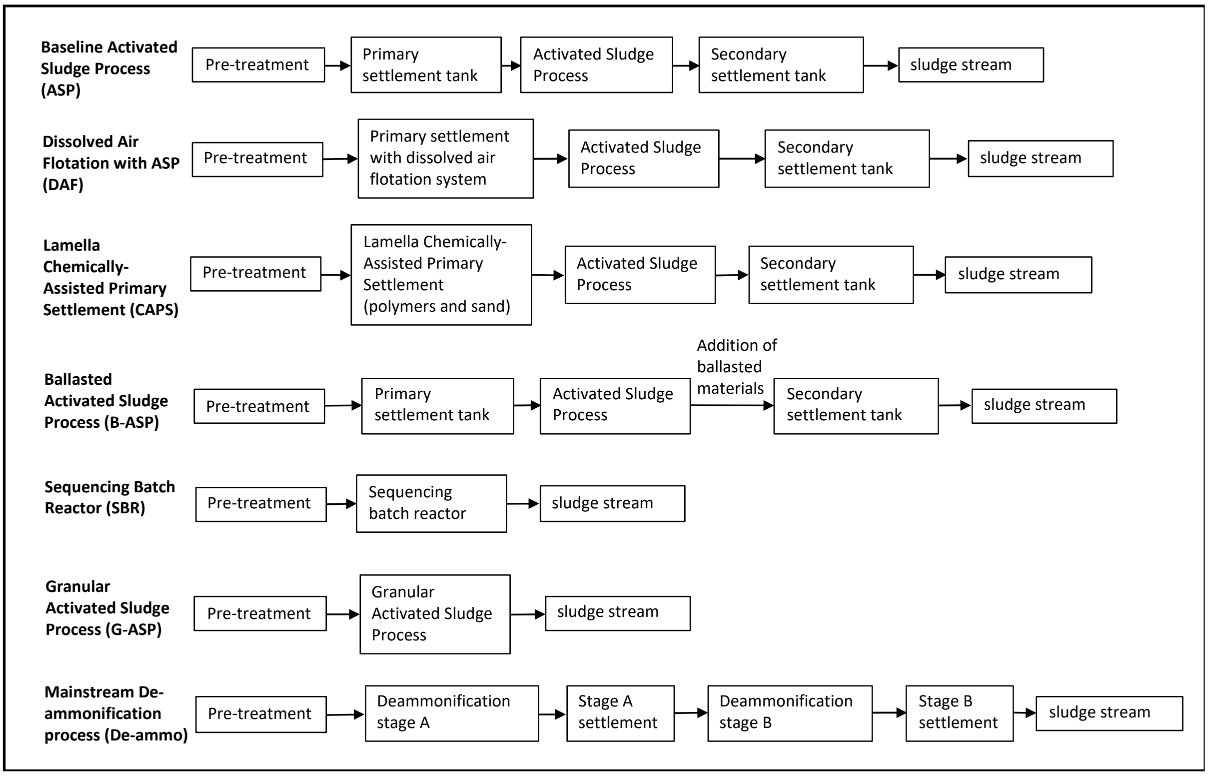


Figure 32. Basic process flow design of seven investment options for STW A. Adapted from Ling *et al.* (2021).

An optioneering study[21] on this case has been previously done by a group of stakeholders in the company and a decision was already made. The information in this business case was reused to test the MCDA model with indicators and weights developed. And the decision result derived from this study was then compared to that concluded in the previous optioneering study. The objectives of this pilot study were to discuss and provide insights into the reliability of the MCDA model for identifying the most preferred solution and its feasibility to perform sustainability assessments.

[21] In the previous stakeholder group meeting, performance ratings from 1 to 5 were provided by stakeholders for each wastewater treatment technology with respect to 15 performance indicators developed by the group. Score of each alternative was aggregated by adding rating scores of all indicators followed by a heuristic process to rank them and select the best option.

6.3.1 Composite score and option ranking

The next stage was to collect the performance data of the indicators. Given that this business case was a desktop study for estimation purposes only, the performance data of quantitative indicators (except for Opex and Capex) were not available. Attempts were made to recruit the same group of stakeholders involved in STW A optioneering study to collect ratings based on expert knowledge. However, some contacts were no longer available in the organisation. Instead, the performance ratings on a 5-point scale provided in the previous optioneering study report were reused as secondary data. Opex and Capex of each option were already estimated quantitatively by stakeholders using a business costing estimation model and they were converted into the same performance rating scale using a linear transformation. The average performance ratings of indicators are summarised in Table 18.

Table 18. The average performance ratings of indicators provided from the previous optioneering study of STW A. The lowest rating '1' refers to the poorest performance of that indicator whereas '5' refers to the best. It was assumed that the rating scale was of an interval scale.

Indicators	ASP	DAF	CAPS	B-ASP	SBR	G-ASP	De-Ammo
Energy neutrality	3.00	2.67	2.67	2.67	2.00	4.00	3.40
Chemical consumption	3.33	2.67	1.83	1.67	3.50	3.67	3.80
Total emission	2.00	2.80	2.80	2.00	2.20	3.80	2.60
Odour	2.67	3.17	3.33	2.50	3.17	3.50	2.40
Operability	5.00	2.83	2.67	2.83	3.33	2.83	2.80
Capex	4.41	5.00	3.17	3.58	4.30	1.00	3.58
Opex	3.70	2.42	1.00	2.39	2.76	5.00	2.39
Flexibility	1.43	3.00	2.86	2.57	2.29	3.71	2.83
Compliance	4.33	3.00	2.83	3.17	2.17	2.67	3.00

AHP was considered less practical for comparing performances between options because it would require 21 pairwise comparisons between the 7 options with respect to each indicator, and hence, each decision-maker would need to perform a total of 210 pairwise comparisons (with respect to a total of 10 indicators). That would place a huge demand on decision-maker's information process capacities and time requirements. Instead, a linear additive model was used as an alternative to synthesise performance ratings of indicators $(v_1, v_2, \dots, v_n)$ and their corresponding weights $(w_1, w_2, \dots, w_n)$ into a composite score s_i for the i-th option, denoted as

$$s_i = \sum_{i=1}^{n} w_n v_n = w_1 v_1 + w_2 v_2 + \cdots + w_n v_n \quad \text{Equation 12}$$

Since the performance ratings of all indicators were homogenous (from 1 to 5), the normalisation of the performances was not needed. The group weighting (AIP-GM) developed in the previous section was applied to the weighted additive model. In this pilot study, Public value was excluded due to a lack of data on its performance rating. As such, the indicator weights were re-normalised and shown in (Table 19). The exclusion of the weight of Public value should not considerably affect the composite scores as the weight value was very low (0.046 as shown in Table 17).

Table 19. Re-normalised indicator weights (after Public value was excluded)

Indicators	Weights	Rank
Energy neutrality	0.085	6
Chemical consumption	0.078	7
Total emission	0.119	3
Odour	0.061	9
Operability	0.067	8
Capex	0.111	4
Opex	0.126	2
Flexibility	0.088	5
Compliance	0.266	1

6.3.2 Sensitivity analysis

Sensitivity analysis should be conducted in AHP to check the sensitivity of option ranking. This study included two parts of sensitivity analysis. Firstly, the ranking of options was compared using the aggregated group weightings and individual weightings to examine the consistency of option rankings. Additionally, composite scores without applying any weighting were also included for comparison. The second part was to identify the most critical indicator by calculating the minimum changes in indicator weights that lead to rank reversals. As the aim of this case study is to identify the best option, only the rank reversal between the top two options were concerned. The steps for identifying the most critical indicator were based on the theorems developed by Triantaphyllou and Sánchez, (1997). If the i-th option is the best and the j-th option is the second-best by their composite scores ($S_i > S_j$), then the minimum change $\delta_{k,i,j}$ in the weight of indicator C_k (with respect to k-th indicator) to cause rank reversal between top two options can be determined, as follows:

Given that $\delta_{k,i,j}$ is defined as: $\delta_{k,i,j} = w_k - w_k^*$, whereas w_k^* is the new weight of k-th indicator that cause a rank change,

if the performance of the j-th option is better than the i-th option with respect to the k-th indicator (i.e. $v_{jk} > v_{ik}$), then

$$\delta_{k,i,j} < \frac{(S_j - S_i)}{(v_{jk} - v_{ik})} \quad \text{Equation 13}$$

if the performance i-th option is better than the j-th option with respect to the k-th indicator (i.e. $v_{ik} > v_{jk}$), then

$$\delta_{k,i,j} > \frac{(S_j - S_i)}{(v_{jk} - v_{ik})} \quad \text{Equation 14}$$

To calculate the minimum change $\delta_{k,i,j}$ for rank reversal, both equations are required to meet the condition of $\frac{(S_j - S_i)}{(v_{jk} - v_{ik})} \leq w_k$, whereas w_k is the weight of k-th indicator. Additionally, the minimum change $\delta_{k,i,j}$ can also be expressed in the relative term as:

$$\delta'_{k,i,j} = \left(\frac{\delta_{k,i,j}}{w_k}\right) \times 100 \quad \text{Equation 15}$$

6.3.3 Results[22]

6.3.3.1 Composite scores and option ranking

A composite score was calculated for each wastewater treatment option by aggregating performance ratings of indicators and weights. Figure 33 shows the composites score of these options and score each sustainability criteria. Composite scores were then used to rank options and identify the options with the highest overall desirability. The ASP option was scored as the best option based on its composite score (3.48) followed by G-ASP (3.25) and DAF (3.07). Specifically, ASP has the highest scores for the criteria of Social impact[23](0.49), Economic viability (0.96) and Resilience (1.28). This was reflected by high performance ratings on the indicator of Operability and Compliance for ASP. ASP was considered the least risky

[22] This section is based on the results in my previous publication (Ling *et al.*, 2021).

[23] The 4 top criteria were renamed as “Environmental impact”, “Social impact”, “Economic viability” and “Resilience” in the publication (Ling *et al.*, 2021).

option due to the vast experience the water company had in designing and operating this type of process. It appears that these factors were the substantial drivers of the greater desirability of ASP compared to other options. In comparison, G-ASP was the second-best option. G-ASP scored the best in the Environmental impact criteria because it is designed to treat wastewater with a very low operational footprint. Although it scored very high on Energy neutrality and Opex, the overall score of G-ASP was compromised by a low score on Capex and a mediocre score on Compliance.

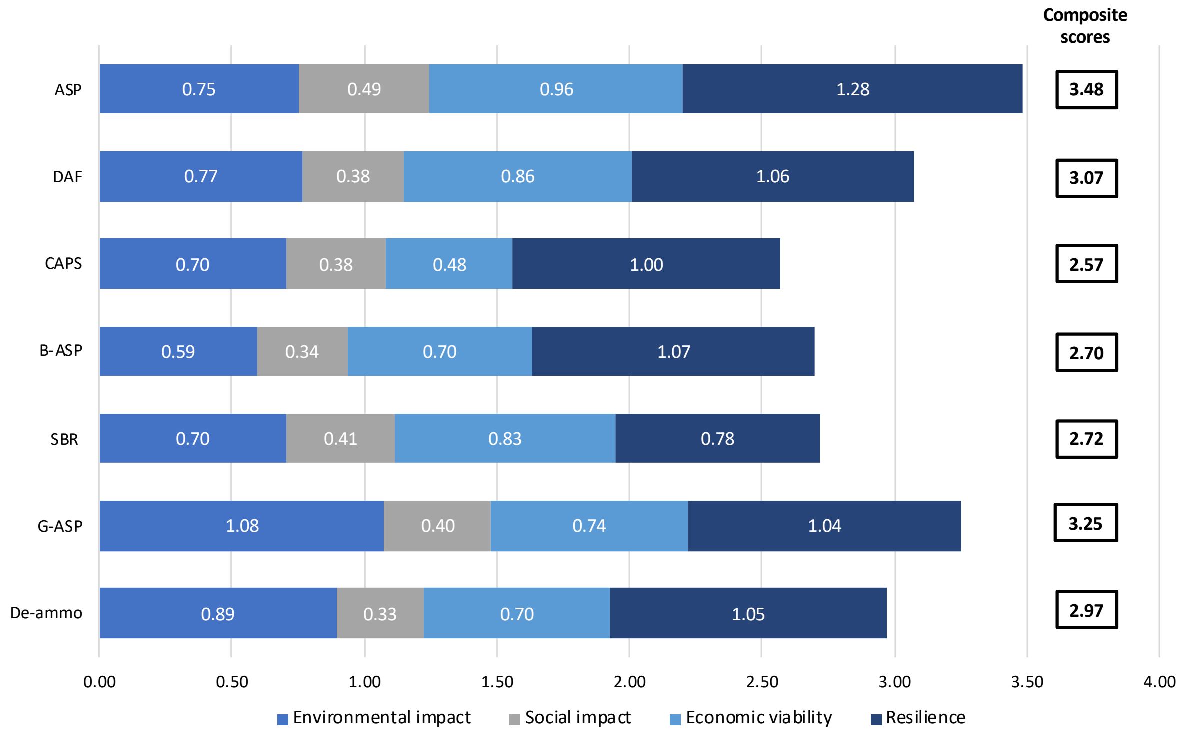


Figure 33. Composite scores of all wastewater treatment options and scores in individual sustainability criteria. (Ling *et al.*, 2021)

The ranking based on the composite score was compared to the result of the previous optioneering study. The result of Spearman's rank correlation showed a significant positive correlation (ρ=0.75, p=0.052) between two sets of ranking (Table 20), indicating a strong similarity. In addition, the ranks of the top three options are identical. The largest discrepancy is the rank of De-ammo. In the previous study, De-ammo was perceived as a competitive alternative with great energy savings and a low footprint. However, the option was given a much lower rank in the previous study due to the concern over an external supply chain risk. Overall, the similarity of the ranking suggests the assessment methods used in the pilot study is capable of providing consistent decision support information. This was expected because the same performance indicators (but different aggregation methods) were used. However, the additional benefits of using this assessment approach are to providing more granular information (i.e. performance in specific pillars) and a consistent approach for decision-makers to rank and select alternatives repeatedly, rather than a heuristic and arbitrary approach. The benefits of this assessment approach were further validated in the second pilot study in Chapter 7 and discussed in usability testing in Chapter 8.

Table 20. The ranking of wastewater treatment options for STW A derived by the MCDA approach compared to the ranking from the previous optioneering study as a reference (Ling *et al.*, 2021).

	ASP	DAF	CAPS	B-ASP	SBR	G-ASP	De-ammo	Spearman's coefficient
Ranking based on composite score	1	3	7	6	5	2	4	ρ=0.75 (p=0.052)
Ranking (previous optioneering study)	1	3	5	6	4	2	7	

6.3.3.2 Sensitivity analysis

By re-applying different weightings profiles of participants developed from online questionnaires in Table 15, the option rankings derived from individual weightings are shown in Table 21. Three out of nine weighting profiles (Input 2,5 and 6) had a rank reversal between the best option (ASP) and the second-best option (G-ASP) compared to the group weighting reference (at the top row). Additionally, rank reversal between the top two options also occurred in the no weighting scenario (second row). The results suggested that the rankings of options were generally consistent but the ranks of the top two options may be susceptible to rank reversals.

The other part of the sensitivity analysis was to calculate the minimum change in the group weight that can cause a rank reversal between the top two options. Table 22 shows that the minimum weight change δ in both absolute and relative terms according to Equation 13-15. In the absolute term, Capex was the most critical indicator, with the smallest value δ of 0.066. By the definition of Equation 14, if the weight value of Capex (0.111) is decreased by at least 0.066, the rank between the best option and second-best option would reverse. In the relative term, Compliance was the most sensitive indicator as a 53% change in the weight value would cause the rank reversal. However, given that the original weight of Compliance has the largest weight (0.266) out of all indicators, a 53% change in its original value is, comparatively, not as sensitive in the absolute term. This suggests that Capex is still the most critical indicator in terms of the absolute value δ and its weight allocation should be revised in future studies. Overall, most indicators can withstand a value change in their weights without causing rank reversals.

Table 21. The comparison of option rankings between different individual weighting profiles developed in Table 15 using the composite scores. Ranking of composite scores without applying any weighting was also included. The best option in each profile was highlighted in green and the worst was in red.

	ASP	DAF	CAPS	B-ASP	SBR	G-ASP	De-ammo
Group weightings	1	3	7	6	5	2	4
No weightings	2	3	7	6	5	1	4
Input 1	1	3	6	5	7	2	4
Input 2	2	3	7	6	5	1	4
Input 3	1	3	7	6	5	2	4
Input 4	1	4	6	5	7	2	3
Input 5	2	4	6	7	5	1	3
Input 6	2	3	7	6	5	1	4
Input 7	1	2	7	4	5	6	3
Input 8	1	3	6	5	7	2	4
Input 9	1	2	7	6	5	3	4

Table 22. The minimum changes required in indicator weights to cause a rank shift between the best option (ASP) and the second-best option (G-ASP). (Ling *et al.*, 2021).

	Group Weights w_k	Absolute Changes $\delta_{k,i,j}$	Relative Changes %
Energy neutrality	0.085	−0.224	263
Chemical consumption	0.078	−0.659	844
Total emission	0.119	−0.124	105
Odour	0.061	−0.270	442
Operability	0.067	0.103	154
Opex	0.126	−0.172	137
Capex	0.111	0.066	59
Flexibility	0.088	−0.098	112
Compliance	0.266	0.135	51

6.3.4 Discussion

It was easy and simple to combine AHP (for weighting development) and a simple additive model for calculating composite scores. The strength of simplicity of AHP has been widely acknowledged (Communities and Local Government, 2009; Karimi *et al.*, 2011; Velasquez and Hester, 2013). The methods were reliable in this study as the results of composite score and option rankings were very similar to the real decision results made by company stakeholders (Table 20). Additionally, the MCDA model also provided additional information in individual sustainability criteria as visualised in a stacked bar chart (Figure 33). Online questionnaires were an efficient alternative to focus groups for collecting AHP responses when a face-to-face setting was not feasible. However, there were also practical limitations of using online questionnaires. First, the implementation of the AHP scale was difficult because of compatibility problems in Microsoft Forms®. As such, the number of scale points was reduced. Second, some participants found the format of pairwise comparisons confusing. It can be difficult to explain the mathematical operation to participants if they request to check the background calculations since they were built and hidden in the Excel® spreadsheet. Although AHP is commonly perceived as an easy method, it was difficult to provide detailed information and training of the method through online questionnaires.

It was observed there were some concerns over the criteria structure and weights developed by AHP. First, given the requirement of pairwise comparisons, it seemed impractical when the number of items to be compared is large. A large number of comparisons are required by decision-makers and this is one of the main limitations of AHP due to a high resource intensity (Olson, 1988; Németh *et al.*, 2019; Odu, 2019). As a result, it can be difficult to maintain the consistency of responses made by decision-makers (Song and Kang, 2016). Second, the weights of indicators are directly influenced by the weights of the top criteria. This implies that different allocations of indicators in the criteria structure could change the final weights of indicators considerably, which was also highlighted by Song and Kang (2016). This implies that a new AHP may be required to develop a new set of weights if the criteria hierarchy changes. Therefore, AHP might not be the most flexible weighting method when there are hundreds of STWs managed by the water company with different investment needs (Ling *et al.*, 2021). Third, although AHP was relatively reliable to attribute weights, rank reversal still occurred between different weighting profiles as shown in Table 21. And the ranks of the top two options were very sensitive to changes in weight value of some indicators such as Capex (δ=0.066) and Flexibility (δ=-0.098) in Table 22.

In terms of the option ranking of wastewater treatment technologies in this pilot study, it is worth noting that, although ASP was included in the assessment as a 'baseline' solution by stakeholders, it had the highest aggregated score compared to other novel treatment technologies (Figure 33). Specifically, ASP was allocated with the highest score in the 3 out of 4 pillars (Social impact, Economic viability and Resilience) among all alternatives. This was due to high ratings given for ASP with respect to the indicators in those categories. Particularly, Operability (indicator in Social impact) of ASP was given the highest rating "5" and it also scored "4.33" on Compliance (indicator in Resilience). The rating on Compliance was also amplified by its high weight value allocated by AHP (Table 19). This reflects a strong preference of stakeholders towards ASP because of the vast knowledge and experience of operating ASP and the confidence of its operational (and environmental) compliance. Some stakeholders commented that they found it "*surprising*" because they were expecting innovative technologies to outperform ASP as the conventional technology. The observation revealed there was the underlying risk attitude towards wastewater technological selection and it may have, subconsciously, influenced the preference towards conventional alternatives even with the presence of novel technologies. This suggests that the selection of technology may be dominated by the strong preference towards compliance and operability of ASP and this requires further review on the ratings and weightings of indicators as well as the stakeholders' risk attitude. Nonetheless, the results of the sustainability assessment still initiated discussions and provide useful insights for stakeholders as a learning opportunity.

Overall, the development of assessment indicators and weights provided useful insights into using MCDA to perform sustainability assessments to compare wastewater treatment options. AHP, combined with online questionnaires, was easy to set up and perform. However, there were limitations in terms of its practicality, flexibility and rank reversal issue. The following changes were proposed to improve the robustness of the multi-criteria assessment methodology on the basis of the findings in this chapter:

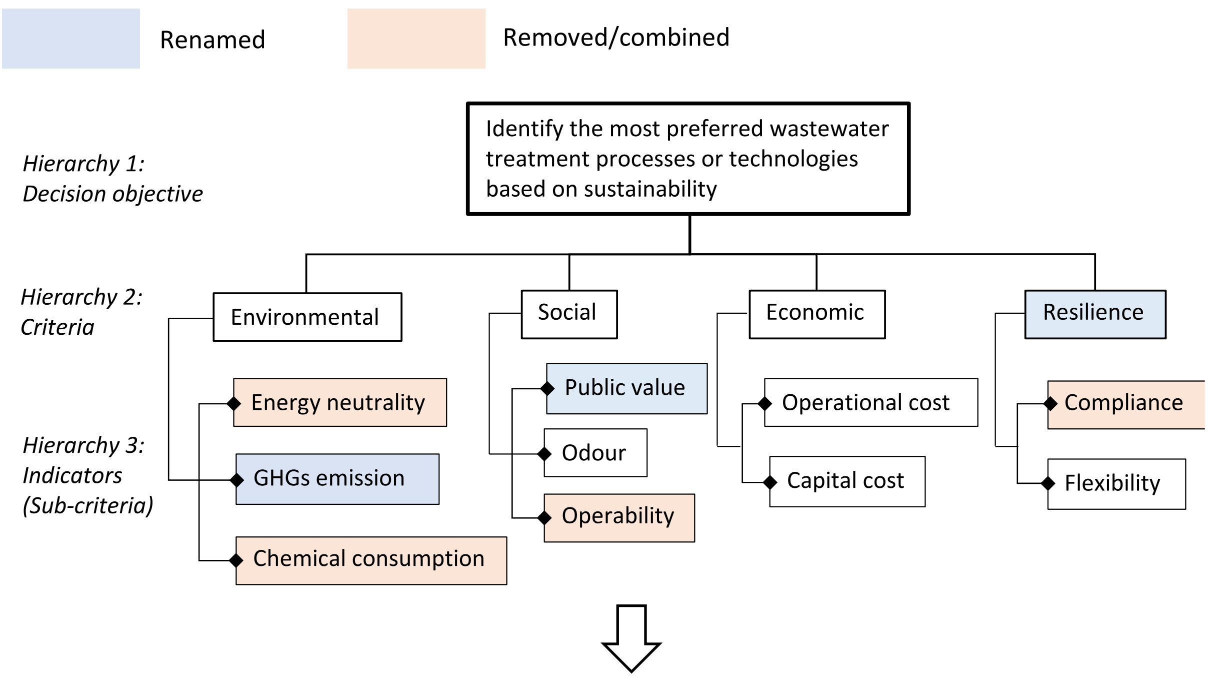
1. The criteria structure and list of indicators will be revised and updated
2. Alternative MCDA methods to AHP will be investigated for their potential suitability.

Chapter 7 Optimisation and validation of methods

The objective of this chapter is to make improvements and update the methodology of the assessment tool used in Chapter 6. This chapter first draws on the insight from the first pilot study to revise assessment indicators and criteria. Another pilot study was then applied to compare other MCDA models or methods reviewed in chapter 5 and proposed the final assessment methods before building the user interface.

7.1 Update on assessment criteria and indicators

As a continuous development process, the criteria structure was further revised and several changes were made. Based on the reflection of the first pilot study followed by consultations with company stakeholders, the naming and definitions of some indicators used for the first pilot study were not clear. Therefore, the main objectives for the revision were to provide clearer definitions and a more accurate representation of the sustainability criteria hierarchy. An updated criteria structure was proposed and shown in Figure 34. The specific changes and their justifications are described in the sub-sections below except for economic indicators which remain unchanged.



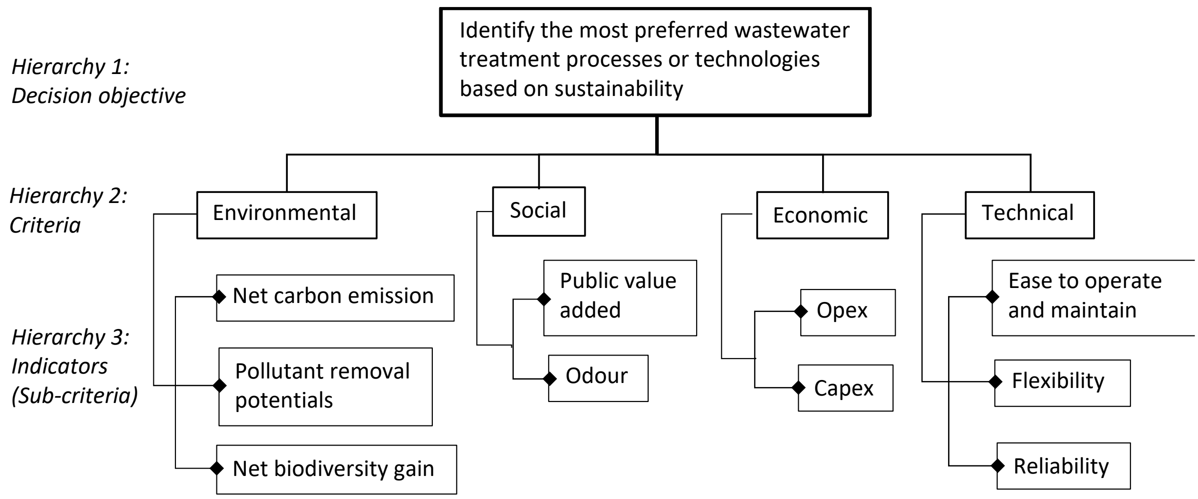


Figure 34. The top structure shows the previous criteria hierarchy used in Chapter 6 with highlighted changes. The bottom structure is the updated criteria hierarchy.

7.1.1 Environmental indicators

A new indicator named *Net carbon emission* was added to merge Energy neutrality, GHGs emission and Chemical consumption from the previous criteria structure. It was realised there was redundancy and double counting between them as they can all be expressed as the term of carbon emission. Energy consumption of the wastewater treatment process can be converted into GHGs emissions if electricity is sourced from fossil fuels (Pagilla *et al.*, 2012). Similarly, Chemical consumption can also be converted to emissions in terms of the manufacturing and transportation of the chemicals (i.e. embedded emission). Therefore, both power and material consumption indicators were merged into one indicator (Net carbon emission).

Pollutant removal potentials were re-added as an environmental indicator. It was previously merged with *Reliability* as one indicator (named *Compliance*) to reflect the overall ability and confidence of the treatment process or technology to achieve the effluent standard and the risks of failure. However, this makes the indicator difficult to measure because it was broad and intangible. Given this consideration, it was decided to retain *Pollutant removal potentials* and *Reliability* as two separate indicators in the criteria hierarchy.

A new indicator called *Net biodiversity gain* was added to reflect emerging priority in the investment decision. From the notes of personal communications and the finding of the previous case study in Chapter 4, biodiversity has become increasingly important as the strategic priority expands from environment protection to enhancement. The water company has set specific targets towards biodiversity improvement in its latest Biodiversity Policy, introducing A 5% biodiversity net gain was set on 253 sites of biodiversity interest and 10% net gain for engineering projects (Thames Water Utilities Limited, 2019c, 2020b). This also implies an increasing preference towards the practice of natural-based solutions that could enhance biodiversity. However, this only applies to a certain number of STWs to which the biodiversity interest status applies.

7.1.2 Social indicators

Within the social criterion, *Public value* was slightly renamed to *Public value added*. The definition of this indicator remains almost unchanged. It aims to indicate the level of additional values or benefits provided for customers and local communities such as leisure and visual amenities. *Biodiversity gain* was previously nested in the definition of this indicator but it was removed from the social criterion because it is more relevant to the environmental criterion (Pagilla *et al.*, 2012).

Operability was renamed to *Ease to operate and maintain* and was moved under the technical criterion. There are many variations in its definitions depending on the perspective upon which it is interpreted. In this study, it is defined as the level of ease, resources and skills required to operate and maintain the treatment process. Although it can be interpreted as the level of education required to operate (Srdjevic *et al.*, 2012), it was considered more relevant to the technical aspect of the process and technology.

7.1.3 Technical indicators

The name of the criterion was changed from *Resilience* to *Technical*[24] to serve as a 'pillar' in addition to the Three-Pillar of sustainability. This change was made to be more consistent with the categorisation and naming in the literature (Balkema *et al.*, 2002; Foxon *et al.*, 2002; Ren and Liang, 2017; Cossio *et al.*, 2020). Technical indicators are used to indicate various performances or technical aspects pertaining to the operation and maintenance of the asset (wastewater treatment facilities). Within the technical criterion, it includes *Ease to operate*

[24] Some studies used the term 'technological' or 'functional' indicators.

and maintain, *Flexibility* and *Reliability*. As mentioned earlier, *Ease to operate and maintain* was renamed after *Operability* and moved under the technical criterion. The previous indicator *Compliance* was divided into *Reliability* and *Pollutant removal potentials*. Whilst *Pollutant removal potentials* are the measurement or estimation of the ability to remove certain pollutants from the wastewater, *Reliability* indicates the potential risks or likelihood of failures and its impact on the effluent qualities (Balkema *et al.*, 2002; Molinos-Senante *et al.*, 2014). *Flexibility* remained unchanged from the previous criteria structure.

7.2 Pilot study 2: Application of selecting an advanced tertiary treatment option for phosphorus removal

7.2.1 Background

This study was based on a small-sized STW where a tighter effluent standard of phosphorus needs to be achieved. To meet the new consent, the site will be upgraded with an additional filter technology as an advanced tertiary treatment. Performance data were collected from the pilot-scale trial and upscaled to the full scale by the amount of flow. An illustration of the process is shown in Figure 35. Eight treatment scenarios were selected for this pilot study based on the combinations of two types of filter technologies[25], two dosing chemicals and two dosing ratios (Table 23).

Besides selecting the best phosphorus treatment option, this pilot study also aimed to provide practical insights into the selection of suitable MCDA models for the sustainability assessment tool through a comparative study. The comparative study applied those MCDA models and methods reviewed in chapter 5 for calculating the composite score. As MAVT and ELECTRE were considered unfeasible (referring to section 5.6.4) in preliminary screening, the rest MCDA models (SAW, AHP, SMART and TOPSIS) were selected for the comparative study. The option rankings derived from those models were compared to examine their consistencies. The consistency of results between different MCDA methods is a key factor when selecting a suitable MCDA model or method (Yoon and Hwang, 1995; Guitouni and Martel, 1998; El Amine *et al.*, 2014). Additionally, the feasibility and ease of use of these methods was also discussed to inform the selection (Roszkowska, 2011; El Amine *et al.*, 2014).

[25] The specific name and brand of each filter technology was labelled due to business confidentiality.

Table 23. Description of 8 treatment options as the combinations of 2 filter types, chemicals dosed and dosing ratios (based on molar ratio).

Options	Filter type	Dosing chemical	Dosing ratio
1	A	Ferric sulphate	4:1 (mol Fe: mol P)
2	B	Ferric sulphate	4:1
3	A	Ferric sulphate	6:1
4	B	Ferric sulphate	6:1
5	A	Polyaluminium chloride (PAC)	4:1
6	B	Polyaluminium chloride	4:1
7	A	Polyaluminium chloride	6:1
8	B	Polyaluminium chloride	6:1

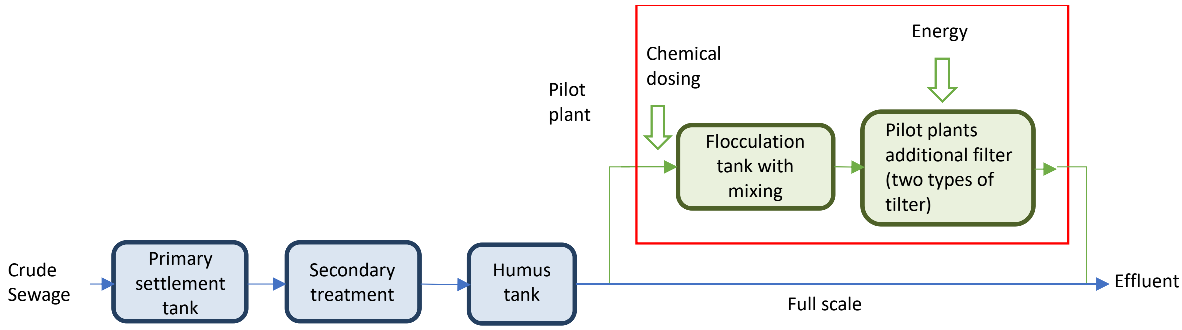


Figure 35. The flow diagram of the full treatment flow and the parallel pilot plant for tertiary treatment filters. The scope for assessment was based on the pilot plant section (Red box).

7.2.2 Data collection of indicators

The performances of quantitative indicators were calculated based on the average data collected from the pilot plant trial over 4 months. Costing data was sourced from the internal financial estimation system of the water company. The performance of indicators for the sustainability assessment was summarised in Table 24. Some indicators listed in the new criteria structure were excluded from the sustainability assessment. They were considered either not directly applicable to the scope of this study (such as *Biodiversity net gain*, *Odour* and *Public value added*) or the performance does not vary between different scenarios (such as *Flexibility*). The sub-sections below describe the methods and scope of the data collection of each indicator.

Table 24. The performances of indicators of 8 advanced phosphorus removal options for the multi-criteria sustainability assessment.

Indicators	Methods of collection	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Net carbon emission (tCO^2e/yr.)	In field measurement and estimation	25.5	24.6	36.8	35.9	255.6	254.7	374.8	373.9
TP removal (%)	In field measurement	90.2	80.4	93.6	91.2	91.5	93.0	88.9	31.9
Ease to operate and maintain	Expert judgement	4	3	4	3	4	3	4	3
Reliability	Expert judgement	4	2	4	2	4	2	4	2
Opex (£k/yr.)	Internal estimation	115	115	171	170	98	97	146	145
Capex (£k)	Internal estimation	850	650	850	650	850	650	850	650

7.2.2.1 Environmental indicators

The estimation of the carbon emission can be divided into three major components as capital carbon, operational carbon and embodied carbon, adopting a whole life cycle perspective (Prescott, 2009). There was little existing data of capital carbon emission (construction of the treatment assets) for all treatment options, and therefore, capital carbon was excluded from the scope of measurement in this study. The operational carbon mainly refers to the energy consumption of operating the treatment filters. As the assessment scope mainly focused on the tertiary treatment, the fugitive emissions (such as Nitrous oxides from aeration lanes in secondary treatment) were not included. The measurement of the embodied carbon mainly includes the emissions associated with the chemicals and materials used for dosing (Mo and Zhang, 2012; Pagilla *et al.*, 2012). The embodied carbon emissions between scenarios vary depending on the type of chemical used and the dosing quantity. Although attempts were made to estimate emissions associated with transport due to the amount of extra solids produced from the backwash in the treatment filter, there were uncertainties around estimating the number of tankers required for transportation (depending on the type of tanker and its volume), amount of extra sludge produced at the full treatment and routes of transportation (depending on which sludge treatment centre for delivery). Therefore, transportation was also excluded from the estimation of net carbon emission. In summary, the total emission in this study was the sum of operational emission from electricity use and

embodied carbon from chemical uses. The standard energy consumption of each filter technology was provided by the supplier. The data of chemical usage was provided by the pilot plant operator and recorded *in situ*. Due to confidentiality, the original data for energy and chemical consumption was not available in this thesis. The annual usages of energy and chemicals therefore can be converted to carbon emission by multiplying their corresponding emission factors (EFs) in
Table 25, as follows:

$$E_{energy}(CO_2 eq) = \text{Energy (kWh/d)} \times EF_{UK\ Electricity} \times 365 \quad \text{Equation 16}$$

$$E_{chemicals}(CO_2 eq) = \text{Quantity of chemicals (mol/d)} \times EF_{chemicals} \times 365 \quad \text{Equation 17}$$

Table 25. Emission factors of the electricity sourced from the grid and the two types of coagulants for chemical dosing. These references are aligned with the Carbon Account Workbook used by the water company for emission reporting.

	Electricity from UK grid ($kgCO_2e/kWh$)	Ferric sulphate ($kgCO_2e/mol\ Fe^{3+}$)	PAC ($kgCO_2e/mol\ Al^{3+}$)
Emission factors (EFs)	0.233[a]	0.013[b]	0.161[b]

a. (Department for Business Energy and Industrial Strategy, 2020); b. (INCOPA, 2014)

As this study was to select a treatment option that would achieve a greater removal of phosphorus from wastewater, the removal potential of Total Phosphorus (TP) was included as the indicator for Pollutant removal potential. The concentrations of TP were recorded before the pilot plant and the effluent over the course of the pilot trial. The removal rate can be calculated from the average concentrations as:

$$\text{TP removal} = 1 - \left(\frac{\text{ave. TP}_{\text{effluent}}}{\text{ave. TP}_{\text{Before pilot plant}}}\right) \times 100 \quad \text{Equation 18}$$

7.2.2.2 Technical and economic indicators

For Reliability and Ease to operate and maintain, the asset manager of the pilot plant was invited to rate each scenario based on a 5-point scale. The variation in ratings given (in Table 24) mainly depends on the type of treatment filter for the operation. In terms of economic indicators, the baseline Capex of the treatment filters were estimated and provided by the internal financial system. Opex of each treatment option was calculated as the sum of the maintenance cost, cost of purchasing chemicals and electricity. The maintenance cost was estimated by the internal financial system and chemicals and electricity costs were estimated from the usages recorded in the pilot plant trial using the internal cost conversion factors. The conversion factors were also not shown in this thesis due to confidentiality.

7.2.3 MCDA Methods

7.2.3.1 SAW

SAW is a linear additive weighted model that aggregates the performances of indicators by using a linear normalisation technique. Linear normalisation techniques include the Linear Sum method, the Linear Max method and the Linear Max-Min method, as shown in Table 26 below. All methods normalise performance values to a homogenous range from 0 to 1 so they can be further aggregated into a composite score. All three linear normalisation techniques were applied to this study and the ranking of options were compared to select the most suitable linear normalisation technique in SAW.

Table 26, There are 3 linear normalisation methods (Cinelli *et al.*, 2014; Miranda and Prasanna, 2014; Vafaei *et al.*, 2018a). a_{ij} refers to the performance of j-th indicator (or attribute) with respect to i-th option and v_{ij} is the value after normalisation.

1. Linear sum	Positive indicator	$v_{ij} = \frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}}$
	Negative indicator	$v_{ij} = \frac{(a_{ij})^{-1}}{\sum_{i=1}^{m} (a_{ij})^{-1}}$
2. Linear Max	Positive indicator	$v_{ij} = \frac{a_{ij}}{a_{max}}$
	Negative indicator	$v_{ij} = 1 - \frac{a_{ij}}{a_{max}}$
3. Linear Max-Min	Positive indicator	$v_{ij} = \frac{a_{ij} - a_{min}}{a_{max} - a_{min}}$
	Negative indicator	$v_{ij} = \frac{a_{max} - a_{ij}}{a_{max} - a_{min}}$

7.2.3.2 AHP

The essence of AHP relies on developing ratio scales through pairwise comparisons. Although AHP is more compatible with qualitative inputs using the AHP judgement scale, it also allows quantitative inputs such as the measurement of specific performance. Quantitative measurement requires to be converted into reciprocal ratios so they can feed into the AHP matrix (Wedley,1990). For pairwise comparisons between i-th option and j-th option with respect to a positive indicator:

$$v_{ij} = a_i/a_j \quad \text{Equation 19}$$

given that a_i and a_j are their performance respectively and v_{ij} is the ratio scale for the reciprocal matrix. This is equivalent to the judgement scale when deciding "how important is option i compared to option j?". For the same pairwise comparison with respect to a negative indicator, then the reciprocal is taken:

$$v_{ij} = a_j/a_i \quad \text{Equation 20}$$

As such, all performances data of indicators and options can be converted into ratio scales for the AHP matrix. The 'priorities' of each option can be derived by applying the same geometric mean method (similar to Equation 5 in chapter 6). The priorities of individual indicators were aggregated using a weighted additive model into composite scores.

7.2.3.3 SMART

As mentioned in Chapter 5, SMART normalise performance scores through linear approximation. The best and the worst options were identified regarding each indicator. The best performances were allocated with a value of 100 and a value of 0 for the most inferior one. A linear value function was then built on this basis to convert performances of other options to a value between 0 and 100 with respect to each indicator.

7.2.3.4 TOPSIS

There were 5 steps in TOPSIS to determine the best option based on the Euclidean distances to the ideal and worst ideal scenarios (Hwang and Yoon, 1981). The first step was to apply a vector normalisation to convert the performance (x_{ij}) of i-th option with respect to j-th indicator. Previous studies suggest that vector normalisation is the best normalisation technique for TOPSIS (Chakraborty and Yeh, 2009; Vafaei *et al.*, 2018a). The vector normalisation operates as:

$$a_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m}(x_{ij})^2}} \quad \text{Equation 21}$$

Then the second step was applying weights w_j to assessment indicators if weights are available:

$$a'_{ij} = a_{ij} \times w_j \quad \text{Equation 22}$$

The third step was to identify the ideal scenario and the worst ideal scenario. The 'ideal scenario', hypothetically, contains the best performances of individual indicators whereas the 'worst ideal scenario' consists of the worst performance of them. It is worth noting that the definition of the best or worst performance depends on whether it is a positive or negative indicator. For example, the best performance for Net carbon emission (i.e., a negative indicator) should be the option with the lowest level of emission. In contrast, the highest value would be considered the best for positive indicators such as Pollutant removal potentials. The fourth step is to calculate the Euclidean distances of option i to both the ideal scenario (denoted as D^{+}) and the worst ideal (D^{-}) scenario, as follows:

$$D_i^{+} = \sqrt{\sum_{i=1}^{m} (a'_{ij} - a_j^{+})^2} \quad \text{Equation 23}$$

$$D_i^{-} = \sqrt{\sum_{i=1}^{m} (a'_{ij} - a_j^{-})^2} \quad \text{Equation 24}$$

The last step was calculating the ratio between D_i^{+} and D_i^{-}. All alternatives can be ranked based on the principle that closer distance to the ideal scenario or further to the worst ideal scenario, the more desirable the option is. This ratio can be expressed as the Similarity Index S_i:

$$S_i = \frac{D_i^{-}}{D_i^{+} + D_i^{-}} \quad \text{Equation 25}$$

7.2.3.5 Weighting development

Two new weighting methods were used for weighting calculation for this pilot study: the direct rating method and the ratio weight method. For the direct rating method, the asset manager of the pilot plant trial was asked to provide a rating for each assessment indicator based on a 10-point scale. The greater importance of the indicator, the higher rating was given. The weights values of indicators were calculated by normalising the ratings and all weights added to 1. The ratio weight method requires the decision-makers to first rank all indicators based on importance and then allocate the least important indicator with a score of 10. And then scores can be attributed for the rest of the indicators based on the ascending importance. However, this was considered complicated because indicators must be ranked first. For this study, a baseline value of 100 instead of 10 was given to the first indicator, and then the asset manager can allocate scores to other indicators based on the relative importance to the baseline indicator in any order. Similarly, allocated scores were then normalised to 1. To check consistency, The asset manager was also asked to directly rank the importance of all indicators before using these two methods. This ranking was used as a reference to compare the results derived by these two weighting methods and examine the similarity of ranking.

7.2.4 Results and discussion

7.2.4.1 Comparing different normalisation techniques in SAW

Once the raw performance data (in Table 24) were normalised, the normalised matrices and the aggregated scores with option ranking are shown using the Linear Sum (Table 27), Linear Max (Table 28) and the Linear Max-Min method (Table 29). No weights were applied at this stage (i.e. each indicator was given a weight value of 1). Although the ranks of the best and the worst options were identical across three methods, results indicated that the Linear Sum produced a different normalised matrix compared to the Linear Max and the Linear Max-Min methods. For negative indicators such as Net carbon emission, the normalisation in the Linear Sum method fits a power function whereas the other two fits a linear model (example see Figure 36). For positive indicators, it was observed that all three methods fit a linear model. This can be explained by their mathematical operations. The Linear Max method normalised the best performances in positive indicators to a value of 1 but always produces a value of 0 in negative indicators. The Linear Max-Min method always normalises performances into a value range from 0 to 1 regardless of the type of indicator.

Table 27. Normalised performance matrix with option ranking using the Linear Sum method.

	Option 1	**Option 2**	**Option 3**	**Option 4**	**Option 5**	**Option 6**	**Option 7**	**Option 8**
Net carbon emission	0.265	0.275	0.183	0.188	0.026	0.027	0.018	0.018
TP removal	0.137	0.122	0.142	0.138	0.138	0.141	0.135	0.048
Operability and Maintenance	0.143	0.107	0.143	0.107	0.143	0.107	0.143	0.107
Reliability	0.167	0.083	0.167	0.083	0.167	0.083	0.167	0.083
Opex	0.137	0.138	0.092	0.093	0.161	0.162	0.108	0.109
Capex	0.108	0.142	0.108	0.142	0.108	0.142	0.108	0.142
Aggregated scores	0.956	0.867	0.835	0.751	0.744	0.662	0.679	0.507
Ranking	1	2	3	4	5	7	6	8

Table 28. Normalised performance matrix with option ranking using the Linear Max method

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Net carbon emission	0.932	0.934	0.902	0.904	0.318	0.321	0	0.002
TP removal	0.964	0.859	1.000	0.974	0.978	0.994	0.950	0.341
Operability and Maintenance	1	0.750	1	0.750	1	0.750	1	0.750
Reliability	1	0.500	1	0.500	1	0.500	1	0.500
Opex	0.326	0.330	0	0.004	0.427	0.431	0.147	0.151
Capex	0	0.235	0	0.235	0	0.235	0	0.235
Aggregated scores	4.222	3.609	3.902	3.368	3.722	3.230	3.096	1.980
Ranking	1	4	2	5	3	6	7	8

Table 29. Normalised performance matrix with option ranking using the Linear Max-Min method

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Net carbon emission	0.997	1	0.965	0.968	0.341	0.343	0	0.003
TP removal	0.945	0.786	1	0.961	0.966	0.990	0.924	0
Operability and Maintenance	1	0	1	0	1	0	1	0
Reliability	1	0	1	0	1	0	1	0
Opex	0.757	0.766	0	0.010	0.990	1	0.340	0.350
Capex	0	1	0	1	0	1	0	1
Aggregated scores	4.699	3.553	3.965	2.938	4.297	3.333	3.264	1.352
Ranking	1	4	3	7	2	5	6	8

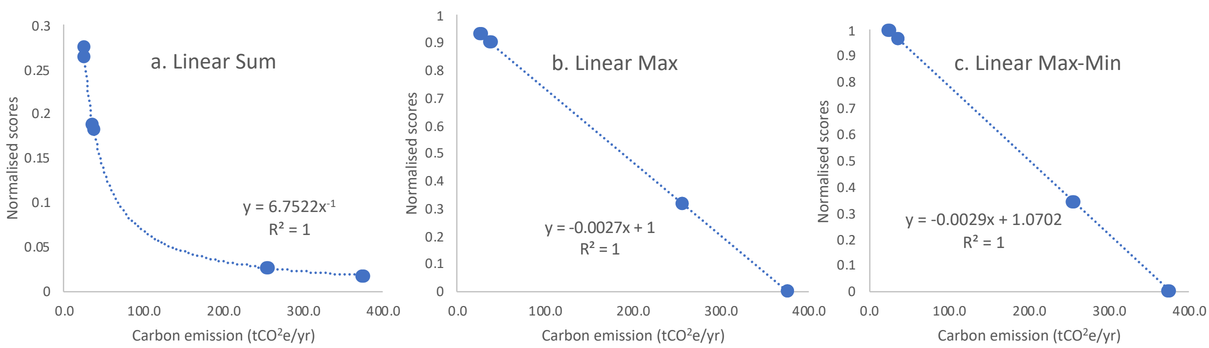


Figure 36. The trendlines and the fitted model for the normalisation of Net carbon emission using three normalisation methods.

There were concerns over whether the Linear Max and the Linear Max-Min method reliably normalise performances into proportionate values. It was observed that the normalised performances always contained extreme values of 0 or 1 in their normalisation matrices. For example, option 7 has the highest carbon emission among all options (Table 24). As a negative indicator, both Linear Max and Linear Max-Min methods normalised it into a value of 0 (Table 28 and Table 29). The question was: should the highest carbon emission be allocated with 0 at all? This is especially the case for the Linear Max-Min method. For indicators where there were only the two levels of performance among all options (Such as Capex), the better performance (£650) was normalised into 1 whereas the poorer one (£850) was normalised into 0, regardless of the degree of difference between the two performance scores. However, the question lies in whether the difference of £200 is proportionate to a full swing from 0 to 1? And what if the cost difference was only £10? Therefore, Linear Max and Linear Max-Min methods normalises disproportionately when the range of performance is small. In contrast, the Linear Sum method produces a more proportionate normalisation. The comparative study by Vafaei *et al.* (2018b) also highlight that Linear Sum is the best normalisation method whilst the Linear Max-Min is not suitable. Additionally, from a feasibility point of view, the Linear Sum method is easier and requires less time compared to other normalisation methods (Miranda and Prasanna, 2014). It was then decided to use the Linear Sum method for normalisation in SAW in the following comparative study between different MCDA models and the future implementation of the sustainability assessment tool.

7.2.4.2 Weight assignment

For this pilot study, the direct rating and the ratio weight methods were used to assign indicator weights. The results of both methods (Table 30) showed that Pollutant removal potentials were assigned the largest weight values as the most important indicator. Net biodiversity gain and public value added were assigned the lowest weight values. The results of Spearman's rank coefficient indicated that the rankings of indicator importance were highly similar to each other (Table 31). The ranking derived from the ratio weight method showed a higher correlation coefficient (ρ=0.942) than the direct rating method (ρ=0.833) when compared to the direct ranking made by the asset manager. This suggests that the ratio weight method may produce results that are more similar to the direct ranking. However, both methods assigned very similar weight values and their difference seems negligible.

In terms of feasibility, both weighting methods were simple to use. Compared to weights development using AHP, they required less time and cognitive capacities to perform. Another simple and common weighting method is the point allocation method which is to assign 100 points among all indicators. However, point allocation is increasingly difficult when the number of items for weight allocation is larger than 6 (Odu, 2019). Based on the experience of using both direct rating and ratio weight methods, the direct rating method was considered more suitable. It allows the adoption of an absolute scale, which remedies the issue of rank shifting when assigning weights based on relative importance. This means the rating allocated to each indicator are supposed to be independent of each other. And the removal of indicators should not affect the ratio between their weights. This was considered useful to apply to different projects where some indicators might be not applicable. By retaining the same rating scale of indicators, their weights can be re-normalised without undertaking another weight development process. This concept of 'absolute mode' has also been suggested by García-Cascales and Lamata (2012) to resolve the rank reversal problem. For this pilot study, the weights indicators were re-normalised and shown in Table 32 as some indicators were excluded.

Based on the discussion above, it was decided to use the direct rating method for the weighting development in the sustainability assessment. However, the direct rating method needs to be further tested to confirm its suitability. This method was further used and discussed alongside the usability testing of the sustainability assessment tool in Chapter 8.

Table 30. The inputs and results of two weighting methods and the importance rankings based on weights value. The asset manager was also asked to directly rank the importance of indicators as a reference.

		Direct rating method			Ratio weight method		
	Direct Ranking	Scale inputs	Weights	Ranking	Ratio inputs	Weights	Ranking
Net carbon emission (tCO2e/yr.)	2	8	0.100	5	100	0.103	4
Pollutants removal potentials	1	10	0.125	1	120	0.124	1
Net biodiversity gain	9	6	0.075	9	80	0.082	8
Public value added	9	6	0.075	9	80	0.082	8
Odour	5	7	0.088	8	90	0.093	7
Ease to operate and maintain	5	8	0.100	5	100	0.103	4
Flexibility (Adaptability)	8	8	0.100	5	80	0.082	8
Reliability	5	9	0.113	2	100	0.103	4
Opex (£k/yr.)	2	9	0.113	2	110	0.113	2
Capex (£k)	2	9	0.113	2	110	0.113	2

Table 31. The Spearman's rank correlation coefficients and the *p*-values between 3 rankings of indicators based on their importance. (Significance level α=0.05)

	Direct rank	Direct rating method	Ratio weight method
Direct rank	1.000	-	-
Direct rating method	ρ=0.833 (p=0.003)	1.000	-
Ratio weight method	ρ=0.942 (p<0.001)	ρ=0.891 (p=0.001)	1.000

Table 32. The weights were re-normalised by retaining their rating of importance. This weighting profile was used for aggregating into composite scores in this pilot study.

Indicators	**Scale ratings**	**Weights**
Net carbon emission (tCO2e/yr.)	8	0.151
TP removal (%)	10	0.189
Operability and Maintenance	8	0.151
Reliability	9	0.170
Opex (£k/yr.)	9	0.170
Capex (£k)	9	0.170

7.2.4.3 Comparing option rankings between MCDA models

The comparative study of different MCDA models reached a consensus that option 1 (Type A filter with Ferric Sulphate at the 4:1 dosing rate) was the most preferred option and option 8 (Type B filter with PAC at the 6:1 dosing rate) was the worst option (Table 33). The aggregated scores for individual sustainability criteria are also shown in Figure 37. Options with using PAC as a dosing chemical (i.e. option 5-8) tends to have a poorer environmental performance due to a larger embodied carbon emission than ferric sulphate (
Table 25). Options installed with filter type A (option 1,3,5,7) also scored better in the technical criterion, which can be explained by higher ratings on the Ease to operate and Reliability given by the asset manager (Table 24). The composite scores without applying weights were also calculated as a reference and shown in (Table 34). The option ranking of the 4 MCDA models remained almost the same compared to that with weights applied. Specifically, ranking in SMART and TOPSIS remains unchanged whilst there was a minor rank reversal between option 4 and option 5 in SAW and AHP.

When comparing MCDA models, SAW and AHP showed identical composites scores and option rankings. TOPSIS showed a similar option ranking to SAW and AHP. This observation is also similar to the comparative study of MCDA by Zanakis *et al.* (1998), although ELECTRE was not included in this pilot study. SMART produced very different results compared to other MCDA models. The operation in SMART is similar to the Linear Max-Min normalisation method. Specifically, the worst performances were converted into a value of 0 whereas the best ones were converted into a value of 100. As such, it converted the performances of each indicator into a linear scale that always contains normalised values of 0 and 1, for the worst and the best performance, respectively.

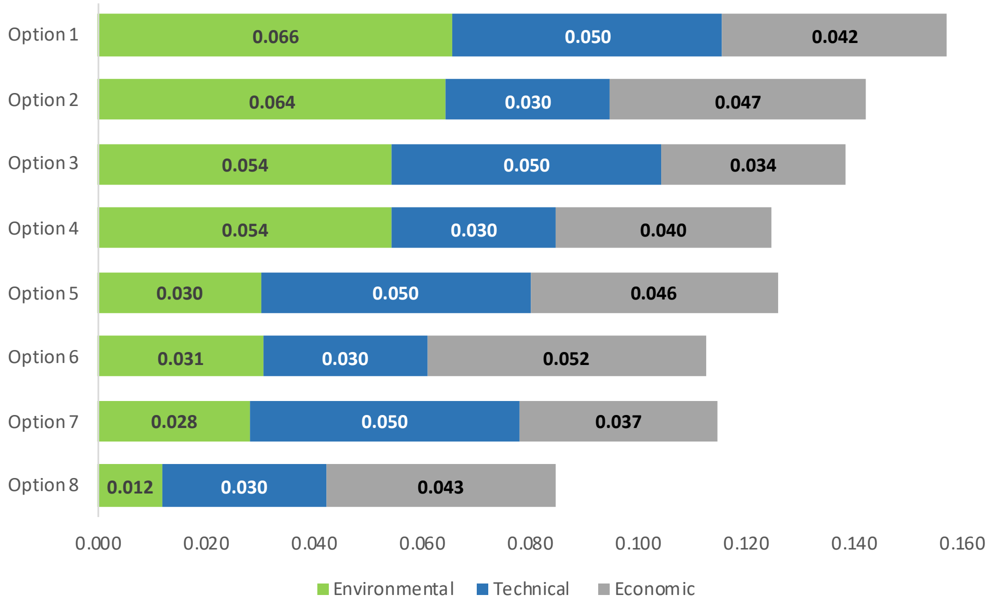


Figure 37. Aggregated weighted performance scores in each individual sustainability criteria (Environmental, Technical and Economic) based on SAW and AHP.

Table 33. The composite scores and ranking of 8 options using 4 MCDA models with weights applied from the last section (Table 32).

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
SAW	0.157	0.142	0.138	0.125	0.126	0.112	0.115	0.085
(Rank)	1	2	3	5	4	7	6	8
AHP	0.157	0.142	0.138	0.125	0.126	0.112	0.115	0.085
(Rank)	1	2	3	5	4	7	6	8
SMART	77.8	59.9	65.5	49.9	72.3	57.8	55.3	23.0
(Rank)	1	4	3	7	2	5	6	8
TOPSIS	0.854	0.690	0.736	0.641	0.572	0.496	0.416	0.156
(Rank)	1	3	2	4	5	6	7	8

Table 34. The composite scores and ranking of 8 options using 4 MCDA models without applying weights. (i.e. an equal weight was used for all indicators)

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
SAW	0.159	0.144	0.139	0.125	0.124	0.110	0.113	0.085
(Rank)	1	2	3	4	5	7	6	8
AHP	0.159	0.144	0.139	0.125	0.124	0.110	0.113	0.085
(Rank)	1	2	3	4	5	7	6	8
SMART	78.3	59.2	66.1	49.0	71.6	55.6	54.4	22.5
(Rank)	1	4	3	7	2	5	6	8
TOPSIS	0.862	0.704	0.747	0.654	0.542	0.468	0.380	0.149
(Rank)	1	3	2	4	5	6	7	8

Based on the results of this study, SAW (with a Linear Sum normalisation) was considered the most suitable MCDA model to be implemented. It provided consistent results compared to other MCDA models. Additionally, it is easy to understand and use (Zanakis *et al.*, 1998; Podvezko, 2011; Velasquez and Hester, 2013). This implies less time and resources required to build, use and edit the sustainability assessment tool. Another alternative MCDA method is TOPSIS. It is relatively simple to understand the logic in TOPSIS (Roszkowska, 2011; García-Cascales and Lamata, 2012) and results can be readily visualised (such as Figure 38). It was easy to identify that option 1 was the most desirable based on the principle of the closest distance to the ideal scenario or the furthest distance to the worst ideal scenario. However, the major drawback compared to SAW is the time and knowledge required to operate extra steps to calculate the Euclidean distances to the ideal and the worst ideal scenarios. This reduces the practical attractiveness of TOPSIS because it can be complex to understand and use for company end-users who have no prior knowledge of TOPSIS. Moreover, as mentioned in Chapter 5, rank reversals can still occur in TOPSIS because the definition of the ideal and the worst ideal scenarios is sensitive to the number of options compared (García-Cascales and Lamata, 2012). Another disadvantage of TOPSIS is that it is difficult to show the results of individual sustainability criteria as performances scores are fully aggregated into a similarity index based on distances. AHP was considered less feasible for the same reasons mentioned in the discussion in Chapter 6. AHP heavily relies on pairwise comparisons which can be resource-intensive when there were many items to be compared (Németh *et al.*, 2019; Odu, 2019). Due to the issue of rank reversals, the removal or addition of an indicator or alternative can alter the rankings of options. SMART was also not preferred because its normalisation technique produced a skewed value range, similar to the reason given for the Linear Max-Min normalisation method. Although the linear approximation simplifies the transformation to single-dimension values, it is disproportionate to allocate two extreme values of 100 and 0 for two performance levels of an indicator regardless of the degree of difference. Swing weights are often combined with SMART to remedy this shortcoming (Edwards and Barron, 1994). The swing weighting method takes the range of difference between performances of the indicator into account by asking decision-maker, if hypothetically, the performance of any indicator can be improved from the worst to the best, which indicator should be improved first? This is continued for the rest of the indicators until all indicators have been ranked in such way. The top indicator is then allocated with a 100-point swing and other indicators are allocated with scores compared to this full swing. The detailed explanation of the steps of swing weighting can be found in Edwards and Barron (1994) and Odu (2019). The drawback of swing weight is that its requires proper explanation for decision-makers to use it correctly (Németh *et al.*, 2019) and it can be resource-intensive (Tervonen *et al.*, 2017).

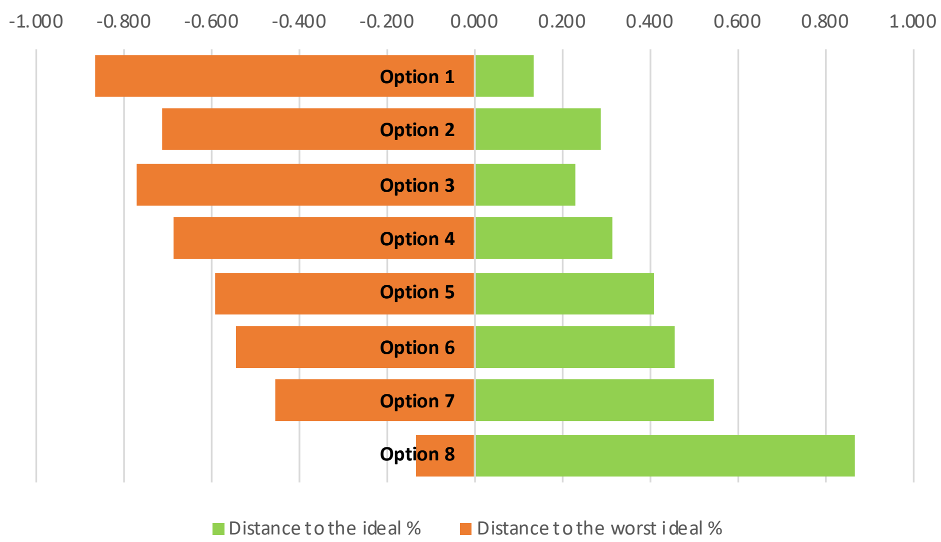


Figure 38. The distances of 8 options in percentage to the ideal and the worst ideal scenarios in TOPSIS. The closer to the ideal (or further to the worst idea), the more desirable the option is.

The results of the comparative study were also presented to the asset manager who was involved in the phosphorus repilot plant trial. The rationales and calculations of each MCDA model were explained and the asset manager was asked about the opinion on ease-of-use of each model. The feedback suggests that SAW seems to be the "*easiest*" and TOPSIS was the "*hardest*" to understand. Both AHP and SMART were given a "*medium*" rating. The feedback supports the evidence that SAW is most feasible MCDA model for the sustainability assessment tool from an ease-of-use aspect.

7.3 Summary of pilot studies

Although the literature review in Chapter 5 initiated the discussion on the suitability of different MCDA models, two pilot studies in Chapter 6 and Chapter 7 provided practical insights into the feasibility of potential methods through iterative testing and reflection. AHP was initially used to develop weights for indicators (Chapter 6) as well as a method for calculating composite scores (Chapter 7). Despite its popularity and ease-of-use, there were two major drawbacks of AHP. Firstly, AHP can be resource-intensive when dealing with a large number of pairwise comparisons. The second shortcoming is the rank shifting and reversal as a result of any addition or removal of indicators or alternatives from the comparison matrix. Additionally, such sensitivity also entails that AHP should be repeatedly performed to reflect any changes in the criteria hierarchy. However, this is not practical in the water company because of the diverse priorities of different STW projects. These shortcomings were the reasons why AHP was also not considered for comparing alternatives and calculating composite scores. From a feasibility aspect, although it was relatively easy to administer AHP through online questionnaire, this approach could lead to a lack of understanding and inconsistent use of AHP.

In the results of the comparative study in Chapter 7, although different models gave rise to consistent rankings of options, SAW with the Linear Sum normalisation technique was considered the most suitable. The Linear Sum technique is relatively simple to operate and produces a proportionate normalisation matrix for SAW. The two weighting methods (The direct rating and ratio weight methods) provided very similar weights and importance order compared to the direct ranking. The direct rating method was preferred to the ratio weight method because it allows weight allocation based on an absolute scale. The findings from two pilot studies underpinned the final methodological design for the sustainability assessment tool. The next step was to assemble the tool consisting of the updated criteria structure, SAW as the MCDA method and the direct rating weighting method into a user interface, which serves to communicate with the end-users.

Chapter 8 Testing of the sustainability assessment tool

8.1 Building the user interface

8.1.1 Content of the interface

Excel® was used as the platform to design and create the sustainability assessment tool. The interface of the tool consists of 7 worksheets of which 4 are directly related to the procedures of the MCDA assessment (Figure 39). Sheet 1 is a general user guide with a brief introduction to the purpose of the assessment tool and the procedures to complete the assessment. Its interface is shown in Figure 40. Sheets 2 to 4 are input sheets. To assist navigation for users, the cells that required inputs were highlighted in yellow (noted in the user guide).

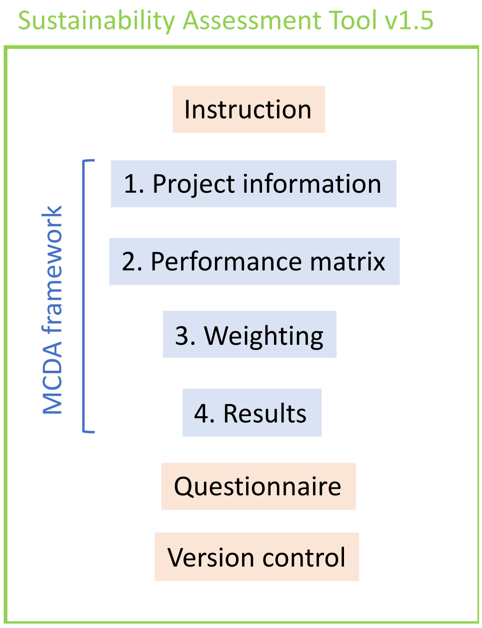


Figure 39. Content of the tool interface.

Worksheet 2 is the project information sheet (Figure 41). This sheet involves documenting the date of assessment, the name of the STW (or project name), the relevant scope of assessment with respect to the wastewater treatment process and the objective of the assessment. This project information sheet also requires users to include a brief description of each potential option to be compared.

Worksheet 3 is the performance matrix sheet that aims to compile the performance data of the assessment indicators (Figure 42). This worksheet has 4 sub-sections: a brief introduction to the objective of the sheet with basic instructions; the latest updated criteria structure for MCDA; the description of each option; and the data input section (highlighted in yellow). The definition of each indicator was attached to each cell as a note (indicated by a red marker on the top right corner of the cell).

Worksheet 4 is the weighting calculation sheet (Figure 43). Similar to previous input sheets, basic instructions are provided to guide the rating of indicators. A drop-down list was built into the sheet to provide a list of input (i.e. rating scale) for the users. It was emphasised that ratings should be given for all indicators so that a new group weighting profile can be developed alongside the following usability testing. As the last input sheet, an automated message box was built into the sheet to indicate its completion once all inputs have been provided.

Worksheet 5 for the sustainability assessment is the results sheet. MCDA formulae (such as data normalisation and SAW) were built into the sheet and results are automatically displayed once all inputs have been provided. The content of the sheet includes three graphs summarising the results of MCDA. The interface is demonstrated in Figure 44 with random inputs. The first graph is a bar chart for comparing the composite scores of all options. The second graph is a 'breakdown' of the first graph by retaining the aggregated scores in each sustainability criterion. The third graph is a Pareto Efficiency graph which compares options based on the ratios between their aggregated score (without economic indicators) and the whole life cost. This chart provides a more realistic outlook in projects where costs are a major constraint and trade-offs with the performance of other indicators are deemed unsuitable. The tabular results for the visualisation are also presented alongside the graphs for transparency.

The assessment tool also contains two peripheral Worksheets (6 & 7). Worksheet 6 is temporary, containing a questionnaire to collect feedback for the usability testing and will be removed once the usability testing research has been completed. Worksheet 7 is a version control sheet which documents the key changes made to the assessment tool up to the latest version. Both sheets are hidden to users but can be set to visible using the '*unhide*' function in Excel®. Additionally, all sheets in the assessment tool are '*protected*' by default to prevent unauthorised and unintended changes made by users during the usability testing.

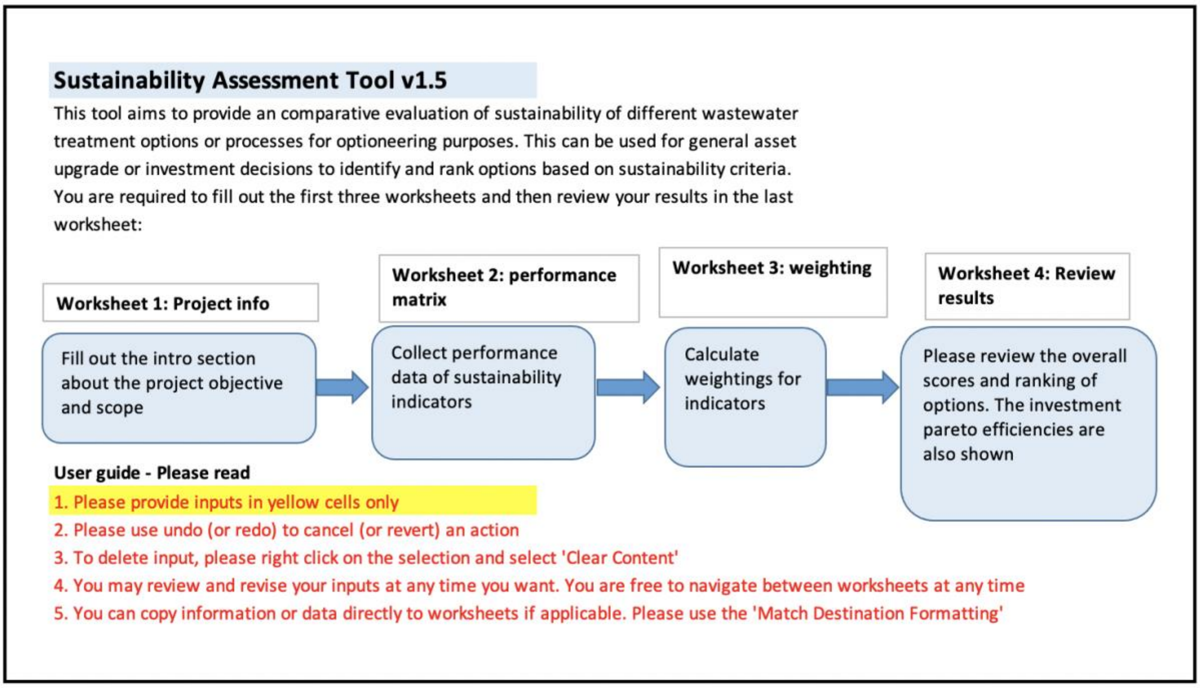


Figure 40. Interface of the introduction sheet in the assessment tool.

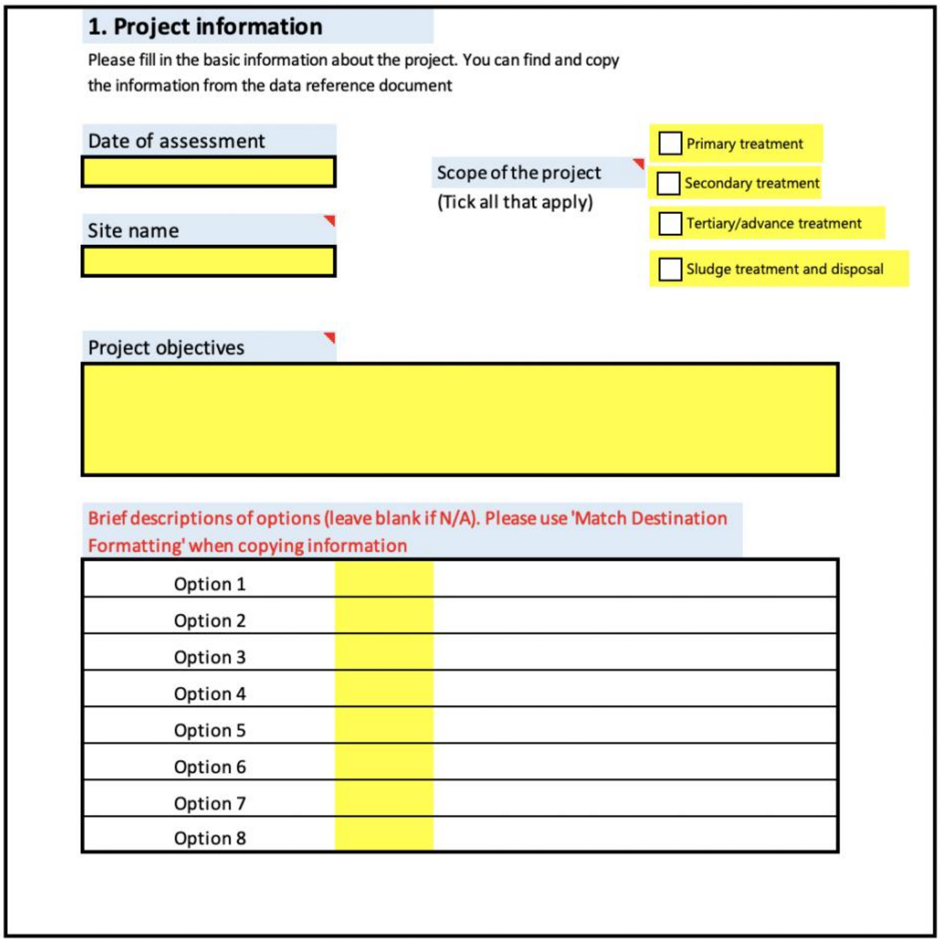
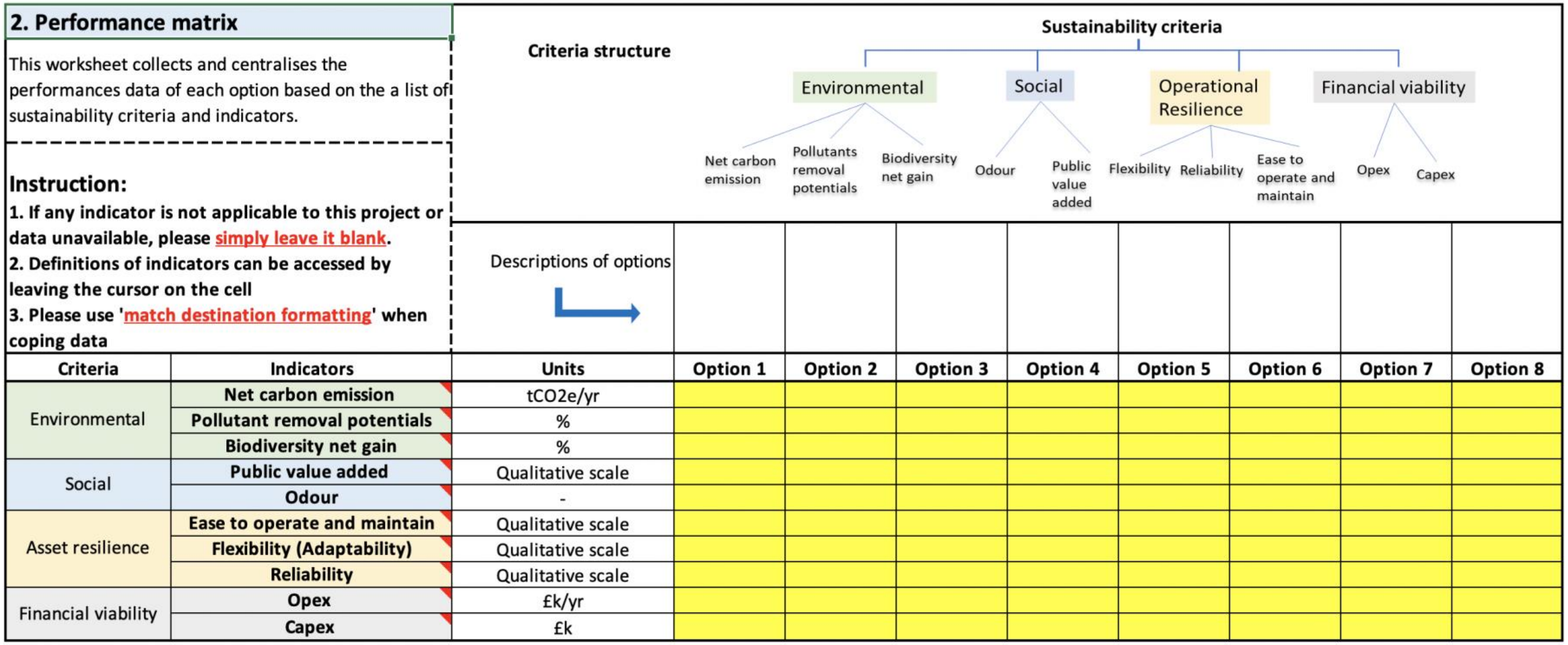


Figure 41. Interface of the project information worksheet in the assessment tool.


2. Performance matrix

This worksheet collects and centralises the performances data of each option based on the a list of sustainability criteria and indicators.

Instruction:

1. If any indicator is not applicable to this project or data unavailable, please simply leave it blank.
2. Definitions of indicators can be accessed by leaving the cursor on the cell
3. Please use 'match destination formatting' when coping data

Criteria	Indicators	Units	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Environmental	**Net carbon emission**	tCO2e/yr								
	Pollutant removal potentials	%								
	Biodiversity net gain	%								
Social	**Public value added**	Qualitative scale								
	Odour	-								
Asset resilience	**Ease to operate and maintain**	Qualitative scale								
	Flexibility (Adaptability)	Qualitative scale								
	Reliability	Qualitative scale								
Financial viability	**Opex**	£k/yr								
	Capex	£k								

Figure 42. Interface of the performance matrix sheet in the assessment tool

3. Weighting calculation

Instruction

This spreadsheet calculates the weight of each indicator. Please provide to your rating from **1 to 10** (1, 1.5, 2, 2.5 ..., 10) using the drop down list to reflect the importance of **all the indicators below**.
1 represents the lowest importance and 10 represents the highest.

If you think they are all equally important, please input the same ratings for all of them

Indicators	Rating	Weights	Rank
Net carbon emission		?	?
Pollutant removal potentials		?	?
Biodiversity net gain		?	?
Public value added		?	?
Odour		?	?
Ease to operate and maintain		?	?
Flexibility (Adaptability)		?	?
Reliability		?	?
Opex		?	?
Capex		?	?

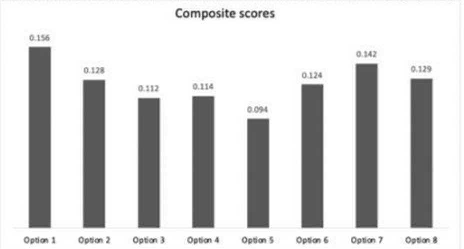
Figure 43. Interface of the weighting calculation sheet in the assessment tool.

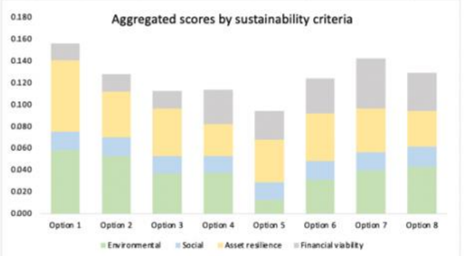
4. Results and visualisation

The total score of each option is automatically calculated based on your inputs. The tabluar results are shown on the right and graphs are provided on the left side. Please review your results.

a. Composite scores

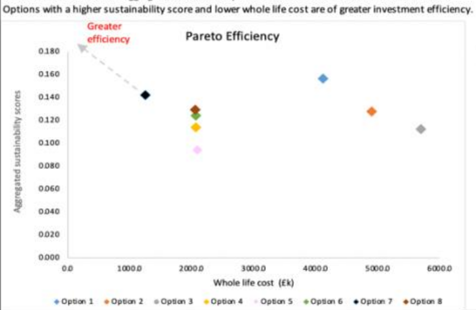
This is the final score of options after performances of indicators are normalised and aggregated by weig





b. Pareto efficiency

This the ratio between the aggregated sustainability score and the whole life cost of the solution. Options with a higher sustainability score and lower whole life cost are of greater investment efficiency.



Descriptions of options

Noramlised matrix	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Weights
Net carbon emission	0.297	0.247	0.066	0.093	0.053	0.038	0.119	0.087	0.104
Pollutants removal potentials	0.167	0.165	0.128	0.109	0.000	0.163	0.105	0.163	0.122
Biodiversity net gain	0.083	0.083	0.167	0.167	0.083	0.083	0.167	0.167	0.085
Public value added	0.130	0.130	0.111	0.111	0.111	0.130	0.130	0.148	0.061
Odour	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.073
Ease to operate and maintain	0.186	0.140	0.163	0.070	0.116	0.140	0.116	0.070	0.110
Flexibility (Adaptability)	0.196	0.196	0.130	0.087	0.109	0.130	0.087	0.065	0.104
Reliability	0.200	0.050	0.100	0.100	0.125	0.125	0.150	0.150	0.122
Opex	0.072	0.060	0.051	0.144	0.144	0.144	0.240	0.144	0.110
Capex	0.073	0.088	0.098	0.146	0.098	0.146	0.176	0.176	0.110
Aggregated sustainability score	**0.156**	**0.128**	**0.112**	**0.114**	**0.094**	**0.124**	**0.142**	**0.129**	
Rank	**1**	**4**	**7**	**6**	**8**	**5**	**2**	**3**	
Senstivity check	0.153	0.128	0.114	0.115	0.096	0.122	0.141	0.130	
Baseline rank	1	4	7	6	8	5	2	3	

Your results

Aggregated scores by criteria	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Environmental	0.058	0.053	0.037	0.037	0.013	0.031	0.039	0.043
Social	0.017	0.017	0.016	0.016	0.016	0.017	0.017	0.018
Asset resilience	0.065	0.042	0.044	0.029	0.039	0.044	0.040	0.033
Financial viability	0.016	0.016	0.016	0.032	0.027	0.032	0.046	0.035

	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Aggregated sustainability score	0.156	0.128	0.112	0.114	0.094	0.124	0.142	0.129
Whole life cost (£k/ 40 yrs)	4120.0	4900.0	5690.0	2060.0	2090.0	2060.0	1250.0	2050.0

Figure 44. Interface of the result sheet with random inputs as an example.

Table 35. Descriptions of 10 usability heuristics principles (Nielsen, 1994) and their application in the design of this assessment tool.

Usability heuristic principles	Descriptions	Application in the design of this tool
1. Visibility of system status	Provide prompt feedback on the user's action to inform what is happening	A status bar was added to inform when all inputs are complete.
2. Match between system and the real world	The system should use language, words and terms that the user can understand and be familiar with	Detailed definitions were provided for any special terminology used through notes attached to the cell.
3. User control and freedom	Users should have the option to exit or revert (e.g. undo or redo) from an unwanted state with clearly marked exits. Information should be provided in a natural and logical order	The spreadsheet has all necessary buttons to cancel or revert an action and exit the tool. Users can also use the 'clear content' function to erase the whole data input. This information is noted in the instruction sheet for the tool.
4. Consistency and standards	Be consistent with the meaning of words, situations or actions in the interface system	Wording and any terms were cross-checked between different worksheets to ensure consistency.
5. Error prevention	The system design should attempt to obviate any error in the first place.	Input requirement was mentioned in the instruction of each task sheet. Users will be informed by an error message if they input in the wrong data format or attempt to put the data in the wrong place.
6. Recognition rather than recall	The design should avoid heavy requirements on the user's short-term memory. The information in the interface system should be recognisable and retrievable	All the inputs remain visible and accessible to users after the completion of each tab. Users were free to review any worksheet if needed.
7. Flexibility and efficiency of use	Make the interface easy and flexible to navigate and complete tasks	The number of tasks in the assessment was minimised as far as possible. Each worksheet was clearly labelled on the tabs.
8. Aesthetic and minimalist design	Only relevant and essential information should be visible for the users in the interface. Prioritise the information in the interface to assist the objective of the primary task	The view settings of each worksheet were modified. Gridlines, formulae bars and headings were hidden to reduce information load. The main body of the sheet was positioned in the central space with a reasonable size.
9. Help users recognise, diagnose, and recover from errors	Errors message should be clear, avoid negative language, and offer a solution. Shneiderman (1982) suggests that good error messages are positive, precise, and constructive	Error messages were customised in the 'data validation' function in Excel to provide a clear and constructive solution to the issues.
10. Help and documentation	The interface system should provide all necessary information and explanation to users.	Key instructions were provided in the instruction sheet of the assessment tool and at the beginning of each input sheet.

8.1.2 Applying usability heuristic principles

The design of the tool's interface was supported by usability heuristic principles, which are guidelines for making a usable and user-centred system design. Such heuristic principles are easy and quick to use and particularly useful in the early to middle stage of the design process (Nielsen and Molich, 1990; Simeral and Branaghan, 1997). Table 35 summarises the description of each usability heuristic principle and how it was applied to the design of this tool.

8.2 Usability testing

8.2.1 Background

Usability testing was undertaken with potential end-users to assess the usability of the tool in a simulated environment. The testing examined two research questions: (1) How easily and successfully do end-users navigate the interface and complete the sustainability assessment independently? and, (2) and how useful is the assessment tool? These questions serve 3 purposes:

1. To assess potential users' experience with the tools' overall ease of use, usefulness, and the extent of user satisfaction
2. To identify design flaws in the interface of the tool and propose improvements
3. To establish a benchmark for usability testing and determine if more testing is needed

8.2.2 Design and methods

8.2.2.1 Recruitment

The required user roles for recruitment to the testing were Asset Planner, Asset manager and System planner as they were identified as relevant end-users in based on the stakeholder analysis conducted in Chapter 4. Participants were also required to have some experience in data management or analysis using Excel to ensure that participants have the minimum skills to use the assessment tool. A pilot trial was conducted with an asset manager to suggest a provisional length of the testing session and to test the feasibility of the test environment and materials. Any critical issues detected were resolved prior to the formal usability testing (for a full summary of the pilot trial please see **Appendix 7**). A total of 16 potential end-users (excluding the pilot trial) were then selected and invited through emails. The whole testing consisted of two case studies to be applied to recruit 8 participants for each case (Table 36). This should be sufficient to uncover most usability problems because four or five participants are often proposed as the minimum number (Virzi, 1992; Nielsen and Landauer, 1993). This recruitment strategy was based on an independent group design, which means a unique group of users was recruited for the two different cases. This design was selected to minimise the potential "*transfer-of-learning*" effect (Rubin and Chisnell, 2008, p. 75). The potential limitation is that each participant may only test the tool once.

Table 36. Summary of the recruitment for usability testing.

Case A: Phosphorus removal options comparisons (previous pilot study 2)	Case B: Selecting a secondary treatment scheme for a new STW (previous pilot study 1)
Participant 1 (pilot trial) Participant 2 Participant 3 Participant 4 Participant 5 Participant 6 Participant 7 Participant 8 Participant 9	Participant 10 Participant 11 Participant 12 Participant 13 Participant 14 Participant 15 Participant 16 Participant 17

8.2.2.2 Design of test session

As per Chapter 3, the type of usability test used for this research was the 'assessment' type which aims to assess the interface and reveal any major design flaws. The test sessions were conducted remotely using Microsoft Teams®. The reasons for using a remote set-up were: 1) the ongoing social distancing condition due to the Covid-19 pandemic prevented a face-to-face setting; 2) all users have the access to the software; 3) Microsoft Teams® provides an efficient platform to digitally distribute the test materials and enable recording (audio and screen recording) of the test session; 4) cost-effectiveness (Dray and Siegel, 2004; Bastien, 2010). Screen sharing and recording were used to collect observational data and monitor the interaction between the users and the tool. As the facilities (e.g. laptop, software, broadband) were available for all internal employees of the company, remote testing was an easier and cheaper alternative than running tests in a usability laboratory.

Test sessions were administered synchronously with presence of a moderator (the researcher) virtually. This was to ensure assistance can be offered immediately if participants experienced technical difficulty or requested interventions. Asynchronous tests are difficult to conduct without the specialised software to record and analyse data and they tend to reveal fewer usability issues (Andreasen *et al.*, 2007). The presence of the moderator has three purposes: 1) to provide a brief introduction and training at the start of the test session; 2) to give a post-task debriefing and 3) to build rapport with participants for discussing the results.

Each test session was provisionally designed to be approximately 45 minutes long. The outline of each test session included, sequentially:

- Orientation (10 minutes): This part welcomed the participant, introduced the purpose of this study, introduced the assessment tool, and provided a brief description of the case to be applied. The task materials (i.e. the assessment tool) were not directly shown to the participants at this stage to maintain the same level of knowledge for all participants. At the end of orientation, participants were requested to read and sign the consent form (**Appendix 8**). The delivery of the orientation was supported by a presentation (screen-shared with the participant). A written orientation script (**Appendix 9**) was used to convey consistent information to all participants.
- Task assessment (15 minutes): the moderator sent the test materials to the participant after agreeing and signing the consent form. Two files were sent to each participant: the assessment tool (an Excel® file) and the data reference document for the case study to be used (a Word® file). The start of the task was marked by the action of the participant opening both documents and the end of the task was marked by the verbal signal given by the participant when all tasks had been completed.
- Post-task debriefing (15 minutes): the participant was asked to complete a questionnaire followed by a verbal discussion with the moderator of some open-end questions. Participants could also use this opportunity to discuss and comment on any other aspects of the tool that were not included in the questionnaire.
- An extra 5 minutes were reserved as contingency.

8.2.2.3 Descriptions of tasks

After the initial briefing, each participant was asked to complete the task independently with minimal interaction with the moderator. The completion of each assessment necessitated five tasks: (1) Reading the instruction and learning about the tool; (2) Filling out the project information sheet; (3) Filling out the performance matrix sheet; (4) Filling out the weighting calculation sheet; and (5) reviewing the results of the sustainability assessment. Although these steps were designed to be completed sequentially, participants were allowed to navigate freely and revise inputs. Individual tasks were not tested in separate usability tests because users need to complete all these steps together to obtain the result of the assessment.

For this testing of the assessment tool, the materials in previous pilot studies in Chapters 6 and 7 were reused as 2 case studies. The project information and performance of indicators were included in a data reference document which was sent to participants along with the assessment tool before the start of testing. The time and needs for the collection of data and information (e.g. indicators) by users are thus excluded from the usability testing for two reasons:

1. it did not fall into the primary scope of the testing
2. it is impractical to include the data collection as a task in usability testing because the time spent to collect and calculate the performance data of assessment indicators can vary considerably depending on other factors such as the information availability on the specific project and the role of the user in the company etc.

Providing the same dataset for all participants normalised the time required to complete the assessment tool itself, it was acknowledged that this may not represent the total amount of time needed to complete the entire sustainability assessment process.

8.2.2.4 Data collection of usability testing

Two types of data were collected from the usability testing: performance data and preference data. Performance measurements were recorded by real-time observations and reviewing the screen recordings. Preference data was collected using the post-tasks questionnaire and debriefing. Measurements of performance data included: (1) time duration for completing the whole assessment as well as individual tasks; (2) the number of errors made. Errors here refer to misuses of the functions in the tool (e.g. the user put the wrong information in the wrong place) and omissions of key information (e.g. the user omits to provide a row of mandatory input for data analysis); (3) the number of prompts made by the moderator or requested by the participant. Prompts refer to verbal assistance provided by the moderator in situations where the participant required help to proceed. Interesting questions or comments made by participants during the task were also observed and recorded.

Once the participant has signalled the completion of all tasks, they were asked to give a rating for the 6 statements in the questionnaire (Worksheet 6) using a 5-point Likert scale (Table 37). The statements were based on various aspects of usability such as ease of use, utility and user satisfaction. The statements in the questionnaire were:

1) *The interface of the tool is easy to navigate*
2) *The tool is easy to use*
3) *I can mostly understand what I need to do to complete the analysis*
4) *I found the results and graphs useful*
5) *The tool would be useful for comparing wastewater asset options and inform investment decision*
6) *I would recommend this decision support tool to other colleagues working with wastewater asset planning*

Table 37. The Likert scale used for rating in the post-task questionnaire.

1	2	3	4	5
Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**

Participants were also verbally debriefed by the moderator after completion of the questionnaire. The purpose of the debriefing was to elicit verbal information regarding the strengths and weaknesses of the interface and provide insights into how to fix the problem (Rubin and Chisnell, 2008, p. 229). The alternative way to collect verbal data is the think-aloud approach. The think-aloud approach involves participants verbalising their thoughts and logic while performing tasks. The strength of the think-aloud approach is that it can uncover a great depth of interaction between the user and the interface and expose a problem promptly. Compared to post-tasks debriefing, think-aloud captures 'live' data so participants do not need to recall their thought process later on (Baauw and Markopoulos, 2004). However, the main drawback is that it diverts part of cognitive resources when completing a task, which may compromise the participant's performance. Therefore, the think-aloud approach was not used in the usability testing because it may compromise the performance of using the assessment tool given that the assessment tool is new to all users.

8.2.2.5 Data analysis

The performance data recorded and observed from the usability testing were analysed using descriptive statistics. Mean values and standard errors for the completion time were calculated for each case study and both cases together, as well as the completion time for individual tasks in the usability testing. The total count of errors and prompts observed in each assessment sheet was also compiled to highlight where most errors and prompts occurred. Descriptions of errors and prompts were also recorded as observational data to provide explanations. For the preference data, central tendencies (the mean and median) of the ratings in questionnaires were calculated among all participants. The median was calculated because Likert scale is often seen as an ordinal scale and the mean may not be suitable. The audio recordings of debriefings were transcribed based on three pre-defined codes "Strength", "Concerns and problems" and "Suggestions". The qualitative data from the debriefing sessions were then used to explain and corroborate the results of the questionnaires. Overall, findings from performance data and preference data were reviewed to identify the strengths and flaws of the tool design with suggestions for improvement.

In addition to the results on usability testing, the option rankings and weightings by participants in each test session of the sustainability assessment were also recorded. These option rankings were compared across participants to check their consistency. As the performance data provided for the indicators was the same for all participants, any discrepancy in option ranking would be caused by the variations in the weighting profiles of the different participants. This will indicate whether the weighting method embedded in the tool provides for consistent option ranking by different individual users. The usability testing was also used as an opportunity to develop a new group weighting profile by aggregating individual weighting profiles from all the testing sessions. This group weighting profile was compared to the previous profile (see Chapter 6) and potentially could be used as a generalised weighting profile for future applications of the sustainability assessment tool.

8.2.2.6 Ethics

An ethics review was undertaken prior to the recruitment for the usability testing. Specifically, a self-assessment form (i.e. SAFE form) was completed by the researcher following the requirement of the Research Ethics Committee at the University of Surrey. It was assessed that this part of the study did not require a full review from the ethics committee. Participants were requested to fill out a consent form which includes details of how and when their data will be collected, accessed and analysed, and destroyed. To reduce the risk of being identifiable, user audios from screening recordings were transcribed immediately after the test session using self-dictation (i.e. the researcher dictates exactly what he hears from the audio recording and then transcribes) and the voice of participants was removed. Recordings files were then deleted once the measurements of observations data were recorded and transcriptions were completed.

8.2.3 Results

8.2.3.1 Performance data

A total of 8 participants and 5 participants accepted the invite and took part in the usability testing for case A and case B, respectively. The time duration for completing the assessment using the tool was recorded and compiled from the testing sessions. Table 38 shows the average time to complete the assessment for all usability tests (n=13) as well as for case A (n=8) and case B (n=5) separately. On average, it required 12.1 mins to complete the whole assessment using the tool. The time for completing individual tasks was also shown in Figure 45. The results highlighted participants spent the most time filling the project information sheet ($\bar{t}$ = 3.7 mins) and the performance matrix sheet ($\bar{t}$ = 3.6 mins). This is expected as both sheets require reading the data reference document to select the correct information for the input. The other input sheet (weighting calculation sheet) required much less time to complete ($\bar{t}$ = 1.9 mins). It was observed that the time participants spent reviewing the result sheet varied greatly. This is reflected by a larger standard error in Figure 45. Therefore, the completion time excluding reviewing results was also calculated. The total completion time is more consistent and standard errors are smaller as shown in Table 38.

Table 38. Average time (mins) for completing the assessment using the tool with standard errors. The completion time without reviewing results was also shown.

Completion time - **Including** reviewing results	Total (n=13)	Case A (n=8)	Case B (n=5)
Mean (mins)	12.1	12.4	11.7
Standard error	0.7	1.1	0.5
(Completion time - **Excluding** reviewing results	Total (n=13)	Case A (n=8)	Case B (n=5)
Mean (mins)	10.4	10.5	10.2
Standard errors	0.4	0.7	0.2

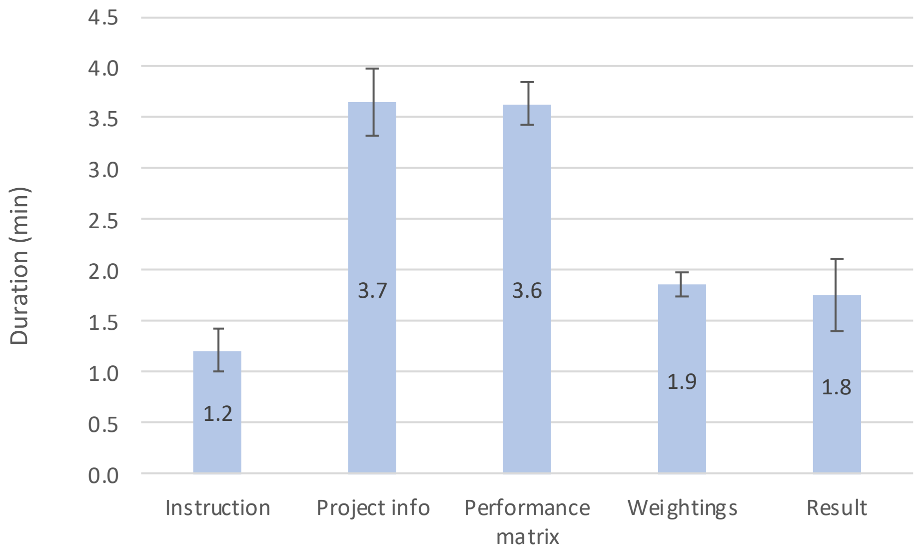


Figure 45. The average completion time (mins) of individual tasks in the assessment with standard error bars (n=13).

The number of errors and prompts made by the participant were observed and recorded. Figure 46 shows the count of errors and prompts and their distribution in different task sheets. A total of 22 errors and 9 prompts was recorded from all testing sessions. The descriptions of frequent errors and prompts are also shown in Table 39. Most errors and prompts occurred in the input sheets (Project information sheet, Performance matrix sheet and Weighting sheet). There was a spike of errors in the Project information sheet. Most errors observed were related to the unexpected change in the formatting of the sheet when copying information or copying the wrong information from the data reference document. Some participants omitted the 'Match Destination Formatting' in the guideline when copying and this caused a glitch in the table format aesthetically. An interesting observation was that some participants decided to manually type the information despite 'copy and paste' being mentioned in the instructions. Some participants attempted to fill out the project section without fully reading the data reference document and, occasionally, inaccurate information was filled in the 'project objectives' section.

Figure 46. Counts of errors and prompts made by participants in individual tasks of the assessment.

Table 39 Description of errors and prompts observed in each task in the assessment.

	Errors	Prompts
Instruction	• Omission of instruction	-
Project info	• Formatting errors when copying and pasting • Omission of the data reference document when filling out the project info • Participant typed the wrong information into the 'objective' tab	• Participants asked whether copy and paste were allowed and affect formatting
Performance matrix	• Formatting errors when copying and pasting	• Participants asked whether copy and paste were allowed and affect formatting
Weightings	• Did not provide ratings for all indicators	• Participants asked whether the weighting was based on their own judgements • Participants asked whether the same rating can be used more than once
Results	• Omission of the final graph	-

In terms of prompts, many participants asked whether it was allowed to directly copy and paste information and data. Prompts were given to iterate the 'Match Destination Formatting' guideline in the instruction. Other prompts were related to the weighting calculation sheet. Some participants were confused with whether to provide ratings based on his/her own opinions or use existing weighting. Other participants were not sure if it was allowed to use the same rating number more than once. This suggests that the guidelines provided in the tool may not be clear and sufficient.

8.2.3.2 Preference data

The results of post-tasks questionnaires were collated and analysed using descriptive statistics. The central tendencies (mean and median) of the rating given for each statement is shown in Table 40. The frequency of rating numbers for each statement is also visualised in Figure 47. Statement 2 ("*The tool is easy to use*") received the highest rating for all central tendency measures and all participants agreed or strongly agreed with this statement. Statement 1 ("*The interface of the tool is easy to navigate"*) had the second-highest mean (4.3) with 6 participants strongly agreed and another 6 participants agreed with this statement. The majority of participants also agreed or strongly agreed with the utility of the tool (Statement 4 and 5) with a slightly lower mean. Most participants also agreed or strongly agreed that they would recommend the tool (S6). Comparatively, the statement with the poorest ratings was Statement 3 ("*I can mostly understand what I need to do to complete the analysis"*). It had the lowest mean and the least number of participants selecting "strongly agree" comparing to other statements.

Table 40. Central tendencies (mean and median) of the ratings on a Likert scale regarding the ease to use and the usefulness of the tool. (n=14, includes the pilot trial)

Label	Statements	Mean	Median
S1	*The interface of the tool is easy to navigate*	4.3	4
S2	*The tool is easy to use*	4.6	5
S3	*I can mostly understand what I need to do to complete the analysis*	3.9	4
S4	*I found the results and graphs useful*	4.2	4
S5	*The tool would be useful for comparing wastewater asset options and inform investment decision*	4.1	4
S6	*I would recommend this decision support tool to other colleagues working with wastewater asset planning*	4.1	4

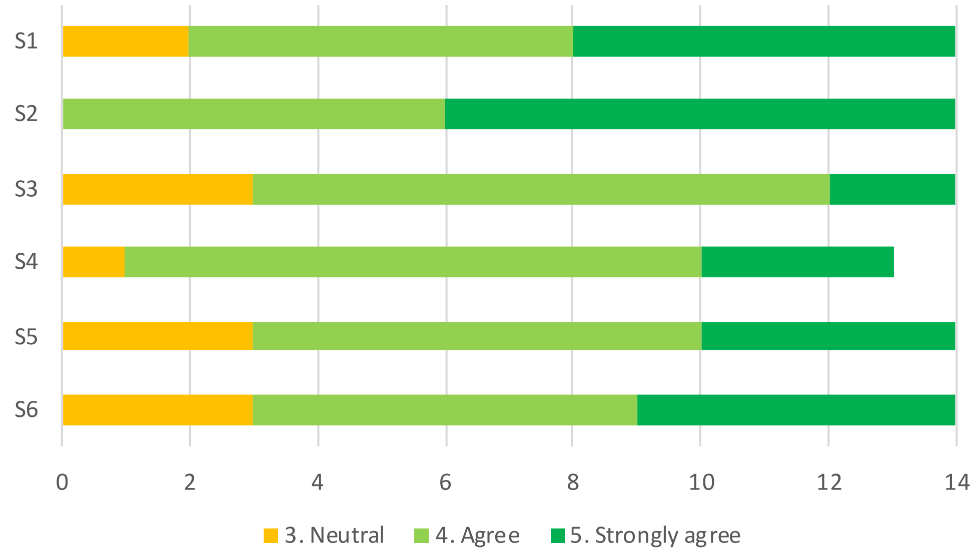


Figure 47. The frequency (counts) of each rating scale number for each statement (n=14). The scale number 1 and 2 (i.e. 'strongly disagree' and 'disagree') were not selected by any participant. Q4 has one less input as the participant in the pilot trial did not rate on this statement due to a critical issue with the graph.

Table 41. Key highlights on the strengths, concerns and areas of improvement mentioned in the post-tasks debriefing.

<table>
<tr><th colspan="2">Strength</th></tr>
<tr><td colspan="2">Overall ease of use and navigation
Graphs were useful
The way options were ranked was interesting and useful for comparing options</td></tr>
<tr><th>Areas of concerns and problems</th><th>Areas of improvement suggested</th></tr>
<tr><td>The potential issue with copying and pasting information data and confusion with ‘Match Destination Formatting’
Data availability and collection for the use of this tool
Lack of guidance on providing rating and scoring
Limited experience with testing this tool</td><td>Provide guidance for how to rate qualitative indicators and weights
Provide clear definitions of indicators
Provide additional ranking for individual sustainability criteria
Potentially develop a fixed weighting profile for some indicators as a baseline</td></tr>
</table>

To understand the reasons behind the ratings, the debriefing sessions were transcribed and coded to three pre-defined codes: strength, concerns or problems, and suggestion improvement.

summarises the details in each aspect. First, many participants commented that they found the tool easy to use and navigate. Some commented they found the graphs useful and interesting for comparing options. In particular, they found the second graph more useful because it dissected the composite scores of options in the first graph into the performance of individual sustainability criteria (see Figure 44). The graph provides more detail and insight than the single aggregated scores.

There were some concerns raised in the debriefings. One issue was the 'copying and pasting' action when filling in the input sheets. Some participants mentioned they were hesitant with whether directly copying and pasting information from the data reference document was allowed. Some were not sure what it meant by 'matching destination formatting'. This provides a reasonable explanation for the high number of errors observed in the project information sheet (Figure 46). Another concern was the source and availability of the data used for the assessment. Although many commented positively on the ease of use of the tool, they also enquired about the source of the data and the requirement of time and resources that would be needed realistically for the primary data collection process. For example, one participant mentioned it could take up to two weeks to acquire the information on Capex from the technology supplier. This concern was acknowledged in the design of the usability testing as it was not considered feasible to include the data collection process in the usability testing session. Some participants also expressed concern over the lack of guidance on how to rate the qualitative indicators and weighting. They argued that ratings by different users can be subjective and a consistent approach should be implemented. Moreover, some participants were hesitant to agree on the overall usability of the tool (e.g. those selected "neutral" for S5 and S6 in Figure 47) because their experience of using the assessment tool was limited to only one case study.

In terms of the areas of improvement, many participants suggested that a greater level of guidance was recommended for providing ratings for the qualitative indicators and weightings. Alternatively, an interesting suggestion was to develop a fixed weighting profile as a baseline for some indicators so the weights are more consistent and less subjective. This should also be supported by a clearer definition of each indicator. Although many commented on the usefulness of the second graph in the result section, some also suggested including the ranking of options based on individual sustainability criteria (i.e. Environmental, Social, Economic and Technical) so it would be easier to compare options.

Table 42. Option rankings derived from 8 usability tests where case A (i.e. pilot study 2 in chapter 7) was applied to the assessment tool. The description of each option can be found in Table 23 in chapter 7. The best option is highlighted in green, the worst in red.

(Case A)	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8
Baseline	1	2	3	4	5	7	6	8
Participant 1	1	2	3	5	4	7	6	8
Participant 2	1	2	3	4	5	7	6	8
Participant 3	1	2	3	5	5	7	6	8
Participant 4	1	2	3	5	4	7	6	8
Participant 5	1	2	3	4	5	7	6	8
Participant 6	1	2	3	4	5	7	6	8
Participant 7	1	3	2	5	4	7	6	8
Participant 8	1	2	3	4	5	7	6	8

Table 43. Option rankings derived from 8 usability tests where case B (i.e. pilot study 1 in chapter 6) was applied to the assessment tool. The description of each option can be found in Figure 32 in chapter 6. The best option is highlighted in green, the worst in red.

(Case B)	Option 1 (ASP)	Option 2 (DAF)	Option 3 (CAPS)	Option 4 (B-ASP)	Option 5 (SBR)	Option 6 (G-ASP)	Option 7 (De-ammo)
Baseline	1	7	4	3	5	6	2
Participant 9	1	7	6	2	5	3	4
Participant 10	1	7	3	5	4	6	2
Participant 11	1	7	5	2	6	4	3
Participant 12	1	7	4	3	5	6	2
Participant 13	1	7	4	3	5	6	2

8.2.3.3 Weighting and option ranking

The option rankings derived from each usability test are shown in Table 42 for case A and Table 43 for case B. They were compared to the baseline rankings which no weights were applied. The results indicate that all participants consistently identified the best and the worst options using the sustainability assessment tool. For case A (Table 42), there were few rank reversals and most of them occurred between option 4 and option 5. Comparatively, there were more rank reversals for case B (Table 43) despite the ranks for the best and the worst option were identical across all participants. This suggests that the weightings developed have the potential to provide consistent score aggregation for selecting the best option.

The usability testing also provided an opportunity to develop a new group weighting profile. Each participant has provided ratings for all indicators and these ratings were normalised into weights. The summary of individual weighting profiles is shown in **Appendix 10**. The individual weightings were then aggregated using the geometric mean into a group weighting (Table 44). The result showed that Pollutant removal potential had the highest weight value (0.129) followed by Reliability (0.122). This is also consistent with the results of the previous weighting development (Chapter 6) where Compliance (which combined Pollution removal and Reliability) was considered the most important indicator. Similar to group weighting developed in Chapter 6, Public value added received the lowest weight. In terms of the weighting method, the new group weighting was developed by the direct rating method instead of AHP. It seems two weighting methods can produce very similar ranking preferences, albeit with different weight values. In terms of ease to use, the direct rating method was easier and flexible to use than AHP. The group weighting profile in Table 44 will be applied as a baseline weighting profile to sustainability assessments in the future.

Table 44. Group weighting of assessment indicators aggregated from individual weighting and their ranking.

	Group weighting	Rank
Net carbon emission	0.105	6
Pollutant removal potentials	0.129	1
Biodiversity net gain	0.058	9
Public value added	0.057	10
Odour	0.081	8
Ease to operate and maintain	0.114	3
Flexibility (Adaptability)	0.088	7
Reliability	0.122	2
Opex	0.111	4
Capex	0.108	5

8.3 Discussion

8.3.1 Usability heuristics and evaluation

The 10 usability heuristics provided good guidance for designing the user interface of this tool. The results of usability testing do not reveal major usability flaws relevant to those heuristics but identified areas for improvement such as providing more detailed guidance on rating and definitions. While the ten usability heuristics principles by Nielsen (1994) have widely applied to various domains of studies, the disadvantage is that they can be too general to apply to studies with specific-domain usability context, which resulted in attempts to adapt and extend those usability heuristics (Jimenez *et al.*, 2016). From the experience of building and testing this assessment tool, some case-specific usability heuristics were highlighted in addition to Nielsen's usability heuristics. For example, the issue with 'copying-and-pasting' information to the worksheet suggests that the level of familiarity with the software can be inconsistent across users. Therefore, it should not be assumed that every user has the same level of understanding of all the functions in the tool. A good usability heuristic would be to provide easily understandable instruction and function and detailed explanation should be provided. This is also partially linked to the second ("*Match between system and the real world*") and tenth heuristics ("*Help and documentation*") shown in Table 35. The importance of providing sufficient guidance for users has also been mentioned in other studies (Krovvidy *et al.*, 1991; Heller *et al.*, 1998; Hamouda *et al.*, 2009). Another usability heuristic is to allow transparency in the sustainability assessment tool. The observation from the usability testing was that some users would request the source of data and enquire about the methods for deriving the results. This is necessary because users may intend to validate the logic and reasoning behind the selection of the best alternative. The need for transparency when developing and using sustainability assessment tools has also been mentioned in other studies (Cappuyns, 2016; Huysegoms and Cappuyns, 2017). Another heuristic is to use automatic computation, which often co-exists with a "black-box" design (i.e. calculation hidden from users). When a complex or unfamiliar assessment methodology is selected for the assessment tool, the automation would ease the operation of the tool and reduce the cognitive requirement of the user. The results of usability testing in this study suggest that despite users do not have prior experience of using the MCDA model, they could complete the assessment within a short timeframe using the built interface. However, this should be complemented by the former heuristic on transparency so the option to review the computation process remains available to users.

This research also highlights the significance of usability testing in the development process of the sustainability assessment tool. The evaluation of usability should be an inseparable part of the product design process because usability is an important attribute of the quality of the 'product' (Jimenez *et al.*, 2016). Hamouda *et al.* (2009) Also mentioned that the verification and validation of a decision support product is an important step in its development. The quality of the tool interface directly influences its usability and the ability of the users to communicate with it. Therefore, it is recommended to integrate the methodology of usability testing as part of the development process and the implementation of the assessment tool. Moreover, this research also highlights the suitability of the remote and synchronous testing approach. Although usability testing is often conducted in a usability laboratory, remote usability testing can be as effective as the traditional testing approach (Hartson *et al.*, 1996; Thompson *et al.*, 2004; Andreasen *et al.*, 2007; Madathil and Greenstein, 2011). The use of Microsoft Teams® in this research as a remote testing environment was proven cost-effective, flexible and easy to set up. Specifically, the remote setting offers convenience to participants and significantly reduces budget and time requirements (Dray and Siegel, 2004; Bastien, 2010; Madathil and Greenstein, 2011; Nelson and Stavrou, 2011). Microsoft Teams® also allows screen sharing and video capturing functions, which are often the requirement for remote usability testing (Andreasen *et al.*, 2007). The practical application of Microsoft Teams® in this research validated its suitability to perform remote usability testing internally in a corporate environment. Additionally, synchronous testing was considered more suitable than asynchronous testing because it could capture real-time and qualitative feedback such as in post-tasks debriefings (Thompson *et al.*, 2004; Bastien, 2010).

8.2.2 Reflection on other observational data

The usability testing also revealed some interesting observations regarding the behavioural aspects when using the tool. When the results were presented to the users by the tool, many of them attempted to review the graph and validate the results. This included, for example:

- Cross-checking the data and information between the graphical and tabular results.

- Reviewing the weighting calculation and the input provided by the users. Some users commented that it would be useful to understand how the rating scale affected the weighting allocated. Some users also tweaked the input slightly to test the sensitivity of weighting on the final ranking of technology alternatives.
- Reviewing and validating the performance rating sheet. When users did not entirely agree with the results or found it interesting, it was observed that some users decided to review the performance ratings of each technology alternative and tried to identify which performance indicators have contributed to the final ranking. Occasionally, users realised they no longer agreed with the input and decided to make changes to the performance ratings.

Additionally, the time spent on reviewing results also varies greatly across users (Figure 45). It was observed that some users showed greater engagement with the results of the tool than others. This may be affected by how the users perceived the results differently and the cognitive capacity of individuals when using the tool. Those observations highlight potential beneficial uses of the assessment tool. First, the tool has the potential to lead to deeper inquiries into the user's preference model and identify potential biases. The transparency of the tool interface allows the users to review and move between the result and the input sections. Additionally, the tool would also be useful when multiple users (in a group) are involved in using the assessment tool simultaneously to collectively reveal, share and discuss their personal preference, which would help reduce personal biases and make sound decisions. The observations also suggest potential challenges when using the tool. Despite the flexibility for users to review and adjust their inputs, there is a risk that users may intentionally create a profile of option ranking primarily based on personal preference and expectation, especially if they do not agree with the results of the tool. Better and additional guidance is needed to help the users interpret and use the tool more consistently, which should be included in the continuous testing and improvement of the tool.

8.3.3 Further improvement of the tool

Overall, the results of usability testing indicate that the assessment tool is easy to use and operate by new users. The tool is useful for comparing multiple options and for providing visualisation of the results. There were no major design flaws identified from the usability testing.

To address the areas of improvement suggested in the testing, a detailed user handbook will be created to include specific instructions for each step of using the assessment tool. The handbook also includes guidelines to aid the user to collect relevant project information and data for indicators. The handbook will also provide guidelines for rating qualitative indicators and their weighting. For the purpose of transparency, the underlying MCDA methodology will be explained in detail in the handbook so the tool can be edited and modified by the end-users in the future. This handbook and the sustainability assessment tool (Excel file) are the major research deliverables for the water company partner in this research. Due to business confidentiality, the handbook is not included in this thesis.

Chapter 9 Research conclusions

9.1 General discussion

9.1.1 The Multi-Criteria Decision Support Tool (MCDST)

The features of the assessment tool conform to the key ingredients when building a Multi-Criteria Decision Support Tool (MCDST) (Jelassi *et al.*, 1985). Firstly, the tool can deal with multiple and conflicting criteria at the same time. The adoption of SAW provides a simple and effective approach to aggregate multiple sustainability criteria and indicators when evaluating wastewater treatment processes. The tool aggregates the performance of these indicators into a sustainability score while retaining the score in each sustainability criterion (Environmental, Social, Economic and Technical). The Linear-Sum normalisation technique homogenises the value scales of different indicators into an unform scale. Secondly, the tool develops weightings for assessment indicators. The tool applies the direct rating method to develop weights of indicators for individual stakeholders and aggregates them into a group weighting profile by using the geometric mean method. The performances of indicators and weightings are aggregated into a composite sustainability score using the SAW model. Composite scores are used to compare and rank wastewater treatment processes to determine the most desirable option. Thirdly, the decision support tool should also provide for an interactive process with decision-makers. In this research, a physical interface was built in Excel® to support and streamline the assessment process. The findings of usability testing suggest that users can perform the analysis using the built-in instructions with little or no prior experience and knowledge of MCDA. Additionally, the tool provides a learning opportunity for the users to understand the multi-criteria nature of the problem and discuss any inconsistency between their perception and the results of the tool. For example, during the usability testing, some participants appeared intrigued by the option ranking and attempted to review the calculation in the background. The option to review the scores in an individual sustainability criterion offers accessible insights into the strength and weakness of each wastewater treatment option. Decision-makers can potentially capitalise on this information not only to identify the best alternative but make strategic improvements. This echoes the popular recognition in the literature that the core value of MCDA is to help decision-makers understand and analyse the complexity of the decision (Saaty, 1980; Zanakis *et al.*, 1998; Communities and Local Government, 2009; Cinelli *et al.*, 2014).

Overall, the experience in this research also indicates a potential synergy between the disciplines of sustainability assessment, MCDA and decision support systems. The salient benefit of MCDA is to provide a coherent and logical process for dealing with multiple and conflicting decision criteria, which can help solve complex and multi-disciplinary sustainability problems. The practice of using MCDA to aid sustainability assessment has been widely acknowledged and discussed (Gasparatos *et al.*, 2008; Lai *et al.*, 2008; Cinelli *et al.*, 2014; De Feo *et al.*, 2018). This research reiterates the strength of MCDA for conducting sustainability assessment and further highlights other desirable characteristics that enable the implementation of MCDA as a decision support tool. A good decision support tool requires a user-centred design to enable interactions and communications with users and effective communication is an important element when performing sustainability assessments (Waas *et al.*, 2014). This was reflected by the continuous involvement of company stakeholders in the development process of the assessment tool.

9.1.2 Usability

The output supports decision-makers to rank and select the best treatment option in a systematic and coherent manner. The adoption of SAW and the direct rating weighting method in the MCDA assessment enabled a simple yet efficient process to integrate multiple sustainability criteria. Although AHP was used formerly for developing weights, it was not considered feasible when there were many indicators to be compared or the number of indicator changes. The direct rating method offered a usable alternative to develop weights that lead to consistent option ranking (as shown in results in Chapter 8). The results of usability testing suggest that the tool is easy to use and useful, which validated the benefit of the assessment tool. Overall, the participants in usability testing provided positive feedback based on their experience with no major design flaws identified. The areas for improvement include the provision of more detailed instructions and guidelines on how to provide ratings for qualitative indicators and their weightings for the users.

9.1.3 Assessment methodology

Besides the usability aspect of the tool, the success of the MCDST tool also hinges on its robustness. The selection of specific assessment methods or models has undergone rounds of literature reviews to identify the current state of knowledge and was followed by pilot studies for testing and validation. The pilot studies and usability testing suggest that the assessment methodology provides consistent results of weighting and option rankings. Although AHP was previously selected as the weighting method in MCDA, the sensitivity analysis in the first pilot study suggests that rank reversals can still occur between the top two options (Table 21 in Chapter 6). Another issue with AHP was that it cannot be easily applied to other cases where the number of indicators or technologies may change. Due to the fundamental operation of pairwise comparisons in AHP, its result is very sensitive to changes in the criteria structure. Therefore, AHP is not suitable for this assessment tool given the vast number of STWs and their site-specific characteristics in the water company. Despite that the comparative study in Chapter 7 showed that the 4 MCDA methods led to very similar option ranking, SAW was the most feasible method due to its ease of use.

9.2 Contribution

The physical deliverables of this research include the production of the sustainability assessment tool (i.e. the Excel® spreadsheet file) and a user handbook. These two files were handed over to the relevant department in the water company at the end of the research project. The novelties of this research are presented and discussed in the following section.

9.2.1. MCDST to inform technological selection

The research culminated in the development of a new MCDST that evaluates wastewater treatment processes or technologies from a sustainability perspective for the water company. There have been few DSTs focusing on sustainability as part of decision-making in the water company and this tool offers an alternative way to inform wastewater asset decisions. Although there are similar MCDSTs built for water and wastewater management in other studies (Adewumi *et al.*, 2013; Castillo *et al.*, 2016; Kalbar *et al.*, 2016; Sadr *et al.*, 2018), the development of this tool has been tailored to the context of the water company addressing its specific challenges and needs with a unique tool interface. The tool is expected to be used in future decisions when comparing potential wastewater treatment processes for a STW. The tool provides a user-friendly and coherent process for decision-makers to compare them and identify the best option from a holistic sustainability perspective.

The tool also provides managerial insights, through a transparent process from the inputs to the results. As suggested from the usability testing, the result of the tool was able to lead to deeper inquiries to understand and validate the rankings between different wastewater technologies options. Particularly, the graph which shows both the aggregated composite scores and the individual sustainability pillars was perceived as useful by the users because they can quickly identify the strengths and weaknesses of each alternative. The result section also allows decision-makers to reveal their underlying preference (reflected by their performance ratings and weighting allocation) and discuss other users to minimise personal bias and resolve their difference. This reiterates one of the strengths of a MCDA approach, which is to catalyse dialogues among decision-makers and enhance mutual understanding when solving a complex decision problem (Saaty, 1980; Zanakis *et al.*, 1998; Communities and Local Government, 2009; Cinelli *et al.*, 2014).

9.2.2 A practical framework for developing a MCDST

The experience of the research offers practical insights into the process of developing a MCDST that accounts effectively for sustainability aspects of decision-making in the water company. A generalised framework is proposed based on the procedures deployed in this research (Figure 48). The framework serves to guide the future development of MCDST particularly in a corporate environment. The framework combines key steps in MCDA and additional user-research. Specifically, "*understand the decision context*" is the first step to explore and learn about the context where the needs of sustainability assessment arise. The acquisition of this knowledge subsequently informs the selection of assessment indicators and potentially a suitable MCDA model, with information drawn from the literature review. Once the assessment methodology is proposed and validated, it is embedded into a DST with a physical interface that enables interaction with the users. The final validation of the tool is to evaluate the tool by testing it with real users. The framework also incorporates stakeholder engagement during the development process enabled by an 'immersive' and interactive research process. Qualitative data collection methods such as interviews were deployed at various stages of the development to provide insights and evidence for making decisions about the methodological design. They also served to build and maintain rapport with company stakeholders throughout the research.

This framework shows a certain degree of resemblance to multi-criteria based assessment frameworks from other studies (Guitouni and Martel, 1998; Ashley *et al.*, 2003; Lai *et al.*, 2008). It also partially mirrors the general prototyping framework for building a DST mentioned in Power, 2002, p. 63 but with extended actions that also accommodate the requirement of MCDA. The similarities also include the practice of qualitative research (Decision mapping) to first develop an understanding of current decisions and initiate stakeholder engagement. This also refers to the 'Information-structuring' challenge proposed by Hugé *et al.*, (2011) and Waas *et al.*, (2014). The outputs of decision mapping were used to inform the subsequent selection of criteria and indicators (another example see Ashley *et al.*, 2003). However, the framework in this research (Figure 48) highlights the inclusion of building a user interface for communication as part of a user-centred design of MCDA. There is the novelty of including the element of usability testing to evaluate the 'communication' aspect of the sustainability assessment tool, which integrates the requirement of a good decision support tool. This would be applicable and useful for future applications where an 'in-house' sustainability assessment tool needs to be created to support repeated decision-making in a company.

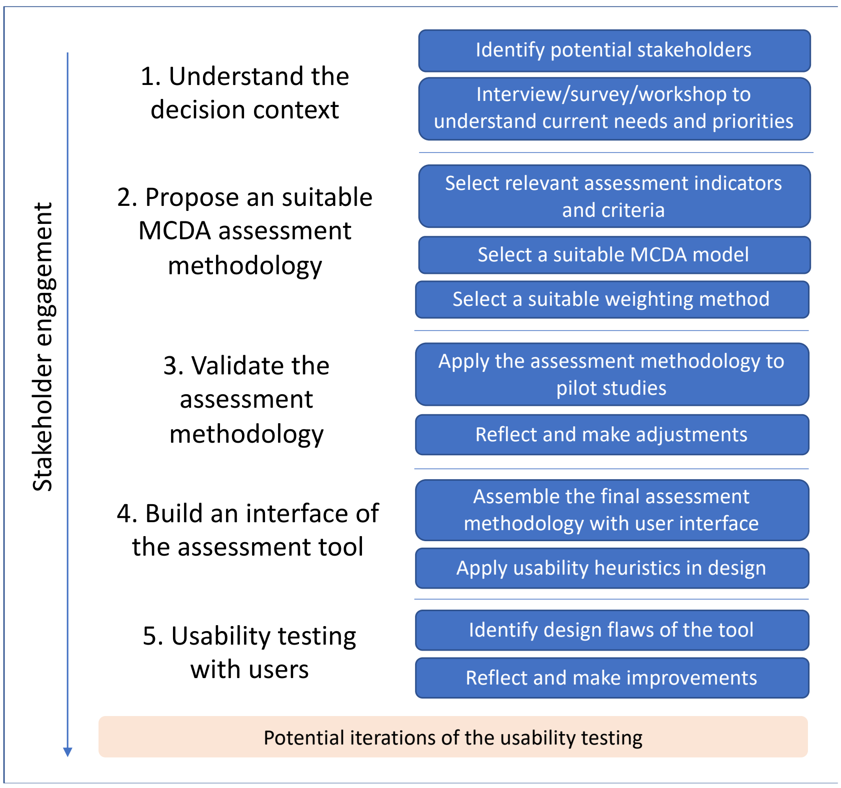


Figure 48. A generalised framework for developing a MCDST in a corporate environment.

9.2.3. Practical insights when developing a MCDST

Additionally, the experience in this research highlights that feasibility is an important factor when designing a sustainability assessment tool. The availability of resources such as time and expertise greatly influence the suitability of the methodology that is incorporated into the tool design. For example, although AHP is considered the most popular and widely applied MCDA model in general, it was not practical for this assessment tool because the time required to perform it would increase exponentially with the number of criteria and alternatives. This considerably reduced the practical attractiveness of AHP. Another example was the number of indicators in the criteria hierarchy. There were 9 assessment indicators included in the hierarchy, but the list could potentially be expanded for a more extensive representation of the complexity of the decision problem. However, a greater number of indicators or criteria does not necessarily improve the assessment result and can also increase the burden of cognitive processing (Ling *et al.*, 2021). Muga and Mihelcic (2008) suggest that the number of indicators should be limited in number and easy to handle. Saaty and Ozdemir (2003) also suggests that as the number of criteria increases so does the likelihood of error when making judgements. Therefore, a balance between the complexity and feasibility of the method needs to be carefully calibrated. This requires the researcher to evaluate the feasibility of the assessment method in a 'real-world' setting.

The experience of this research also highlights the importance of stakeholder engagement and participation in the development and the use of the assessment tool. The stakeholder engagement was crucial to establish the decision context where the assessment tool was embedded and to determine the specific designs of the tool that reflect this context, which has been recognised as one of the three challenges in sustainability assessments, namely the "interpretation" challenge (Hugé *et al.*, 2011; Waas *et al.*, 2014). The multiple instances of stakeholder engagement in the development process also created opportunities to build rapport between project stakeholders and the researcher. Such rapport provided additional resources and information-sharing which, in turn, facilitated subsequent research tasks such as recruitment for AHP questionnaires (for weighting development) and usability testing. The rapport with key stakeholders also served as human resources for making informal consultations and collecting feedback as part of continuous cycles of development and testing of the tool. However, one particular challenge of conducting research development in a large corporation is that the group of key stakeholders may quickly change due to job recruitment changes and organisational reforms. It was difficult to maintain the same group of individuals throughout the research project and it was important to regularly update the list of stakeholders as changes occurred. This suggests that it may be useful to document the organisational changes and investigate how would they impact the direction of the research and subsequent research decisions in a longitudinal study.

9.3 Limitations

9.3.1 Generalisability

There is limited generalisability of the research findings because they are, at least to a degree, context-dependent given the research objectives in this thesis. The specific assessment methodology (MCDA model, assessment indicators and criteria etc.) used in this research may not be directly applicable to other water companies. There are strategies to improve the generalisability such as triangulation or conducting multiple studies (Coghlan and Brydon-Miller, 2014). There were only two pilot studies in this research to test the assessment tool, and one round of usability testing to validate its usability. Thus, one way of improving generalisability is to apply the assessment tool and framework to a greater number of studies (e.g. a variety of STWs projects), a wider decision scope (e.g. water resource management), and even in other organisational contexts (e.g. other water companies and industries). However, testing a tool is a lengthy process that requires iterative and continuous effort. Therefore, future studies will expand on the findings of this research to test the assessment tool and methods in wider contexts.

Although the specific assessment tool and methods may not be directly applicable to other companies, the procedures and insights on how to build a similar MCDST based on sustainability are transferrable to wider contexts and industries. Specifically, the research offers the procedural process for selecting and building a MCDA approach, with a framework to develop tailored assessment indicators (in the scope of wastewater technological selection) for its operation. The research also encouraged the adoption of social research methods (such as interviews and usability research) to tailor sustainability assessment to the organisational context and support stakeholder participation in the development process. In terms of the output of the tool, the results of MCDA with a composite indicator approach can offer useful insights into the overall ranking and suitability of alternatives but also retain granularity of the individual sustainability 'pillars' for making strategic recommendations. The decision-makers can utilise this information to determine the best course of action as well as improve the performance of existing assets (e.g. proposing energy reduction measures to improve the carbon emission profile). Besides the technical decision support from using the MCDA, the research highlights the 'learning' benefits as a result of its development and use, which lead to more meaningful discussions and actions among decision-makers (Waas *et al.*, 2014; Sala *et al.*, 2015).

Despite the limited generalisability of the research outcomes, this research initiated a robust first step towards better understanding and practices of developing a multi-criteria sustainability assessment and decision support tool in a large water corporation. The procedural framework proposed in Figure 48 offers a process that can be replicated, expanded and extended for projects in other water companies and industries and provide practical aid to complex decision-making.

9.3.2 Data availability

One limitation of the research was the availability of data for the assessment indicators. As mentioned, the collection of data by the users for the performance indicators was not required in the usability testing (data was provided from previous internal reports or pre-populated by the researcher). This was to normalise the amount of time and resources required to run the usability testing but may have underestimated the real length of time to use the sustainability assessment tool. In reality, some quantitative data is not always readily available. For instance, the costing data (Opex and Capex) is processed and owned by another department of the company and the acquisition of quotes from asset suppliers could take weeks. Data of other indicators such as pollutant removal potentials can also be difficult to obtain if field measurements need to be taken. To mitigate the limitation of data availability, future studies may investigate how uncertainty modelling such as fuzzy logic can be incorporated into the assessment methodology. Developed by Zadeh (1965), fuzzy logic is particularly useful when there is vague and imprecise data. Fuzzy logic has been applied as an extension to existing MCDA methods in various fields such as sludge management (An *et al.*, 2018) and energy policy-making (Kaya *et al.*, 2019). The Grey System Theory, developed by

JuLong (1982), may also be applied to the assessment methodology for dealing with a lack or incomplete information. Both fuzzy logic and the Grey System Theory were excluded from the current version of the assessment tool due to high technical complexity and requirement for specialised knowledge and time for their operation.

9.3.3 Group and individual decision-making

In this research, stakeholder participation was conducted mostly at the individual level. There were limited opportunities for developing and testing the assessment tool in a group decision-making environment. Attempts were made to recruit a panel of participants as a focus group to understand this interaction aspect. However, it was practically challenging to recruit for focus groups due to clashes of time availability, compounded by the implications of the social distancing condition during the Covid-19 pandemic. As a result, the tool was primarily built for individual users at the local management level in the company (e.g. regional asset managers and planners). As a DST, it aims to inform the decisions by providing additional insights in the early phase of the investment process, rather than directly driving the investment decision. In this water company, the final decision point occurs at the level of the investment board as a group and it is beyond the scope of this DST.

However, once the condition permits, it would be interesting to investigate how effective would MCDA be when applied in a group decision-making environment. It would also be insightful to compare the group weighting of indicators by using mathematical aggregation (i.e. Geometric mean) and reaching a consensus between decision-makers (i.e. the 'sharing' aggregation method mentioned in section 6.2.2.).

9.3.4 Alternative user interface and system

There are potential opportunities for utilising other computer software or systems to operate the MCDA. A comprehensive summary of pre-made software for individual MCDA methods can be found in (Ishizaka and Nemery, 2013). Pre-made software was not used for this research because they have little room for modification to adapt to the specific assessment methods selected for the context. Furthermore, installing third-party software on company devices (e.g. laptops) requires additional approvals due to internal data and software policy. This was a practical challenge when implementing a tool inside a company. Alternatively, the user interface can also be devised using a programming language such as Python® which offer great flexibility (for example see Papathanasiou and Ploskas, 2018). However, this puts a high requirement of programming skills on the researcher and it was not feasible to programme a tool interface within the project timeframe. In the future, it may be useful to utilise the programming capabilities in the company to build an advanced tool interface that is able to simplify the user navigation and enable more sophisticated calculations in the background.

9.4 Conclusions

In summary, this research has developed a multi-criteria sustainability assessment tool that specifically aims to evaluate the sustainability of different wastewater treatment processes for a water company. A suite of sustainability indicators was developed and tailored to the decision context in the water company. SAW and the direct weighting were selected as the MCDA methodology to calculate sustainability composite scores for comparing and ranking potential wastewater treatment options. This was assembled into a user interface in Excel to facilitate user interaction. The results of usability testing suggest that the assessment tool was perceived as useful and easy to use by the users. The tool has the potential to deal with complex investment problems with multiple decision priorities from the sustainability perspective and incorporate users' preferences into decision-making coherently. The research findings highlight the potential of combining sustainability assessment and MCDA to inform wastewater investment decisions and offer practical insights into the development of a decision support tool with a user-centred design. The practical experience in this research recommended that its development process should be tailored to the unique requirements and challenges of the decision context, enabled stakeholder engagement.

9.5 Future studies

The major aspects for future studies focus on: (1) The continuous iteration of the development and testing cycle and optimise the tool with evolving needs; (2) Consolidating the understanding of the robustness and generalisability of the MCDST by applying it to more STWs projects as well as projects in a wider scope; (3) Testing the suitability of alternative MCDA models and extensions to deal with incomplete or imprecise data; (4) More research is also recommended to testing the tentative framework in other contexts to evaluate its applicability and effectiveness. This information will be passed to the project owner in the water company to further optimise the assessment tool and create lasting value.

Bibliography

Adewumi, J. R., Ilemobade, A. A. and Van Zyl, J. E. (2013) 'Application of a multi-criteria decision support tool in assessing the feasibility of implementing treated wastewater reuse', *International Journal of Decision Support System Technology*. doi: 10.4018/jdsst.2013010101.

Ahmed, Y., Gendy, A. and Hagger, S. (2017) 'Sustainability assessment of municipal wastewater treatment.pdf', *International Journal of Environmental and Ecologial Engineering*, 11.

Ainger, C. and Fenner, R. (2014) 'Chapter 2: Key Principles', in *Sustainable Infrastructure : Principles into Practice*, pp. 263–325. doi: 10.1680/sipp.57548.263.

Ainger, C. M. and Fenner, R. A. (2013) *Sustainable Infrastructure: Principles into practice*, *Sustainable Infrastructure: Principles into practice*. doi: 10.1680/sipp.57548.

Akcin, G., Alp, Ö., Gulyas, H. and Büst, B. (2005) 'Characteristic, Analytic and Sampling of Wastewater', *Emwater E-Learning Course*.

Alhojailan, M. I. and Ibrahim, M. (2012) 'Thematic Analysis : A Critical Review of Its Process and Evaluation', *WEI International European AcademicConference Proceedings*.

Allen, C., Metternicht, G. and Wiedmann, T. (2018) 'Initial progress in implementing the Sustainable Development Goals (SDGs): a review of evidence from countries', *Sustainability Science*. doi: 10.1007/s11625-018-0572-3.

El Amine, M., Pailhes, J. and Perry, N. (2014) 'Comparison of different Multiple-criteria decision analysis methods in the context of conceptual design: application to the development of a solar collector structure', in *Proceedings of Joint Conference on Mechanical, Design Engineering & Advanced Manufacturing*.

An, D., Xi, B., Ren, J., Ren, X., Zhang, W., Wang, Y. and Dong, L. (2018) 'Multi-criteria sustainability assessment of urban sludge treatment technologies: Method and case study', *Resources, Conservation and Recycling*. doi: 10.1016/j.resconrec.2016.08.018.

Andreasen, M. S., Nielsen, H. V., Schrøder, S. O. and Stage, J. (2007) 'What happened to remote usability testing?: An empirical study of three methods', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/1240624.1240838.

Ang, F. and Passel, S. Van (2012) 'Beyond the environmentalist's paradox and the debate on weak versus strong sustainability', *BioScience*. doi: 10.1525/bio.2012.62.3.6.

Ashley, R., Blackwood, D., Butler, D., Davies, J., Jowitt, P. and Smith, H. (2003) 'Sustainable decision making for the UK water industry', *Proceedings of the Institution of Civil Engineers - Engineering Sustainability*, 156(1), pp. 41–49. doi: 10.1680/ensu.2003.156.1.41.

Ashley, R., Blackwood, D., Butler, D., Jowitt, P., Davies, J., Smith, H., Gilmour, D. and Oltean-Dumbrava, C. (2008) 'Making Asset Investment Decisions for Wastewater Systems That Include Sustainability', *Journal of Environmental Engineering*. doi: doi:10.1061/(ASCE)0733-9372(2008)134:3(200).

ATKINS (2018) *A framework for the production of Drainage and Wastewater Management Plans*.

ATKINS (2019) *Working together to improve drainage and environmental water quality: An overview of Drainage and Wastewater Management Plans*.

Baauw, E. and Markopoulos, P. (2004) 'A comparison of think-aloud and post-task interview for usability testing with children', in *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community, IDC 2004*. doi: 10.1145/1017833.1017848.

Bal, M., Bryde, D., Fearon, D. and Ochieng, E. (2013) 'Stakeholder Engagement: Achieving Sustainability in the Construction Sector', *Sustainability (Switzerland)*. doi: 10.3390/su5020695.

Balkema, A. J., Preisig, H. A., Otterpohl, R. and Lambert, F. J. D. (2002) 'Indicators for the sustainability assessment of wastewater treatment systems', *Urban Water*, 4(2), pp. 153–161. doi: 10.1016/S1462-0758(02)00014-6.

Bangser, P. M. (1982) 'Inherent Role for Cost-Benefit Analysis in Judicial Review of Agency Decisions: A New Perspective on OSHA Rulemaking, An', *Boston College Environmental Affairs Law Review*.

Barrett, C. B. and Grizzle, R. (1999) 'A holistic approach to sustainability based on pluralism stewardship', *Environmental Ethics*. doi: 10.5840/enviroethics199921139.

Barron, F. H. and Barrett, B. E. (1996) 'Decision quality using ranked attribute weights', *Management Science*. doi: 10.1287/mnsc.42.11.1515.

Basak, I. and Saaty, T. (1993) 'Group decision making using the analytic hierarchy process', *Mathematical and Computer Modelling*, 17(4), pp. 101–109. doi: https://doi.org/10.1016/0895-7177(93)90179-3.

Basiago, A. D. (1998) 'Economic, social, and environmental sustainability in development theory and urban planning practice', *Environmentalist*. doi: 10.1023/A:1006697118620.

Bastien, J. M. C. (2010) 'Usability testing: a review of some methodological and technical aspects of the method', *International Journal of Medical Informatics*. doi: 10.1016/j.ijmedinf.2008.12.004.

Bebbington, J., Brown, J. and Frame, B. (2007) 'Accounting technologies and sustainability assessment models', *Ecological Economics*. doi: 10.1016/j.ecolecon.2006.10.021.

Belton, V. and Gear, T. (1983) 'On a short-coming of Saaty's method of analytic hierarchies', *Omega*, 11(3), pp. 228–230. doi: 10.1016/0305-0483(83)90047-6.

Belton, V. and Pictet, J. (1997) 'A framework for group decision using a mcda model: Sharing, aggregating or comparing individual information?', *Journal of Decision Systems*. doi: 10.1080/12460125.1997.10511726.

Belton, V. and Stewart, T. J. (2002) *Multiple criteria decision analysis An Integrated Approach*, *Springer-Science+Business Media, B.V.* doi: 10.1007/978-1-4615-1495-4.

Bernasconi, M., Choirat, C. and Seri, R. (2010) 'The Analytic Hierarchy Process and the Theory of Measurement', *Management Science*. doi: 10.1287/mnsc.1090.1123.

Biely, K., Maes, D. and Van Passel, S. (2018) 'The idea of weak sustainability is illegitimate', *Environment, Development and Sustainability*. doi: 10.1007/s10668-016-9878-4.

Bond, A. J. and Morrison-Saunders, A. (2011) 'Re-evaluating Sustainability Assessment: Aligning the vision and the practice', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2010.01.007.

Bond, A., Morrison-Saunders, A. and Howitt, R. (2012) *Sustainability assessment: Pluralism, practice and progress*, *Sustainability Assessment: Pluralism, Practice and Progress*. doi: 10.4324/9780203112625.

Bond, A., Morrison-Saunders, A. and Pope, J. (2012) 'Sustainability assessment: The state of the art', *Impact Assessment and Project Appraisal*. doi: 10.1080/14615517.2012.661974.

Bottero, M., Comino, E. and Riggio, V. (2011) 'Application of the Analytic Hierarchy Process and the Analytic Network Process for the assessment of different wastewater treatment systems', *Environmental Modelling and Software*. doi: 10.1016/j.envsoft.2011.04.002.

Bowen, G. A. (2009) 'Document analysis as a qualitative research method', *Qualitative Research Journal*. doi: 10.3316/QRJ0902027.

Bradley, B. R., Daigger, G. T., Rubin, R. and Tchobanoglous, G. (2002) 'Evaluation of onsite wastewater treatment technologies using sustainable development criteria', *Clean Technologies and Environmental Policy*. doi: 10.1007/s10098-001-0130-y.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*. doi: 10.1191/1478088706qp063oa.

Braun, V., Clarke, V., Hayfield, N. and Terry, G. (2019) 'Thematic analysis', in *Handbook of Research Methods in Health Social Sciences*. doi: 10.1007/978-981-10-5251-4_103.

Browne, K. (2005) 'Snowball sampling: Using social networks to research non-heterosexual women', *International Journal of Social Research Methodology: Theory and Practice*. doi: 10.1080/1364557032000081663.

Brugnach, M., Dewulf, A., Pahl-Wostl, C. and Taillieu, T. (2008) 'Toward a relational concept of uncertainty: About knowing too little, knowing too differently, and accepting not to know', *Ecology and Society*. doi: 10.5751/ES-02616-130230.

Brundtland, G. H. (1987) *Our Common Future (The Brundtland Report)*, *Report of the World Commission on Environment and Development: Our Common Future (The Brundtland Report)*. doi: 10.1080/07488008808408783.

Bryman, A. (2016) *Social Research Methods*, *Oxford University Press*.

Bunce, J. T., Ndam, E., Ofiteru, I. D., Moore, A. and Graham, D. W. (2018) 'A review of phosphorus removal technologies and their applicability to small-scale domestic wastewater treatment systems', *Frontiers in Environmental Science*. doi: 10.3389/fenvs.2018.00008.

Burton, D. (2012) 'The Use of Case Studies in Social Science Research', in *Research Training for Social Scientists*. doi: 10.4135/9780857028051.d24.

Cappuyns, V. (2016) 'Inclusion of social indicators in decision support tools for the selection of sustainable site remediation options', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2016.07.035.

Caprani, L. (2016) 'Five ways the sustainable development goals are better than the millennium development goals and why every educationalist should care', *Management in Education*. doi: 10.1177/0892020616653464.

Carius, A., Ivleva, D., Pohl, B., Rüttinger, L. and ... (2018) 'A foreign policy perspective on the SDGs', *Berlin: adelphi …*.

Castillo, A., Porro, J., Garrido-Baserba, M., Rosso, D., Renzi, D., Fatone, F., Gómez, V., Comas, J. and Poch, M. (2016) 'Validation of a decision support tool for wastewater treatment selection', *Journal of Environmental Management*, 184, pp. 409–418. doi: 10.1016/j.jenvman.2016.09.087.

Chakraborty, S. and Yeh, C. H. (2009) 'A simulation comparison of normalization procedures for TOPSIS', in *2009 International Conference on Computers and Industrial Engineering, CIE 2009*. doi: 10.1109/iccie.2009.5223811.

Chen, R. and Wang, X. C. (2009) 'Cost-benefit evaluation of a decentralized water system for wastewater reuse and environmental protection', *Water Science and Technology*. doi: 10.2166/wst.2009.156.

Cinelli, M., Coles, S. R. and Kirwan, K. (2014) 'Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment', *Ecological Indicators*, pp. 138–148. doi: 10.1016/j.ecolind.2014.06.011.

Coghlan, D. and Brydon-Miller, M. (2014) *The SAGE Encyclopedia of Action Research*, *The SAGE Encyclopedia of Action Research*. doi: 10.4135/9781446294406.

Cole, M. A. (1999) 'Limits to growth, sustainable development and environmental Kuznets curves: An examination of the environmental impact of economic development', *Sustainable Development*. doi: 10.1002/(SICI)1099-1719(199905)7:2<87::AID-SD102>3.0.CO;2-5.

Communities and Local Government (2009) *Multi-criteria analysis : a manual*, *Appraisal*. Available at: http://regulatoryreform.com/wp-content/uploads/2015/02/UK-Multi-criteria-analysis-a-manual-2009.pdf.

Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., Lancelot, C. and Likens, G. E. (2009) 'Ecology - Controlling eutrophication: Nitrogen and phosphorus', *Science*. doi: 10.1126/science.1167755.

Corominas, L., Byrne, D. M., Guest, J. S., Hospido, A., Roux, P., Shaw, A. and Short, M. D. (2020) 'The application of life cycle assessment (LCA) to wastewater treatment: A best practice guide and critical review', *Water Research*. doi: 10.1016/j.watres.2020.116058.

Corominas, L., Foley, J., Guest, J. S., Hospido, A., Larsen, H. F., Morera, S. and Shaw, A. (2013) 'Life cycle assessment applied to wastewater treatment: State of the art', *Water Research*. doi: 10.1016/j.watres.2013.06.049.

Cossio, C., Norrman, J., McConville, J., Mercado, A. and Rauch, S. (2020) 'Indicators for sustainability assessment of small-scale wastewater treatment plants in low and lower-middle income countries', *Environmental and Sustainability Indicators*, 6, p. 100028. doi: https://doi.org/10.1016/j.indic.2020.100028.

Creswell, J., & Plano Clark, V. (2009) 'Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage', *Organizational Research Methods*.

Creswell, J. W. (2009) *Research Design: Qualitative, Quantitative and Mixed Approaches (3rd Edition)*, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. doi: 10.2307/1523157.

Cronin, A., Alexander, V. D., Fielding, J., Moran-Ellis, J. and Thomas, H. (2012) 'The Analytic Integration of Qualitative Data Sources', in *The SAGE Handbook of Social Research Methods*. doi: 10.4135/9781446212165.n34.

Crozier, G., Denzin, N. and Lincoln, Y. (1994) 'Handbook of Qualitative Research', *British Journal of Educational Studies*. doi: 10.2307/3121684.

Currivan, D. B. and Gilbert, N. (2008) 'Researching Social Life', *Teaching Sociology*. doi: 10.2307/1318574.

Davison, R. and French, S. (1987) 'Decision Theory: An Introduction to the Mathematics of Rationality', *The Mathematical Gazette*. doi: 10.2307/3616789.

DEFRA (1999) 'A better quality of life - strategy for sustainable development for the United Kingdom', *One Future - Different Paths*.

DEFRA (2009) 'Guidance on how to measure and report your greenhouse gas emissions', *DECC Website*.

DEFRA (2013) *Sustainable Development Indicators*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/223992/0_SDIs_final__2_.pdf.

Denzin, N. K. and Lincoln, Y. S. (2000) 'Introduction: The discipline and practice of qualitative research', in *Handbook of qualitative research (2nd edition)*.

Department for Business Energy and Industrial Strategy (2020) *Greenhouse gas reporting: conversion factors 2020*. Available at: https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2020 (Accessed: 13 August 2021).

Department for Education (2015) *2010 to 2015 government policy: young people: Policy paper*.

Devuyst, D. (2016) 'Introduction to Sustainability Assessment at the Local Level', in *How Green Is the City?* doi: 10.7312/devu11802-intro.

Dixon, J. A. and Fallon, L. A. (1989) 'The concept of sustainability: Origins, extensions, and usefulness for policy', *Society and Natural Resources*, 2(1), pp. 73–84. doi: 10.1080/08941928909380675.

Djukic, M., Jovanoski, I., Ivanovic, O. M., Lazic, M. and Bodroza, D. (2016) 'Cost-benefit analysis of an infrastructure project and a cost-reflective tariff: A case study for investment in wastewater treatment plant in Serbia', *Renewable and Sustainable Energy Reviews*. doi: 10.1016/j.rser.2016.01.050.

Dray, S. and Siegel, D. (2004) 'Remote possibilities?: international usability testing at a distance', *interactions*. doi: 10.1145/971258.971264.

Dyer, J. S. (2008) 'Remarks on the Analytic Hierarchy Process', *Management Science*, 36(3), pp. 249–258. doi: 10.1287/mnsc.36.3.249.

Easterly, W. (2015) 'The SDGs Should Stand for Senseless, Dreamy, Garbled', *Foreign Policy*.

Edwards, W. (1977) 'How to Use Multiattribute Utility Measurement for Social Decisionmaking', *IEEE Transactions on Systems, Man and Cybernetics*, 7(5), pp. 326–340. doi: 10.1109/TSMC.1977.4309720.

Edwards, W. and Barron, F. H. (1994) 'Smarts and smarter: Improved simple methods for multiattribute utility measurement', *Organizational Behavior and Human Decision Processes*. doi: 10.1006/obhd.1994.1087.

Ekins, P. (1993) '"Limits to growth" and "sustainable development": grappling with ecological realities', *Ecological Economics*. doi: 10.1016/0921-8009(93)90062-B.

Ekins, P., Hillman, M. and Hutchison, R. (1992) *Wealth beyond measure : an atlas of new economics*, *The Gaia future series*. doi: 10.1002/eet.3320020108.

Ekins, P., Simon, S., Deutsch, L., Folke, C. and De Groot, R. (2003) 'A framework for the practical application of the concepts of critical natural capital and strong sustainability', *Ecological Economics*. doi: 10.1016/S0921-8009(02)00272-0.

Elkington, J., Emerson, J. and Beloe, S. (2006) 'The Value Palette: A Tool for Full Strategy Spectrum', *California Management Review*.

Erlingsson, C. and Brysiewicz, P. (2017) 'A hands-on guide to doing content analysis', *African Journal of Emergency Medicine*. doi: 10.1016/j.afjem.2017.08.001.

European Commission (1991) 'Urban Waste Water Treatment Directive', *Official Journal of the European Union*.

Feilzer, M. Y. (2010) 'Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm', *Journal of Mixed Methods Research*. doi: 10.1177/1558689809349691.

De Feo, G., De Gisi, S., De Vita, S. and Notarnicola, M. (2018) 'Sustainability assessment of alternative end-uses for disused areas based on multi-criteria decision-making method', *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2018.03.016.

Figueira, J. R., Greco, S., Roy, B. and Słowiński, R. (2010) ‘ELECTRE Methods: Main Features and Recent Developments’, in. doi: 10.1007/978-3-540-92828-7_3.

Figueira, J. R., Mousseau, V. and Roy, B. (2016) ‘ELECTRE methods’, *International Series in Operations Research and Management Science*. doi: 10.1007/978-1-4939-3094-4_5.

Fishburn, P. C. and Lavalle, I. H. (1999) ‘MCDA: Theory, practice and the future’, *Journal of Multi-Criteria Decision Analysis*. doi: 10.1002/(SICI)1099-1360(199901)8:1<1::AID-MCDA205>3.0.CO;2-E.

Forman, E. H. and Gass, S. I. (2001) ‘The analytic hierarchy process - An exposition’, *Operations Research*. doi: 10.1287/opre.49.4.469.11231.

Forman, E. and Peniwati, K. (1998) ‘Aggregating individual judgments and priorities with the Analytic Hierarchy Process’, *European Journal of Operational Research*. doi: 10.1016/S0377-2217(97)00244-0.

Foxon, T. J., McIlkenny, G., Gilmour, D., Oltean-Dumbrava, C., Souter, N., Ashley, R., Butler, D., Pearson, P., Jowitt, P. and Moir, J. (2002) ‘Sustainability criteria for decision support in the UK water industry’, *Journal of Environmental Planning and Management*. doi: 10.1080/09640560220116341.

Fratini, C. F., Geldof, G. D., Kluck, J. and Mikkelsen, P. S. (2012) ‘Three Points Approach (3PA) for urban flood risk management: A tool to support climate change adaptation through transdisciplinarity and multifunctionality’, *Urban Water Journal*. doi: 10.1080/1573062X.2012.668913.

French, S. and Roy, B. (1997) ‘Multicriteria Methodology for Decision Aiding.’, *The Journal of the Operational Research Society*. doi: 10.2307/3010757.

Fukuda-Parr, S. (2016) ‘From the Millennium Development Goals to the Sustainable Development Goals: shifts in purpose, concept, and politics of global goal setting for development’, *Gender and Development*. doi: 10.1080/13552074.2016.1145895.

Fung, F., Bett, P., Maisey, P., Lowe, J., McSweeney, C., Mitchell, J., Murphy, J., Rostron, J., Sexton, D. and Yamazaki, K. (2018) ‘UKCP18 Factsheet: Precipitation’, *Met Office Hadley Centre, Exeter*.

Funtowicz, S. O. and Ravetz, J. R. (1990) *Uncertainty and Quality in Science for Policy*, *Uncertainty and Quality in Science for Policy*. doi: 10.1007/978-94-009-0621-1.

Gallagher, V. C., Hrivnak, M. W., Valcea, S., Mahoney, C. B. and LaWong, D. (2018) ‘A comprehensive three-dimensional sustainability measure: The “missing P” of “people” – a vital stakeholder in sustainable development’, *Corporate Social Responsibility and Environmental Management*. doi: 10.1002/csr.1493.

Gallopín, G. C. (1996) ‘Environmental and sustainability indicators and the concept of situational indicators. A systems approach’, *Environmental Modeling & Assessment*. doi: 10.1007/bf01874899.

García-Cascales, M. S. and Lamata, M. T. (2012) ‘On rank reversal and TOPSIS method’, *Mathematical and Computer Modelling*. doi: 10.1016/j.mcm.2011.12.022.

Garfí, M., Flores, L. and Ferrer, I. (2017) 'Life Cycle Assessment of wastewater treatment systems for small communities: Activated sludge, constructed wetlands and high rate algal ponds', *Journal of Cleaner Production*. doi: 10.1016/j.jclepro.2017.05.116.

Gasparatos, A. (2010) 'Embedded value systems in sustainability assessment tools and their implications', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2010.03.014.

Gasparatos, A., El-Haram, M. and Horner, M. (2008) 'A critical review of reductionist approaches for assessing the progress towards sustainability', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2007.09.002.

Gasparatos, A. and Scolobig, A. (2012) 'Choosing the most appropriate sustainability assessment tool', *Ecological Economics*, 80, pp. 1–7. doi: 10.1016/j.ecolecon.2012.05.005.

Georgescu-Roegen, N. (1973) 'The Entropy Law & The Economic Problem', in *Toward a Steady State Economy*.

GHG PROTOCOL (2015) *GHG Protocol Scope 2 Guidance: An amendment to the GHG Protocol Corporate Standard*, *GHG Protocol Scope 2 Guidance*. Available at: https://ghgprotocol.org/sites/default/files/standards/Scope 2 Guidance_Final_Sept26.pdf.

Gibbert, M., Ruigrok, W. and Wicki, B. (2008) 'What passes as a rigorous case study?', *Strategic Management Journal*. doi: 10.1002/smj.722.

Gibson, R. B. (2006) 'Beyond the pillars: Sustainability assessment as a framework for effective integration of social, economic and ecological considerations in significant desicion-making', *Journal of Environmental Assessment Policy and Management*. doi: 10.1142/S1464333206002517.

Gibson, R. B. (2012) 'Why sustainability assessment?', in *Sustainability Assessment: Pluralism, Practice and Progress*. doi: 10.4324/9780203112625.

Giovannoni, E. and Fabietti, G. (2013) 'What is sustainability? A review of the concept and its applications', in *Integrated Reporting: Concepts and Cases that Redefine Corporate Accountability*, pp. 21–40. doi: 10.1007/978-3-319-02168-3_2.

Glavič, P. and Lukman, R. (2007) 'Review of sustainability terms and their definitions', *Journal of Cleaner Production*, 15(18), pp. 1875–1885. doi: 10.1016/j.jclepro.2006.12.006.

Global Footprint Network (2020) *National Footprint and Biocapacity Accounts*, *Global Footprint Network*. Available at: https://www.footprintnetwork.org/our-work/sustainable-development/.

Govindan, K. and Jepsen, M. B. (2016) 'ELECTRE: A comprehensive literature review on methodologies and applications', *European Journal of Operational Research*. doi: 10.1016/j.ejor.2015.07.019.

Guba, E. G. (1990) 'The alternative paradigm dialog', *The paradigm dialog*.

Guest, J. S., Skerlos, S. J., Barnard, J. L., Beck, M. B., Daigger, G. T., Hilger, H., Jackson, S. J., Karvazy, K., Kelly, L., Macpherson, L., Mihelcic, J. R., Pramanik, A., Raskin, L., Van Loosdrecht, M. C. M., Yeh, D. and Love, N. G. (2009) 'A new planning and design paradigm to achieve sustainable resource recovery from wastewater', *Environmental Science and Technology*, pp. 6126–6130. doi: 10.1021/es9010515.

Guitouni, A. and Martel, J. M. (1998) 'Tentative guidelines to help choosing an appropriate MCDA method', *European Journal of Operational Research*. doi: 10.1016/S0377-2217(98)00073-3.

Hacking, T. and Guthrie, P. (2008) 'A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2007.03.002.

Hamouda, M. A., Anderson, W. B. and Huck, P. M. (2009) 'Decision support systems in water and wastewater treatment process selection and design: A review', *Water Science and Technology*, pp. 1767–1770. doi: 10.2166/wst.2009.538.

Harker, P. T. and Vargas, L. G. (1987) 'THEORY OF RATIO SCALE ESTIMATION: SAATY'S ANALYTIC HIERARCHY PROCESS.', *Management Science*. doi: 10.1287/mnsc.33.11.1383.

Hartson, H. R., Castillo, J. C., Kelso, J., Neale, W. C. and Kamler, J. (1996) 'Remote evaluation: the network as an extension of the usability laboratory', in *Conference on Human Factors in Computing Systems - Proceedings*.

Heller, M., Garlapati, S. and Aithala, K. (1998) 'Expert membrane system design and selection for metal finishing waste water treatment', *Expert Systems with Applications*. doi: 10.1016/S0957-4174(97)00087-0.

Hendricks, V. M. and Blanken, P. (1992) 'Snowball sampling: A pilot study on cocaine use', *1992*.

HM Government (2005) *Securing the future delivering UK sustainable development strategy, CM6467, London*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69412/pb10589-securing-the-future-050307.pdf.

HM Government (2008) *The Climate Change Act*. Available at: https://www.legislation.gov.uk/ukpga/2008/27/contents (Accessed: 28 July 2021).

HM Government (2010) *Healthy Lives, Healthy People: Our strategy for public health in England*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216096/dh_127424.pdf.

HM Government (2017) *Industrial strategy White paper: Building a Britain fit for the future*, *Department for Business Energy & Industrial Strategy*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf.

HM Government (2018) *A Green Future: Our 25 Year plan to improve the environment*, *UK Government*. Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf.

HM Treasury (2011) *The Plan for Growth*, *HM Treasury BIS*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/221514/2011budget_growth.pdf.

Hobbs, B. F., Chankong, V., Hamadeh, W. and Stakhiv, E. Z. (1992) 'Does choice of multicriteria method matter? An experiment in water resources planning', *Water Resources Research*. doi: 10.1029/92WR00712.

Holden, E., Linnerud, K. and Banister, D. (2014) 'Sustainable development: Our Common Future revisited', *Global Environmental Change*, 26(1), pp. 130–139. doi: 10.1016/j.gloenvcha.2014.04.006.

Holland, A. (1997) 'Sustainability: Why strong sustainability is weak and absurdly strong sustainability is not absurd', in *Valuing Nature? Economics, Ethics and Environment.* Routledge, pp. 119–134.

Huang, I. B., Keisler, J. and Linkov, I. (2011) 'Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends', *Science of the Total Environment*, pp. 3578–3594. doi: 10.1016/j.scitotenv.2011.06.022.

Hugé, J., Waas, T., Eggermont, G. and Verbruggen, A. (2011) 'Impact assessment for a sustainable energy future-Reflections and practical experiences', *Energy Policy*. doi: 10.1016/j.enpol.2011.07.023.

Huysegoms, L. and Cappuyns, V. (2017) 'Critical review of decision support tools for sustainability assessment of site remediation options', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2017.03.002.

Hwang, C.-L. and Yoon, K. (1981) 'Methods for Multiple Attribute Decision Making', in. doi: 10.1007/978-3-642-48318-9_3.

In, J. (2017) 'Introduction of a pilot study', *Korean Journal of Anesthesiology*. doi: 10.4097/kjae.2017.70.6.601.

INCOPA (2014) *Life Cycle Analysis of Leading Coagulants: Executive Summary*.

International Organization For Standardization (1998) 'Ergonomics of human-system interaction. Usability: Definitions and concepts', *International Organization (ISO 9241-11)*.

Ishizaka, A. and Nemery, P. (2013) *Multi-Criteria Decision Analysis: Methods and Software*, *Multi-Criteria Decision Analysis: Methods and Software*. doi: 10.1002/9781118644898.

ISO (2006) '14040: Environmental management–life cycle assessment—Principles and framework', *International organization for standardization*.

Ivanco, M., Hou, G. and Michaeli, J. (2017) 'Sensitivity analysis method to address user disparities in the analytic hierarchy process', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2017.08.003.

Ivory, S. and Brooks, S. (2018) 'An updated conceptualisation of corporate sustainability: Five resources sustainability', in *British Academy of Management Annual Conference 2018*.

Jacquelyn, M. and Forwood, G. (2021) *Environmental law after Brexit*, *White & Case LLP*. Available at: https://www.whitecase.com/publications/alert/environmental-law-after-brexit.

Jamieson, D. (1998) 'Sustainability and beyond', *Ecological Economics*. doi: 10.1016/S0921-8009(97)00142-0.

Jansen, S. J. T. (2011) 'The Multi-attribute Utility Method', in *The Measurement and Analysis of Housing Preference and Choice*. doi: 10.1007/978-90-481-8894-9_5.

Jelassi, M. T., Jarke, M. and Stohr, E. A. (1985) 'Designing a Generalized Multiple Criteria Decision Support System', *Journal of Management Information Systems*. doi: 10.1080/07421222.1985.11517713.

Jimenez, C., Lozada, P. and Rosas, P. (2016) 'Usability heuristics: A systematic review', in *2016 IEEE 11th Colombian Computing Conference, CCC 2016 - Conference Proceedings*. doi: 10.1109/ColumbianCC.2016.7750805.

Johnson, R. B. and Onwuegbuzie, A. J. (2004) 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher*. doi: 10.3102/0013189X033007014.

Johnston, P., Everard, M., Santillo, D. and Robèrt, K. H. (2007) 'Reclaiming the definition of sustainability', *Environmental Science and Pollution Research*. doi: 10.1065/espr2007.01.375.

JuLong, D. (1982) 'Control problems of grey systems', *Systems and Control Letters*. doi: 10.1016/S0167-6911(82)80025-X.

Kalbar, P. P., Karmakar, S. and Asolekar, S. R. (2016) 'Life cycle-based decision support tool for selection of wastewater treatment alternatives', *Journal of Cleaner Production*, 117, pp. 64–72. doi: 10.1016/j.jclepro.2016.01.036.

Karageorgos, P., Latos, M., Kotsifaki, C., Lazaridis, M. and Kalogerakis, N. (2010) 'Treatment of unpleasant odors in municipal wastewater treatment plants', *Water Science and Technology*. doi: 10.2166/wst.2010.211.

Karimi, A. R., Mehrdadi, N., Hashemian, S. J., Nabi Bidhendi, G. R. and Moghaddam, R. T. (2011) 'Selection of wastewater treatment process based on the analytical hierarchy process and fuzzy analytical hierarchy process methods', *Iranian Journal of Environmental Health Science and Engineering*.

Kaya, İ., Çolak, M. and Terzi, F. (2019) 'A comprehensive review of fuzzy multi criteria decision making methodologies for energy policy making', *Energy Strategy Reviews*. doi: 10.1016/j.esr.2019.03.003.

Keeney, R. L. and Raiffa, H. (1976) *Decision Analysis With Multiple Conflicting Objectives Preferences and Value Tradeoffs*, *John Wiley and Sons Inc., New YorkWiley & Sons*.

Kiker, G. A., Bridges, T. S., Varghese, A., Seager, P. T. P. and Linkov, I. (2005) 'Application of multicriteria decision analysis in environmental decision making.', *Integrated environmental assessment and management*, pp. 95–108. doi: 10.1897/IEAM_2004a-015.1.

Kim, G., Park, C. S. and Yoon, K. P. (1997) 'Identifying investment opportunities for advanced manufacturing systems with comparative-integrated performance measurement', *International Journal of Production Economics*. doi: 10.1016/S0925-5273(97)00014-5.

Kim, M., Kim, Y., Kim, H., Piao, W. and Kim, C. (2016) 'Operator decision support system for integrated wastewater management including wastewater treatment plants and receiving water bodies', *Environmental Science and Pollution Research*. doi: 10.1007/s11356-016-6272-6.

King, N. (2004) 'Using templates in the thematic analysis of text", in C.Cassell and G.Symon (Eds.) Essential Guide to Qualitative Methods in Organizational Research.', *London: Sage.*

Klapwijk, A. and Eggels, P. G. (1997) 'Sustainability of municipal wastewater treatment', *Science*, 11(2), pp. 10–11. doi: 10.1016/S0273-1223(97)00199-6.

Kleindorfer, P., Kunreuther, H. and Schoemaker, P. (1993) *Decision Sciences: An integrative perspective*. Cambridge University Press.

Koul, A. and John, S. (2015) 'A Life Cycle Cost Approach for Evaluation of Sewage Treatment Plants', *International Journal of Innovative Research in Advanced Engineering*.

Krovvidy, S., Wee, W. G., Summers, R. S. and Coleman, J. J. (1991) 'An AI approach for wastewater treatment systems', *Applied Intelligence*. doi: 10.1007/BF00118999.

Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*, *The Structure of Scientific Revolutions*. doi: 10.7208/chicago/9780226458106.001.0001.

Lafferty, W. M. (2004) *Governance for sustainable development: The challenge of adapting form to function*, *Governance for Sustainable Development: The Challenge of Adapting Form to Function*. doi: 10.4337/9781845421700.

Lai, E., Lundie, S. and Ashbolt, N. J. (2008) 'Review of multi-criteria decision aid for integrated sustainability assessment of urban water systems', *Urban Water Journal*. doi: 10.1080/15730620802041038.

Lane, E. F. and Verdini, W. A. (1989) 'A Consistency Test for AHP Decision Makers', *Decision Sciences*. doi: 10.1111/j.1540-5915.1989.tb01568.x.

Levers, M. J. D. (2013) 'Philosophical paradigms, grounded theory, and perspectives on emergence', *SAGE Open*. doi: 10.1177/2158244013517243.

Ling, J., Germain, E., Murphy, R. and Saroj, D. (2021) 'Designing a sustainability assessment tool for selecting sustainable wastewater treatment technologies in corporate asset decisions', *Sustainability (Switzerland)*. doi: 10.3390/su13073831.

Loetscher, T. and Keller, J. (2002) 'A decision support system for selecting sanitation systems in developing countries', *Socio-Economic Planning Sciences*. doi: 10.1016/S0038-0121(02)00007-1.

Lundie, S., Peters, G. M. and Beavis, P. C. (2004) 'Life cycle assessment for sustainable metropolitan water systems planning', *Environmental Science and Technology*. doi: 10.1021/es034206m.

Mackenzie, N. and Knipe, S. (2006) 'Research dilemmas: Paradigms, methods and methodology', *Issues in Educational Research*.

Madathil, K. C. and Greenstein, J. S. (2011) 'Synchronous remote usability testing - A new approach facilitated by virtual worlds', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/1978942.1979267.

Malmqvist, J., Hellberg, K., Möllås, G., Rose, R. and Shevlin, M. (2019) 'Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies', *International Journal of Qualitative Methods*. doi: 10.1177/1609406919878341.

Mason, J. (2006) 'Mixing methods in a qualitatively driven way', *Qualitative Research*. doi: 10.1177/1468794106058866.

Meadows, D. H., Randers, J., Meadowns, D. L. and Behrens (1972) *The Limits to Growth: A Report to The Club of Rome*. New York: Universe Books. Available at: https://www.clubofrome.org/publication/the-limits-to-growth/.

Mebratu, D. (1998) 'Sustainability and sustainable development: Historical and conceptual review', *Environmental Impact Assessment Review*. doi: 10.1016/S0195-9255(98)00019-5.

Miles, M. B. and Huberman, A. M. (1994) 'Qualitative data analysis: an expanded sourcebook', *SAGE Publications*.

Millard, D. (2011) 'Management Learning and the Greening of SMEs: Moving beyond Problem-Solving', *German Journal of Human Resource Management: Zeitschrift für Personalforschung*. doi: 10.1177/239700221102500207.

Mills, N., Pearce, P., Farrow, J., Thorpe, R. B. and Kirkby, N. F. (2014) 'Environmental & economic life cycle assessment of current & future sewage sludge to energy technologies', *Waste Management*. doi: 10.1016/j.wasman.2013.08.024.

Miranda, T. and Prasanna, V. (2014) 'A Comparison of Various Normalization in Techniques for Order Performance by Similarity to Ideal Solution (TOPSIS)', *INTERNATIONAL JOURNAL OF COMPUTING ALGORITHM*. doi: 10.20894/ijcoa.101.003.003.023.

Mishra, R. and Rasundram, J. (2017) 'Triangulation an Essential Tool To Enhance the Validity of a', *Scholarly Research Journal for Interdisciplinary Studies*.

Mo, W. and Zhang, Q. (2012) 'Can municipal wastewater treatment systems be carbon neutral?', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2012.08.014.

Molinos-Senante, M., Go´mez, T., Garrido-Baserba, M., Caballero, R. and Sala-Garrido, R. (2014) 'Assessing the sustainability of small wastewater treatment systems: A composite indicator approach', *Science of the Total Environment*, 497–498, pp. 607–617. doi: 10.1016/j.scitotenv.2014.08.026.

Molinos-Senante, M., Hernández-Sancho, F. and Sala-Garrido, R. (2010) ‘Economic feasibility study for wastewater treatment: A cost-benefit analysis’, *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2010.07.014.

Molinos-Senante, M., Marques, R. C., Pérez, F., Gómez, T., Sala-Garrido, R. and Caballero, R. (2016) ‘Assessing the sustainability of water companies: A synthetic indicator approach’, *Ecological Indicators*. doi: 10.1016/j.ecolind.2015.10.009.

Montiel, I. (2008) ‘Corporate social responsibility and corporate sustainability: Separate pasts, common futures’, *Organization and Environment*. doi: 10.1177/1086026608321329.

De Montis, A., De Toro, P., Droste-Franke, B., Omann, I. and Stagl, S. (2004) ‘Assessing the quality of different MCDA methods’, in *Alternatives for Environmental Valuation*. doi: 10.4324/9780203412879-14.

Morgan, D. L. (2007) ‘Paradigms Lost and Pragmatism Regained’, *Journal of Mixed Methods Research*. doi: 10.1177/2345678906292462.

Mostafa, N. and Negm, A. (2018) ‘Promoting organizational sustainability and innovation: An exploratory case study from the Egyptian chemical industry’, in *Procedia Manufacturing*. doi: 10.1016/j.promfg.2018.03.143.

Muga, H. E. and Mihelcic, J. R. (2008) ‘Sustainability of wastewater treatment technologies’, *Journal of Environmental Management*, 88(3), pp. 437–447. doi: 10.1016/j.jenvman.2007.03.008.

Murphy, C. K. (1993) ‘Limits on the analytic hierarchy process from its consistency index’, *European Journal of Operational Research*.

Mustapha, M. A., Manan, Z. A. and Wan Alwi, S. R. (2018) ‘A new quantitative overall environmental performance indicator for a wastewater treatment plant’, *Journal of Cleaner Production*, 167, pp. 815–823. doi: 10.1016/j.jclepro.2017.08.169.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A. and Giovannini, E. (2005) *Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Statistics Working Papers*, *Handbook on Constructing Composite Indicators: Methodology and User Guide*. doi: 10.1787/9789264043466-en.

Nelson, E. T. and Stavrou, A. (2011) ‘Advantages and disadvantages of remote asynchronous usability testing using Amazon Mechanical Turk’, in *Proceedings of the Human Factors and Ergonomics Society*. doi: 10.1177/1071181311551226.

Németh, B., Molnár, A., Bozóki, S., Wijaya, K., Inotai, A., Campbell, J. D. and Kaló, Z. (2019) ‘Comparison of weighting methods used in multicriteria decision analysis frameworks in healthcare with focus on low-and middle-income countries’, *Journal of Comparative Effectiveness Research*. doi: 10.2217/cer-2018-0102.

Neumann, J. Von, Morgenstern, O. and Smith, C. A. B. (1947) ‘Theory of Games and Economic Behaviour’, *The Mathematical Gazette*, 29(285), p. 131. doi: 10.2307/3610940.

Niekamp, S., Bharadwaj, U. R., Sadhukhan, J. and Chryssanthopoulos, M. K. (2015) ‘A multi-criteria decision support framework for sustainable asset management and challenges in its application', *Journal of Industrial and Production Engineering*, 32(1), pp. 44–57. doi: 10.1080/21681015.2014.1000401.

Nielsen, J. (1994) 'Enhancing the explanatory power of usability heuristics', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/191666.191729.

Nielsen, J. and Landauer, T. K. (1993) 'Mathematical model of the finding of usability problems', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/169059.169166.

Nielsen, J. and Molich, R. (1990) 'Heuristic evaluation of user interfaces', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/97243.97281.

Nowell, L. S., Norris, J. M., White, D. E. and Moules, N. J. (2017) 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria', *International Journal of Qualitative Methods*. doi: 10.1177/1609406917733847.

Odu, G. O. (2019) 'Weighting methods for multi-criteria decision making technique', *Journal of Applied Sciences and Environmental Management*. doi: 10.4314/jasem.v23i8.7.

Ofwat (2017) *Delivering Water 2020: Our final methodology for the 2019 price review*.

Olson, D. L. (1988) 'Opportunities and limitations of AHP in multiobjective programming', *Mathematical and Computer Modelling*. doi: 10.1016/0895-7177(88)90481-5.

Olson, D. L. (1996) *Decision Aids for Selection Problems*, *Decision Aids for Selection Problems*. doi: 10.1007/978-1-4612-3982-6.

Ossadnik, W., Schinke, S. and Kaspar, R. H. (2016) 'Group Aggregation Techniques for Analytic Hierarchy Process and Analytic Network Process: A Comparative Analysis', *Group Decision and Negotiation*. doi: 10.1007/s10726-015-9448-4.

Padilla-Rivera, A., Morgan-Sagastume, J. M. and Güereca-Hernández, L. P. (2019) 'Sustainability Assessment of Wastewater Systems: An Environmental and Economic Approach', *Journal of Environmental Protection*. doi: 10.4236/jep.2019.102014.

Padilla, E. (2002) 'Intergenerational equity and sustainability', *Ecological Economics*. doi: 10.1016/S0921-8009(02)00026-5.

Pagilla, K., Shaw, A., Kunetz, T. and Schiltz, M. (2012) 'A Systematic Approach to Establishing Carbon Footprints for Wastewater Treatment Plants', *Proceedings of the Water Environment Federation*. doi: 10.2175/193864709793952350.

Papathanasiou, J. and Ploskas, N. (2018) *Multi Criteria Decision Aid - Methods, Examples and Python Implementations*, *Springer*.

Parkin, S., Uren, S. and Sommer, F. (2003) 'Sustainable development: understanding the concept and practical challenge', *Proceedings of the ICE - Engineering Sustainability*, 156(1), pp. 19–26. doi: 10.1680/ensu.2003.156.3.169.

Pasqualino, J. C., Meneses, M., Abella, M. and Castells, F. (2009) 'LCA as a decision support tool for the environmental improvement of the operation of a municipal wastewater treatment plant', *Environmental Science and Technology*. doi: 10.1021/es802056r.

Pauer, F., Schmidt, K., Babac, A., Damm, K., Frank, M. and Von Der Schulenburg, J. M. G. (2016) 'Comparison of different approaches applied in Analytic Hierarchy Process - An example of information needs of patients with rare diseases', *BMC Medical Informatics and Decision Making*. doi: 10.1186/s12911-016-0346-8.

Pavlovskaia, E. (2014) 'Sustainability criteria: their indicators, control, and monitoring (with examples from the biofuel sector)', *Environmental Sciences Europe*, 26(1). doi: 10.1186/s12302-014-0017-2.

Pearce, D., Atkinson, G. and Mourato, S. (2006) *Cost-benefit analysis and the environment: Recent developments*, *Cost-benefit analysis and the environment: Recent developments*. doi: 10.1787/9789264010055-en.

Pickworth, B., Adams, J., Panter, K. and Solheim, O. E. (2006) 'Maximising biogas in anaerobic digestion by using engine waste heat for thermal hydrolysis pre-treatment of sludge', *Water Science and Technology*, 54(5), pp. 101–108. doi: 10.2166/wst.2006.552.

Plakas, K. V., Karabelas, A. J. and Georgiadis, A. A. (2016) 'Sustainability assessment of tertiary wastewater treatment technologies: A multi-criteria analysis', *Water Science and Technology*, 73(7), pp. 1532–1540. doi: 10.2166/wst.2015.630.

Podvezko, V. (2011) 'The comparative analysis of MCDA methods SAW and COPRAS', *Engineering Economics*. doi: 10.5755/j01.ee.22.2.310.

Pope, J., Annandale, D. and Morrison-Saunders, A. (2004) 'Conceptualising sustainability assessment', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2004.03.001.

Popovic, T., Kraslawski, A. and Avramenko, Y. (2013) 'Applicability of Sustainability Indicators to Wastewater Treatment Processes', in *23rd European Symposium on Computer Aided Process Engineering*, pp. 931–936. doi: http://dx.doi.org/10.1016/B978-0-444-63234-0.50156-1.

Porritt, J. (2012) *Capitalism as if the world matters*, *Capitalism as if the World Matters*. doi: 10.4324/9781849770675.

Power, D. J. (2002) 'Decision Support Systems: Concepts and Resources for Managers', *Information Systems Management*.

Pöyhönen, M. and Hämäläinen, R. P. (2001) 'On the convergence of multiattribute weighting methods', *European Journal of Operational Research*. doi: 10.1016/S0377-2217(99)00467-1.

Prasetiyo, B. and Baroroh, N. (2016) 'Fuzzy Simple Additive Weighting Method in the Decision Making of Human Resource Recruitment', *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*. doi: 10.24843/lkjiti.2016.v07.i03.p05.

Prescott, C. (2009) 'Carbon accounting in the United Kingdom water sector: A review', *Water Science and Technology*. doi: 10.2166/wst.2009.708.

Purvis, B., Mao, Y. and Robinson, D. (2019) 'Three pillars of sustainability: in search of conceptual origins', *Sustainability Science*. doi: 10.1007/s11625-018-0627-5.

Raworth, K. (2012) 'A Safe and Just Space for Humanity: Can we live within the doughnut?', *Oxfam Discussion Papers*, p. 26. doi: 10.1080/00420980120087081.

Rebello, T. A., Roque, R. P., Gonçalves, R. F., Calmon, J. L. and Queiroz, L. M. (2021) 'Life cycle assessment of urban wastewater treatment plants: A critical analysis and guideline proposal', *Water Science and Technology*. doi: 10.2166/wst.2020.608.

Redclift, M. (2005) 'Sustainable development (1987-2005): An oxymoron comes of age', *Sustainable Development*. doi: 10.1002/sd.281.

Rees, W. E. (1990) 'The ecology of sustainable development', *Ecologist*.

Rees, W. and Wackernagel, M. (1996) 'Urban ecological footprints: Why cities cannot be sustainable - and why they are a key to sustainability', *Environmental Impact Assessment Review*. doi: 10.1016/S0195-9255(96)00022-4.

Ren, J. and Liang, H. (2017) 'Multi-criteria group decision-making based sustainability measurement of wastewater treatment processes', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2017.04.008.

Rockström, J. … Foley, J. (2009) 'Planetary boundaries: Exploring the safe operating space for humanity', *Ecology and Society*, 14(2). doi: 10.5751/ES-03180-140232.

Rønning, A. and Brekke, A. (2013) 'Life cycle assessment (LCA) of the building sector: Strengths and weaknesses', in *Eco-Efficient Construction and Building Materials: Life Cycle Assessment (LCA), Eco-Labelling and Case Studies*. doi: 10.1533/9780857097729.1.63.

Roszkowska, E. (2011) 'Multi-Criteria Decision Making Models By Applying the Topsis Method To Crisp and Interval Data', *Multiple Criteria Decision Making*.

Roy, B. (1971) 'Problems and methods with multiple objective functions', *Mathematical Programming*. doi: 10.1007/BF01584088.

Rubin, J. and Chisnell, D. (2008) *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests 2nd Ed*, *Handbook Of Usability Testing*.

Saaty, R. W. (1987) 'The analytic hierarchy process-what it is and how it is used', *Mathematical Modelling*. doi: 10.1016/0270-0255(87)90473-8.

Saaty, T. (1994) 'Fundamentals of Decision Making and Priority Theory with the AHP. , Pittsburgh, PA, U.S.A.', *RWS Publications*.

Saaty, T. L. (1980) 'The Analytic Hierarchy Process 2.1', in *The analytical hierarchy Process; planning, priority setting, resource allocation*, pp. 93–121.

Saaty, T. L. (1990) 'An Exposition of the AHP in Reply to the Paper "Remarks on the Analytic Hierarchy Process"', *Management Science*. doi: 10.1287/mnsc.36.3.259.

Saaty, T. L. (2002) 'Decision making with the Analytic Hierarchy Process', *Scientia Iranica*. doi: 10.1504/ijssci.2008.017590.

Saaty, T. L. (2003) ‘Decision-making with the AHP: Why is the principal eigenvector necessary’, *European Journal of Operational Research*, 145(1), pp. 85–91. doi: 10.1016/S0377-2217(02)00227-8.

Saaty, T. L. and Ozdemir, M. S. (2003) ‘Why the magic number seven plus or minus two’, *Mathematical and Computer Modelling*. doi: 10.1016/S0895-7177(03)90083-5.

Sabia, G., De Gisi, S. and Farina, R. (2016) ‘Implementing a composite indicator approach for prioritizing activated sludge-based wastewater treatment plants at large spatial scale’, *Ecological Indicators*, 71, pp. 1–18. doi: 10.1016/j.ecolind.2016.06.053.

Sachs, J. D. (2012) ‘From millennium development goals to sustainable development goals’, *The Lancet*. doi: 10.1016/S0140-6736(12)60685-0.

Sadr, S. M. K., Saroj, D. P., Mierzwa, J. C., McGrane, S. J., Skouteris, G., Farmani, R., Kazos, X., Aumeier, B., Kouchaki, S. and Ouki, S. K. (2018) ‘A multi expert decision support tool for the evaluation of advanced wastewater treatment trains: A novel approach to improve urban sustainability’, *Environmental Science and Policy*. doi: 10.1016/j.envsci.2018.09.006.

Sala, S., Ciuffo, B. and Nijkamp, P. (2015) ‘A systemic framework for sustainability assessment’, *Ecological Economics*. doi: 10.1016/j.ecolecon.2015.09.015.

Salet, W., Bertolini, L. and Giezen, M. (2013) ‘Complexity and uncertainty: Problem or asset in decision making of mega infrastructure projects?’, *International Journal of Urban and Regional Research*. doi: 10.1111/j.1468-2427.2012.01133.x.

Salo, A. A. and Hämäläinen, R. P. (1997) ‘On the measurement of preferences in the analytic hierarchy process’, *Journal of Multi-Criteria Decision Analysis*. doi: 10.1002/(SICI)1099-1360(199711)6:6<309::AID-MCDA163>3.0.CO;2-2.

Dos Santos, P. H., Neves, S. M., Sant’Anna, D. O., Oliveira, C. H. de and Carvalho, H. D. (2019) ‘The analytic hierarchy process supporting decision making for sustainable development: An overview of applications’, *Journal of Cleaner Production*. doi: 10.1016/j.jclepro.2018.11.270.

Saxena, A., Ramaswamy, M., Beale, J., Marciniuk, D. and Smith, P. (2021) ‘Striving for the United Nations (UN) Sustainable Development Goals (SDGs): what will it take?’, *Discover Sustainability*. doi: 10.1007/s43621-021-00029-8.

Schmidt, K., Babac, A., Pauer, F., Damm, K. and von der Schulenburg, J. M. (2016) ‘Measuring patients’ priorities using the Analytic Hierarchy Process in comparison with Best-Worst-Scaling and rating cards: methodological aspects and ranking tasks’, *Health Economics Review*. doi: 10.1186/s13561-016-0130-6.

Scholten, L., Schuwirth, N., Reichert, P. and Lienert, J. (2015) ‘Tackling uncertainty in multi-criteria decision analysis - An application to water supply infrastructure planning’, *European Journal of Operational Research*. doi: 10.1016/j.ejor.2014.09.044.

Scotland, J. (2012) ‘Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms’, *English Language Teaching*. doi: 10.5539/elt.v5n9p9.

Serageldin, I. and Steer, A. (1994) 'Making development sustainable: from concepts to action', *Making development sustainable: from concepts to action*.

Shih, H. S., Shyur, H. J. and Lee, E. S. (2007) 'An extension of TOPSIS for group decision making', *Mathematical and Computer Modelling*. doi: 10.1016/j.mcm.2006.03.023.

Shneiderman, B. (1982) 'Designing Computer System Messages', *Communications of the ACM*. doi: 10.1145/358628.358639.

Simeral, E. J. and Branaghan, R. J. (1997) 'Comparative analysis of heuristic and usability evaluation methods', in *Proceedings/STC, Society for Technical Communication Annual Conference*.

Sinden, A. (2004) 'The economics of endangered species: Why less is more in the economic analysis of critical habitat designations', *Harvard Environmental Law Review*.

Sipahi, S. and Timor, M. (2010) 'The analytic hierarchy process and analytic network process: An overview of applications', *Management Decision*. doi: 10.1108/00251741011043920.

Song, B. and Kang, S. (2016) 'A Method of Assigning Weights Using a Ranking and Nonhierarchy Comparison', *Advances in Decision Sciences*. doi: 10.1155/2016/8963214.

Sprague, R. H. (1980) 'A framework for the development of decision support systems', *MIS Quarterly: Management Information Systems*. doi: 10.2307/248957.

Srdjevic, Z., Samardzic, M. and Srdjevic, B. (2012) 'Robustness of AHP in selecting wastewater treatment method for the coloured metal industry: Serbian case study', *Civil Engineering and Environmental Systems*. doi: 10.1080/10286608.2012.672412.

Starkl, M. and Brunner, N. (2004) 'Feasibility versus sustainability in urban water management', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2004.03.004.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., De Vries, W., De Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B. and Sörlin, S. (2015) 'Planetary boundaries: Guiding human development on a changing planet', *Science*. doi: 10.1126/science.1259855.

Stringer, E. (2010) 'Action research in education', in *International Encyclopedia of Education*. doi: 10.1016/B978-0-08-044894-7.01531-1.

Tajima, R. and Fischer, T. B. (2013) 'Should different impact assessment instruments be integrated? Evidence from English spatial planning', *Environmental Impact Assessment Review*. doi: 10.1016/j.eiar.2013.02.001.

Tanzil, D. and Beloff, B. R. (2006) 'Assessing impacts: Overview on sustainability indicators and metrics', *Environmental Quality Management*. doi: 10.1002/tqem.20101.

Tarpani, R. R. Z. and Azapagic, A. (2018) 'Life cycle costs of advanced treatment techniques for wastewater reuse and resource recovery from sewage sludge', *Journal of Cleaner Production*. doi: 10.1016/j.jclepro.2018.08.300.

Tashakkori, A. and Teddlie, C. (2003) 'Handbook of mixed methods in social and behavioral research', *Sage Publication*.

Teodosiu, C., Barjoveanu, G., Sluser, B. R., Popa, S. A. E. and Trofin, O. (2016) 'Environmental assessment of municipal wastewater discharges: a comparative study of evaluation methods', *International Journal of Life Cycle Assessment*. doi: 10.1007/s11367-016-1029-5.

Tervonen, T., Gelhorn, H., Sri Bhashyam, S., Poon, J. L., Gries, K. S., Rentz, A. and Marsh, K. (2017) 'MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment', *Pharmacoepidemiology and Drug Safety*. doi: 10.1002/pds.4255.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L. and Goldsmith, C. H. (2010) 'A tutorial on pilot studies: The what, why and how', *BMC Medical Research Methodology*. doi: 10.1186/1471-2288-10-1.

Thames Water Utilities Limited (2014) *Business Plan 2015 – 2020: June 2014*. Available at: https://www.thameswater.co.uk/media-library/home/about-us/regulation/our-five-year-plan/previous-plans/overview-explanatory-note.pdf.

Thames Water Utilities Limited (2016) *Thermal Hydrolysis Process, a Case Study*. Available at: https://www.thameswater.co.uk/-/media/site-content/corporate-responsibility/pdfs/thp.pdf.

Thames Water Utilities Limited (2017a) *Corporate Responsibility and Sustainability (CR&S) Report 2016/17: Content*. Available at: https://www.thameswater.co.uk/media-library/home/about-us/responsibility/sustainability/corporate-responsibility-sustainability-report-2016-17.pdf.

Thames Water Utilities Limited (2017b) *London 2100: the case for change. London's wastewater future*. Available at: https://www.thameswater.co.uk/-/media/Site-Content/Thames-Water/London2100/London-2100-Case-for-change.pdf.

Thames Water Utilities Limited (2017c) *Smart Bugs to revolutionise nutrient removal and recovery*. Available at: https://sustainability.thameswater.co.uk/-/media/Site-Content/Corporate-Responsibility/CRS-2017-18/HWDB/Case-studies/Smart-bugs-to-revolutionise-nutrient-removal-and-recovery.pdf.

Thames Water Utilities Limited (2018a) *Environment, Social and Governance Statement*. Available at: https://corporate.thameswater.co.uk/-/media/Site-Content/Corporate-Responsibility/CRS-2017-18/HWDB/Case-studies/Environment-Social-and-Governance-Statement-2018.pdf.

Thames Water Utilities Limited (2018b) *Our approach to managing sewage sludge*. Available at: https://corporate.thameswater.co.uk/About-us/Protecting-our-environment/Managing-our-impact/Managing-sewage-sludge.

Thames Water Utilities Limited (2018c) *Our Sustainability Policy*. Available at: https://www.thameswater.co.uk/media-library/home/about-us/governance/our-policies/Sustainability/sustainability-policy.pdf (Accessed: 14 June 2021).

Thames Water Utilities Limited (2019a) *About Thames Water. Corporate Responsibility and*

Sustainability Report. Available at: https://sustainability.thameswater.co.uk/how-we-do-business/becoming-more-sustainable/about-thames-water (Accessed: 30 March 2018).

Thames Water Utilities Limited (2019b) *Annual Report and Annual Performance Report 2018/19*. Available at: https://www.thameswater.co.uk/media-library/home/about-us/investors/our-results/previous-reports/2018-19/full-report/annual-report-and-annual-performance-report-2018-19.pdf.

Thames Water Utilities Limited (2019c) *Here for you: Our plan for 2020–2025*. Available at: https://www.thameswater.co.uk/media-library/home/about-us/regulation/our-five-year-plan/pr19/our-plan--summary-version.pdf.

Thames Water Utilities Limited (2020a) *Asset management policy 2020*. Available at: (Archive link no long available).

Thames Water Utilities Limited (2020b) *Biodiversity policy*. Available at: https://corporate.thameswater.co.uk/-/media/Site-Content/Thames-Water/Corporate/AboutUs/CRS/2018/Sustainability-Policy-2018.pdf.

Thames Water Utilities Limited (2020c) *Our journey to net zero carbon and beyond*. Available at: https://www.thameswater.co.uk/about-us/responsibility/net-zero-carbon (Accessed: 28 July 2021).

Thames Water Utilities Limited (2021) *Sustainability and reporting*. Available at: https://www.thameswater.co.uk/about-us/responsibility/sustainability.

Thames Water Utilities Limited (2022) *Creating resilient wastewater catchments*. Available at: https://www.thameswater.co.uk/about-us/regulation/drainage-and-wastewater-management.

Thompson, K. E., Rozanski, E. P. and Haake, A. R. (2004) 'Here, there, anywhere: Remote usability testing that works', in *SIGITE 2004 Conference*.

Triantaphyllou, E. and Sánchez, A. (1997) 'A sensitivity analysis approach for some deterministic multi-criteria decision-making methods', *Decision Sciences*. doi: 10.1111/j.1540-5915.1997.tb01306.x.

Ullah, A., Hussain, S., Wasim, A. and Jahanzaib, M. (2020) 'Development of a decision support system for the selection of wastewater treatment technologies', *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2020.139158.

United Nations (2019) *The Sustainable Development Goals report 2019 from the United Nations Department of Social and Economic Affairs*.

United Nations (no date) *SDGs: The 17 Goals*. Available at: https://sdgs.un.org/goals (Accessed: 7 September 2021).

Vafaei, N., Ribeiro, R. A. and Camarinha-Matos, L. M. (2018a) 'Data normalisation techniques in decision making: Case study with TOPSIS method', *International Journal of Information and Decision Sciences*. doi: 10.1504/IJIDS.2018.090667.

Vafaei, N., Ribeiro, R. A. and Camarinha-Matos, L. M. (2018b) 'Selection of normalization technique for weighted average multi-criteria decision making', in *IFIP Advances in Information and Communication Technology*. doi: 10.1007/978-3-319-78574-5_4.

Velasquez, M. and Hester, P. (2013) 'An analysis of multi-criteria decision making methods', *International Journal of Operations Research*.

Verheem, R. (2002) *Recommendations for Sustainability Assessment in the Netherlands. Environmental Impact Assessment in the Netherlands. Views from the Commission for EIA*.

Vildåsen, S. S., Keitsch, M. and Fet, A. M. (2017) 'Clarifying the Epistemology of Corporate Sustainability', *Ecological Economics*. doi: 10.1016/j.ecolecon.2017.03.029.

Virzi, R. A. (1992) 'Refining the test phase of usability evaluation: How many subjects is enough?', *Human Factors*. doi: 10.1177/001872089203400407.

Waas, T., Hugé, J., Block, T., Wright, T., Benitez-Capistros, F. and Verbruggen, A. (2014) 'Sustainability assessment and indicators: Tools in a decision-making strategy for sustainable development', *Sustainability (Switzerland)*. doi: 10.3390/su6095512.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P. and Krayer von Krauss, M. P. (2003) 'Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support', *Integrated Assessment*. doi: 10.1076/iaij.4.1.5.16466.

Wang, X. and Triantaphyllou, E. (2008) 'Ranking irregularities when evaluating alternatives by using some ELECTRE methods', *Omega*. doi: 10.1016/j.omega.2005.12.003.

Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A. and Zioło, M. (2019) 'Generalised framework for multi-criteria method selection', *Omega (United Kingdom)*. doi: 10.1016/j.omega.2018.07.004.

Wedley, W. C. (1990) 'Combining qualitative and quantitative factors-an analytic hierarchy approach', *Socio-Economic Planning Sciences*. doi: 10.1016/0038-0121(90)90028-6.

Wiek, A. and Walter, A. I. (2009) 'A transdisciplinary approach for formalized integrated planning and decision-making in complex systems', *European Journal of Operational Research*. doi: 10.1016/j.ejor.2008.06.013.

von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioral Research*. Cambridge University Press.

Wong-Parodi, G., Mach, K. J., Jagannathan, K. and Sjostrom, K. D. (2020) 'Insights for developing effective decision support tools for environmental sustainability', *Current Opinion in Environmental Sustainability*. doi: 10.1016/j.cosust.2020.01.005.

Woodward, D. G. (1997) 'Life cycle costing - Theory, information acquisition and application', *International Journal of Project Management*. doi: 10.1016/S0263-7863(96)00089-0.

World Health Organization (2015) 'Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals.', *World Health Organisation*.

Xue, X., Schoen, M. E., Ma, X., Hawkins, T. R., Ashbolt, N. J., Cashdollar, J. and Garland, J. (2015) 'Critical insights for a sustainability framework to address integrated community water services: Technical metrics and approaches', *Water Research*. doi: 10.1016/j.watres.2015.03.017.

Yap, J. Y. L., Ho, C. C. and Ting, C. Y. (2019) 'Aggregating multiple decision makers' judgement', in *Lecture Notes in Networks and Systems*. doi: 10.1007/978-981-13-6031-2_26.

Yin, R. (1994) *Case Study Research: Design and Methods (2ª Ed) Thousand Oaks, CA*, *SAGE Publications*.

Yin, R. K. (2014) *Case study research: Design and methods (5th ed.)*, *Thousand Oaks, CA: SAGE Publications*.

Yoon, P. and Hwang, C.-L. (1995) *Multiple Attribute Decision Making: An Introduction*, *Sage Publications Thousand Oaks CA*.

Zadeh, L. A. (1965) 'Fuzzy sets', *Information and Control*. doi: 10.1016/S0019-9958(65)90241-X.

Zanakis, S. H., Solomon, A., Wishart, N. and Dublish, S. (1998) 'Multi-attribute decision making: A simulation comparison of select methods', *European Journal of Operational Research*. doi: 10.1016/S0377-2217(97)00147-1.

Zang, Y., Li, Y., Wang, C., Zhang, W. and Xiong, W. (2015) 'Towards more accurate life cycle assessment of biological wastewater treatment plants: A review', *Journal of Cleaner Production*. doi: 10.1016/j.jclepro.2015.05.060.

Zeng, G., Jiang, R., Huang, G., Xu, M. and Li, J. (2007) 'Optimization of wastewater treatment alternative selection by hierarchy grey relational analysis', *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2005.12.024.

Zhou, P. and Ang, B. W. (2009) 'Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure', *Social Indicators Research*. doi: 10.1007/s11205-008-9338-0.

Zouboulis, A. and Tolkou, A. (2015) 'Effect of climate change in wastewater treatment plants: Reviewing the problems and solutions', in *Managing Water Resources Under Climate Uncertainty: Examples from Asia, Europe, Latin America, and Australia*. doi: 10.1007/978-3-319-10467-6_10.

Appendices

Appendix 1. Ethics application: Contingency Plan

Research tasks	Potential concerns	Mitigation and remedy measures
Understanding the decision context: Research interviews	Security of audio recording and transcription	Files stored in a local encrypted folder in a device owned by the water company
	Confidentiality and anonymity of the data	The names of participants are not asked. Only the principal researcher has the access to the files
	Loss of data due to IT device renewal scheduled in May 2019	a. File stored locally on the old device will be transferred to company's online drive b. Alternatively, an encrypted flash drive can be ordered from IT department to temporarily store the files
	Difficulty of installing and using NVivo® on a company device due to IT policy	a. An early request will be made to IT for approval of installation of NVivo®. If the software is blocked by company's network proxy, coding and thematic analysis will be conducted manually using Microsoft Word® and Excel®
Identification of assessment criteria and indicators and methods for case studies: Focus group	Difficulty of arranging focus groups due to limited time availability of participants	a. Revise the list of participants and look for alternative contacts b. Alternatively, quantitative methods (e.g. online survey) will be considered and the research methodologies will be modified correspondently to justify this change
Trialling the proposed assessment framework	No prior knowledge process modelling of wastewater technologies	Consultation and support will be also sought within the process and modelling department
	Limited availability of data to undertake the evaluation of technologies. This includes but not limited to: a. Lack of data of full-scales trial of technologies b. Lack of Data of laboratory/pilot scale trial c. Financial estimates such as Opex and Capex	The uncertainty will be documented and estimated in the assessment result. Sensitivity check can also be conducted. Additional measures can be considered: a. Prediction models will be sought from empirical studies to upscale the data from pilot scales to full-site scale b. Alternative assessment methods with qualitative inputs and subjective judgement will be considered
Project management	Little dissemination of research findings	Actively looking for internal (e.g. steering group; lunch and learn) and external dissemination opportunities (e.g. university conference and international conference)
	Limited funding for conference attendance	Funding from multiple sources will be considered. Planning will be done to allocate funding for different conferences with good justifications
	Temporarily withdrawal and annual leave	Any foreseeable events will be planned in advance. Unforeseeable reasons that may lead to withdrawal will be discussed with supervisors and programme administrators to ensure the progress of project is not unsatisfactorily compromised
	Visa condition compliance	Supervisor meeting will be organised regularly and timely. Meeting minutes and progress review will be submitted on time

Appendix 2. Ethics application: Participation Information Sheet

INFORMATION SHEET FOR PARTICIPANTS (v0.5 11/09/2018). The file was created based on a templated provided by the Ethics committee at the University of Surrey. Some information has been removed or modified in this thesis for confidentiality reasons.

Title of Study (Provisional): Sustainability impact of current and future wastewater treatment process

Invitation Paragraph

I am a doctorate practitioner at the Centre for Environment and Sustainability at the University of Surrey. I would like to invite you to participate in this research interview which forms part of my research project trialling at Thames Water Utilities. This is a voluntary participation and you have the right to withdraw at any time. Choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to understand why the research is being done, what your participation will involve and how data is managed. Please read the following information carefully and discuss it with me if you wish. Ask me if there is anything that is not clear or if you would like to more information upon request.

What is the purpose of the study?

The study aims to propose an adaptive assessment framework and tool to inform resilient decision making on selecting wastewater treatment processes and technologies, in order to adapt to the emerging long-term business challenges.

This research interview involves information collection of:

1. Crucial business values among key stakeholders and also employees
2. Decision-making process in the wastewater investment

The acquisition of this information will facilitate the formulation of the assessment framework and development of a flexible and adaptive asset management tool that deals with dynamic business constraints. This is also to make sure the research direction is closely aligned with Thames Water's strategic direction.

Who will be invited?

In this study, a diverse group of representatives are selected to reflect the key business values and roles involved in the decision-making process. A stakeholder analysis matrix has been carried out to identify the key business functional groups where potential participants are selected from. The analysis will provide a good understanding of the most influential executive roles in Thames Water that has an impact on this project.

Do I have to take part in the interview?

Participation is voluntary. There are no incentives involved. Participants are free to withdraw from the study at any time without giving reasons. Withdrawal from this part of study will not disadvantage you in any way.

What will happen to me if I take part?

If you decide to take part you will be given this information sheet to keep and will be asked to sign a consent form. An interview will then be scheduled upon agreement. I will send out an appointment invite through Thames Water email system to confirm time, date and location. Ideally, the location of interviews will be at either:

(a) *** OR

(b) *** (Locations are removed for confidential reason)

The interview takes approximately 45-60 minutes. Only one interview is required in this part of the research. There will be no personal information, which could make you identifiable, recorded in the interview. The only potential identifiable information that will be collected is 'job title'. However, the information will be strictly protected and stored in the encrypted device and folder in Thames Water's depository so only the research team has the access to this data.

The interview is audio recorded by an encrypted device owned by Thames Water. Audio files will be deleted immediately once it has been transcribed. Audio will be transcribed by the principal researcher. Transcription will be stored in Thames Water data depository with limited access, which means it will not be shared beyond research team (i.e. researcher and supervisory team).

What are the possible benefits and risks of taking part?

Benefit:

The interview is a crucial part of the study to understand the existing sustainability strategy and key business values at Thames Water Utilities. Your opinions and suggestion will, therefore, underpins the new framework and assists the development of a robust decision support tools that enhance business resilience.

The interview is also an effective way to promote intellectual discussion and stakeholder engagement in the business, which further enhances Thames Water's collaborative capability.

Risk:

There is potential concern over the confidentiality and disclosure of business-sensitive information. The participants will be asked about the information regarding the business strategy and practices. However, personal thoughts and bias may be added on the top of the facts by participants as part of the discussion process. Confidentiality is, therefore, strictly protected to ensure participants are not disadvantaged in any way. Any information, data, and conclusion drawn from this interview will only be used for the research purpose and will not be disclosed beyond the research team. The data will be handled with an encrypted device and stored in a secure depository at Thames Water. Other employees at Thames Water do not have the access to the data. Any potential opportunity of research publication related to this part of the study will be checked with the company's policy to ensure no unintended disclosure of confidential or commercial-sensitive information. If participant has any concern over confidentiality, please contact the principal researcher.

How is the project being funded?

Thames Water Utilities funds this project and owns the intellectual property. University of Surrey acts as the research sponsor for this project.

Who should I contact for further information?

If you have any questions or require more information about this study, please contact me using the following contact details:

-Alex Jiean Ling (Researcher), j.ling@surrey.ac.uk

The University has in force the relevant insurance policies which apply to this study. If you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study then you should follow the instructions given above

Who is Handling My Data and what will happen to my data?

Thames Water Utilities is the data controller of this project. For the data extracted from this interview, only the research team, consisting of the principal researcher and supervisory team, has the direct access to the data. Please note that you have the right to withdraw the data up to two weeks after the interview is conducted. The withdrawal cannot be done once transcription is finalised and transferred to the encrypted storage due to the anonymity nature of the data. The transcriptions will be primarily stored in the company's data depository with encrypted folders. The principal researcher has the responsibility to protect the data and control over the data. On occasion, the access of data can be granted for audit purpose by University of Surrey or Thames Water Utilities.

Confidentiality and anonymity are strictly protected and maintained throughout the research process. No obvious identifiable data or information from the interview will be collected or stored. The only possible 'identifiable' risk is the information of 'job title' may be recorded in the interview. This is justified as the research requires different perspectives from various business roles that are involved in the decision-making process. However, the data access is strictly limited to only the research team to minimise the risk of participants be identified.

What will happen to the results of the study?

I will produce a final report summarising the main findings, which will be shared with you upon your request. The transcription of the interview will also be shared with you before transferred to the data storage. If you think the information extracted from the interview was not appropriate or you are misrepresented in any way, please inform the principal researcher and relevant correction can be made.

If any kind of the research data or deliverables needs to be disseminated or published for research communication purpose (e.g. conference, doctorate thesis, workshop), this will be checked by Thames Water to ensure no confidential or commercial-sensitive information is disclosed in accordance with company's data policy.

What if I want to complain about the way data is handled?

If you wish to raise a complaint about how we have handled your personal data, you can contact the research team, whose detail has been presented in the previous section.
You can contact the Thames Water data protection officer (… contact details removed), who can advise on any concern with data management and handling.

If you want to contact the research sponsor (University of Surrey), you can contact the Data Protection Officer (…contact details removed) who will investigate the matter. If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO) (https://ico.org.uk/).

For contact details of the University of Surrey's Data Protection Officer please visit: https://www.surrey.ac.uk/information-management/data-protection

Limits to confidentiality

Confidentiality will be respected and protected all the time unless there are compelling and legitimate reasons for this to be breached. If this was the case we would normally inform you first of any decisions that might limit confidentiality.

Thank you for reading this information sheet and for considering taking part in this research.

Appendix 3. Ethics Application: Consent Form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Sustainability impact of current and future wastewater treatment processes

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

By ticking/initialling each box you are consenting to this element of the study. It will be assumed that un-ticked/un-initialled boxes mean that you DO NOT consent to that part of the study and you may be deemed ineligible for the study.

1. I confirm that I have read and understood the information sheet dated [11/09/2018 v0.5] for the above study. I have had the opportunity to consider the information and asked questions which have been answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time during the study without giving any reason and without being disadvantaged in any way. Furthermore, I understand that I will be able to withdraw my data up to two weeks after the interview. ☐
3. I consent to the processing of my personal information for the purposes explained to me. I understand that such information will be handled in accordance with current data protection regulations in Thames Water and UK Data Protection Act 2018. ☐
4. I understand that anonymity of the data will be protected and maintained. No special category data will be recorded. However, I am aware that the researcher will collect the information of my business role for purely research purpose. The researcher will employ every measure to minimise the risk of you be identified in any way. ☐
5. I understand that confidentiality of the data is strictly protected and maintained. Whenever the research output needs to be published for research communication (e.g. conference, doctorate thesis), this will be checked by the company to avoid any unintended disclosure of sensitive or confidential information. ☐
6. I understand that only the research team (i.e. principal researcher and supervisors) has the direct access to the data. However, this information may be subject to review by responsible individuals from the University of Surrey and/or regulators and Thames Water for monitoring and audit purposes. ☐
7. I consent to my interview being audio recorded. I am aware that the audio will be recorded by an encrypted device owned by Thames Water and will be deleted immediately once it is transcribed. I am also aware that transcription will be done only by the principal researcher and will not be shared beyond the research team. Transcription will, thereafter, be stored in an encrypted folder on Thames Water's data depository. ☐

8. I wish to receive a copy of the summary of the transcription and research findings. (optional) ☐

___________________	___________________	__________________
Name of Participant	**Date**	**Signature**
___________________	___________________	__________________
Name of Researcher	**Date**	**Signature**

Appendix 4. Ethics application: Recruit Email

The primary recruitment approach is by email.
The recruitment methodology includes:
(a) Stakeholder analysis matrix: the potential participants are prioritised using the influence VS interest matrix. The most crucial business functional groups for this project are identified. This result forms the selection criteria of potential interviewees.

(b) 'Free find': the principal researcher uses his own knowledge to decide the eligibility and select potential participants via internal network and mailing list. The access to the mailing list is granted by the company.

(c) Recommendation 'Snowballing': The participant's contact details are forwarded by the line manager who is the wastewater innovation process manager at the company. The line manager is also part of the supervisory team of the research project and is thus in a suitable position to provide network information to the principal researcher for purely research purpose.

Email template:

Title: Interview Invite for wastewater innovation research

I am a research scientist working in the wastewater process innovation team at Thames Water as part of my EngD study associated with the University of Surrey. My research project aims to design an adaptive decision support tool that informs wastewater treatment technology selection and wastewater asset management, in order to address emerging business challenge such as climate change and population growth. I believe this project has the potential to deliver long-term business strategy in the most sustainable and resilient way.

This project requires in-depth understanding of our current business practices in this topic, which I hope to gain through discussion within our organisation and key stakeholders. Your opinions and shared business values will contribute to the foundation of the research framework. The interview will collect relevant information so that the research outcome is specifically aligned with Thames Water's vision.

Please let me know if you are interested in participating this interview. Please contact myself if you have any concern and concerns or would just like more information. You can also contact my line manager (contact details removed) or my academic supervisor (contact details removed) for additional information.

This project has been reviewed and given an ethical opinion by the University committees to ensure the ethical standard is maintained. Your name and email address were forwarded by my line manager and none of personal information was disclosed beyond Thames Water.

Appendix 5. Ethics Application: Interview schedule

Interview schedule (version0.5 11/09/2018)
Project title: Sustainability Impact of current and future wastewater process
Principal researcher: Jiean Ling
This interview schedule describes the interview process including: (1) location and duration (2) design of interview content (3) interview procedure

1. Location, date, duration
The preferred locations for undertaking one-to-one interview are, ideally, based at company's premises, either

a) Location A (Address removed)

b) Location B (Address removed)

The interviews will be undertaken between the 1st September 2018 and 30th November. However, this may need to be extended practice depending on the time availability and flexibility of participants. The expected number of participants to be recruited is around 15. This is subject to change due to availability of participants and eligibility of interviewees but will not be less than 12 or more than 18. The expected duration of individual interview is approximately 45 – 60 mins.

2. Design of interview content
During the interview, the interviewer (i.e. principal researcher) will introduce the research background, rationale and objective of this project, and use these questions set out below to guide the flow of the discussion. The introduction will be such that the research setting is well clarified but minimises the potential to influence the answers provided by the participants. Questions will be asked in a relaxed atmosphere. The wording and order of the question may be adjusted to suit different communication styles of the individual participants. Some other research materials may be presented to the participant to facilitate the discussion (e.g. a summary diagram of key business strategy).

Three themes of questions will be investigated during the interview. However, the specific questions may change to reflect new or additional topic of interest within the same general research focus. In addition, this list is not exhaustive and there might be divergence in the real conversation.

a) Setting the scene and professional background of participant
    1) Could you please tell me about your current position and daily activities?

b) Decision making strategy and process

    1) What is the decision-making process to evaluate wastewater technology/process for implementation on the sewage treatment work, including sludge treatment as well?

    2) What do you think are challenging aspects when making strategic decision?

3. Interview procedures
The structural flow of the interview process is presented below:

a) Interview invites through internal email. If invites are accepted, the specific date, time and location will be scheduled and confirmed.
b) Consent request: give a short brief about the Participant Information Sheet (PIS) and Consent Form (CF) so participants are familiarised with them. Consent signature is requested
c) Introduction: who am I (i.e. principal researcher); research rationale and objectives; why this interview is needed and why you are invited
d) Interview questions and discussion
e) Interviewee raises questions and concerns for the researcher if necessary

The whole interview process is audio recorded using an encrypted device that is owned by the company. The audio will be deleted once transcribed the principal researcher. The transcription and any related data will be primary stored in encrypted folder in the company's data depository. This data can only be shared within the research team (i.e. project supervisors) for academic discussion and will not be shared beyond the research purpose and domain. This is to ensure the confidentiality is strictly protected and maintained while the research desire is also fulfilled.

Appendix 6. Ethics Application: Risk Assessment

Risk Assessment for Study: Sustainability impact of current and future wastewater treatment process					
Version: 0.5	**Date: 11/09/2018**				
1. Identified Risks	**2. Likelihood**	**3. Potential Impact/Outcome**	**4. Potential Severity of Outcome**	**5. Risk Management/Mitigating Factors**	
Identify risks/hazards present	*Identify how likely the event is i.e.* *Very likely/ Likely/ Possible/ Unlikely*	*Who might be harmed and how?* *Ensure you have considered the research team, participants and anyone not directly involved in the research.*	*Classify the severity of outcomes identified in 3.* *i.e. High/ Medium/ Low*	*Evaluate the risks and decide on the precautions.*	*Standard Operating Procedures*/ risk assessments* *Enter Ref no/ title/ expiry date*
Mental wellbeing of the participants can be affected	Possible	Participant: -Psychological stress and pressure when discussing business matter -Reluctance and stress when being asked about a question	Medium	-Invite participant to read the participant information sheet together so all information is well clarified -Create a relaxing interview atmosphere for the participant using guiding questions -Offer participant the right to withdraw at any point without giving any reason	N/A
Unintended disclosure of company sensitive or confidential information	Possible	Participant: -Reputational damage of the company Researcher: -Breach of company's data policy -May face reputational and legal consequences	High	-Liaise with participant and familiarise with company's data policy and what information can be shared or disclosed through academic publication -Store data in the encrypted folder or device in the secure depository. It will be deleted once the research framework has been finalised -Only the research project team has the direct access to the data unless special demand arises such as for audit purpose	N/A

(Continued)

Risk Assessment for Study: Sustainability impact of current and future wastewater treatment process					
Version: 0.5	**Date: 11/09/2018**				
1. Identified Risks	**2. Likelihood**	**3. Potential Impact/Outcome**	**4. Potential Severity of Outcome**	**5. Risk Management/Mitigating Factors**	
Risk of data loss and participants are identified	Unlikely	Participant: -The data may render them identifiable -Psychological stress when their opinions is leaked and disclosed Researcher: -Reputational damage and maybe legal outcome	Medium	-Interviewed will be audio recorded with an encrypted device owned by the company -Audio will be deleted transcription will be done only by the principal researcher -Transcription will be stored in encrypted folders in the company's data depository -The access to the data is only limited to the researcher team consisting the principal researcher and the supervisors	N/A
Biased opinion/negative comments about company	Possible	-This may disadvantage participants and compromise their position in the company -Psychological distress if this information is leaked	Medium	-Researcher will intervene and redirect the discussion if the conversation involves strong personal opinions and bias -Research will make sure the interview question only aim to collect factual information -Anonymity, confidentiality and data security is safeguarded by the measures mentioned previously	N/A

Appendix 7. Usability test pilot trial result

Action log:
Total meeting length: 35 mins

(00:00-0:59) Browsing the document reference document
(1:00-1:40) Saving the document locally 40 sec
(1:45-4:20) Browsing the instruction sheet in the tool
(4:25-5:25) Browsing and filling out the project information sheet
(5:25-9:10) Browsing and filling out the performance matrix sheet
(9:10-10.30) Browsing and filling out the weighting sheet
(10:30-14:10) Reviewing the results session. A critical issue was detected and the test session was cancelled.

	Time to complete	Number of errors (excluding self-resolved)	Number of prompts requested or made (excluding self-resolved)
Data reference document	1 minute	-	
Tool (1. Instruction)	2mins 35sec	-	
Tool (2. Project info)	2mins	0	1
Tool (3. Performance matrix)	3mins 45 sec	2	3
Tool (4. Weighting)	1min 20 sec	0	
Tool (5. Results)	3mins 40 sec (Cancelled due to a critical issue)	-	-
Total		2	4

Errors:

- (Performance matrix) Participant attempted to fill out the performance of indicators with no data provided
- (Performance matrix) Participant omitted the performance data for the qualitative indicators and provided own judgements

Prompts:

- "*Can I copy and paste the description of the options* (from the data reference document)"?
- "*Can I copy and paste the performance data* (from the data reference document)"?
- Prompt was made to inform the participant to leave it blank if data was not applicable/available
- "*Can I delete the data in the cell? Will it mess (the spreadsheet) up*"?

Issues noted from the pilot trial (Minor, median, critical risk)

Location of the issue	Descriptions of issues	Risk level of issues	Actions to resolve
Data reference file	Participant would like the option descriptions to be next to the performance matrix so it is easier and people won't make a mistake. Same with the options in the performance matrix	Suggestion	Descriptions of options were added to the table of performance data as suggested
Prior to the start of the task	To remind the participant to save the files on their device before opening the file. The file format was altered when opened from 'viewing only' mode.	Medium	This has been added to the orientation script.
Instruction sheet	Minor grammar errors. 'Sustainability of' and 'Fill out'	Minor	Correction made
Instruction sheet; Questionnaire sheet	Replace 'asset planners' as the intended users. Users with other roles will also be using it.	Minor	Correction made
Instruction sheet	Visual error: words were cut off by the size of the text box use.	Medium	The text has been revised to reduce the word count
Instruction sheet	maybe copy the instruction sheet as a picture to prevent unintended to format when scales are changed?	Suggestion	No change made. The full sheet is kept as its original format to allow future editing.
Performance matrix sheet	Prompt was made to inform the participant to leave it blank if data was not applicable/available	Medium	The instruction has been re-worded and highlighted to increase visibility
Performance matrix sheet	Keep the qualitative data visually shown as whole number without the decimals	Suggestion	Cell data formatting has been updated
Weighting sheet	Message box error: Participant did not fill out the weighting input in the vertical order. The completion message box popped up even the input was incomplete	Medium	The condition of the message box has been re-coded so it only shows itself when all input are completed.
Results sheet	Graph operation errors: the normalisation sheet could not process the data input in the performance matrix. This caused erroneous information shown on the graph and misleading result	Critical	The issue seems to be the data format when coping directly from Word file opened in Teams. In the orientation script, participants will be asked to open the document reference file using the desktop function.
Questionnaire sheet	Grammar error 'I found'	Minor	Correction made

Appendix 8. Usability testing: Consent Form

v1.0 (31/05/2021)

Please complete and sign this form after you have listened to an explanation about the research.

Title of Study: Usability test of the sustainability assessment tool

Thank you for considering taking part in this research. This study aims to test the usability of a newly developed decision support tool. You will be asked to complete a task using this tool. Your performance will be monitored using screening recording on Microsoft Teams. You will be asked to share your screen during the test session. Once the task is finished, you will be asked some questions to provide feedback. During the screen recording, you are free to choose to turn off the camera if you wish. The screening recording will be collected and stored securely on the private OneDrive company account. Only the researcher has the access to the files. Once the data from the recording is recorded and transcribed, the screen recording files will be destroyed immediately. No personal information will be recorded in this study and your anonymity will be protected throughout the data collection, analysis and reporting.

By ticking each box you are consenting to this element of the study. It will be assumed that un-ticked/un-initialled boxes mean that you DO NOT consent to that part of the study and you may be deemed ineligible for the study. You will be given a copy of this Consent Form to keep and refer to at any time.

9. I confirm that I have read and understood the information above. I have had the opportunity to consider the information and asked questions which have been answered satisfactorily. ☐
10. I understand that my participation is voluntary and that I am free to withdraw at any time during the study without giving any reason and without being disadvantaged in any way. ☐
11. I understand that confidentiality of the data is strictly protected and maintained. Whenever the research output needs to be published for research communication (e.g. conference, doctorate thesis), this will be checked by the company to avoid any unintended disclosure of sensitive or confidential information. ☐
12. I understand that only the principal researcher has the direct access to the data. However, this information may be subject to be reviewed for ethics audit purposes. ☐
13. I consent that the test session to be screen recorded. I am aware that the recording will be stored in a secure data depository and immediately destroyed once data are recorded and transcribed. ☐
14. I wish to receive a copy of the summary of findings of usability tests. (optional) ☐

Participant name	**Date**	**Signature**

Appendix 9. Orientation plan and script

1. Welcome the participant and briefly introduce the purpose of this study

(Slide 1) Hi How are you doing? I am great thanks. Well thank you for accepting the invite. So today we will be testing a new decision support tool which intends to compare different wastewater treatment processes and technologies. So it would be interesting to invite you to test this tool today and we will collect some feedback. Your feedback will help to improve the usability of the tool before it is finalised. Does that make sense?

(Slide 2) So this session will last about 40 minutes. For the first 10 minutes I will briefly introduce the tool and the description of the task to be tested. I will also send you a consent form to sign before you start using the tool. And then the next 15 minutes you will be using the tool yourself to analyse a business case in a simulated environment. For the last 15 minutes, I will ask you some questions to collect feedback. Is that clear?

2. A brief training session of the tool (with a ppt) and task description

(Slide 3) So a quick introduction on the tool. The tool was built to support our asset investment decision at the Gateway 1. Particularly what this tool does is to compare different potential wastewater treatment options based on a range of sustainability indicators and using a multi-criteria analysis model to suggest the most desirable option for asset planners. Basically it is an optioneering tool.

(Slide 4) So today I would like to invite you to use the tool to compare different phosphorus removal options at (…location removed) sewage treatment works. The site requires an upgrade to meet a tighter phosphorus consent going through AMP 7. We have eight options trialled in pilot plants based on the combinations of two type of tertiary filters, two dosing ratios and two types of dosing chemicals. So the task for you is to use the decision support tool to identify the best treatment option.

(Slide 5) You will be receiving two files to complete the analysis of this case. One of the files is the Excel spreadsheet which is the actual decision support tool you will use for analysis. Another file contains all the data reference of this project. Most of information can be directly copied from the data reference file. You will be undertaking the analysis independently with these materials. You are free to browse those files as long as you wish. The basic instructions are all provided inside the excel spreadsheet and they should be sufficient to guide you to complete the analysis. I won't interrupt you during analysis unless you request additional assistance or explanations from me. You can let me know that you have finished once you have completed the analysis and reviewed the results. Is that clear?

3. Sign the consent form and handover the testing materials

And now, I would like you to read and sign the consent form. I will brief describe what data will be collected from the testing process and how they will be managed.

Now I will send you the materials. You need to save both files to your device locally and open them using the apps to use them.

Could you please share your screen and I will start screen recording once you open the file? You are free to start whenever you are ready.

Appendix 10. individual weighting profiles developed from the usability testing using the direct rating weighting method

	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8	Input 9	Input 10	Input 11	Input 12
Net carbon emission	0.103	0.146	0.127	0.083	0.108	0.097	0.107	0.120	0.067	0.088	0.125	0.116
Pollutant removal potentials	0.128	0.139	0.141	0.119	0.135	0.116	0.118	0.120	0.133	0.125	0.125	0.155
Biodiversity net gain	0.077	0.073	0.042	0.083	0.054	0.090	0.095	0.038	0.040	0.063	0.119	0.016
Public value added	0.077	0.044	0.042	0.077	0.054	0.090	0.095	0.038	0.067	0.063	0.094	0.016
Odour	0.090	0.073	0.042	0.083	0.068	0.084	0.083	0.095	0.133	0.125	0.094	0.047
Ease to operate and maintain	0.115	0.117	0.127	0.119	0.108	0.110	0.089	0.120	0.093	0.125	0.125	0.124
Flexibility (Adaptability)	0.090	0.058	0.070	0.119	0.095	0.090	0.101	0.089	0.067	0.100	0.100	0.093
Reliability	0.115	0.131	0.127	0.107	0.135	0.116	0.101	0.127	0.133	0.125	0.100	0.155
Opex	0.103	0.117	0.141	0.113	0.122	0.103	0.107	0.127	0.133	0.100	0.056	0.140
Capex	0.103	0.102	0.141	0.095	0.122	0.103	0.107	0.127	0.133	0.088	0.063	0.140