Assessing Urban Environmental Sustainability Performance of Greater London

Ömer Faruk Eriş

September 2021

This dissertation is submitted in part requirement for the Master of Science in Smart Cities and Urban Analytics in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL

Submitted:	2nd September 2021
Supervisor:	Mark Tewdwr-Jones
Module Code:	CASA0010
Module Title:	MSc Smart Cities and Urban Analytics
Student ID:	20198890
Word Count:	10,500

Abstract

Due to the overpopulation and accelerated urbanization in recent years, there has been negative stress on the environment. A significant amount of carbon footprint, excessive consumption of natural resources, air pollution affecting the life quality, and excessive waste production are the main ingredients of this stress. Since this situation is not sustainable, several solutions and strategies emerged under the name of sustainable development, especially after the UN's Brundtland report in 1987. Information flow is a very important tool for sustainable development both to question the current situation of the environment and to evaluate the strategies and actions to be taken. For this reason, many sustainable assessment methods have emerged. Perhaps the most popular among them is the composite indicator method, which converts multiple datasets into a single index. Although this method is widely used internationally and referred to in policy documents, it has been observed that there is a research gap and limitations at the regional scale. In this study, urban sustainability, its origins, and its pillars are reviewed. The research gaps and limitations of existing assessment methodologies are evaluated. At the final step, the environmental sustainability performance of the selected region, London and its wards, are attempted to be measured. By applying cluster analysis on the results, regions with similar characteristics are classified to guide the decision makers. It is aimed that the datasets to be used for the composite indicator are publicly available data, and the variables are selected by taking reference from the existing policy documents.

Declaration

I, Ömer Faruk Eriş, hereby declare that this dissertation is entirely my own original work and that all sources have been acknowledged. This dissertation is 10,500 words in length. Word count processed by Microsoft Word.

Signed:

Date: 2nd September 2021

Contents

Figures

Tables

Acronyms

CI	Composite Indicator
EPI	Environmental Performance Index
ESI	Environmental Sustainability Index
EW	Equal Weighted
GLA	Greater London Authority
LCA	Life Cycle Assessment
LEPIS	London Environment Strategy Implementations and Strategies
OECD	The Organisation for Economic Co-operation and Development
PCA	Principal Component Analysis
UA	Uncertainty Analysis
UN	United Nations
WCED	World Commission on Environment and Development
WTO	World Trade Organization

Acknowledgements

Firstly, I would like to thank my supervisor, Mark Tewdwr-Jones, for great advice and encouragement with a perfect blend of insight. I am grateful for Mark's feedback throughout this research. I also thank my colleagues and open-source communities for several valuable programming suggestions.

I would like to express my sincere gratitude to all the staff at UCL'S Centre for Advanced Spatial Analysis for everything they have taught and inspired me. Thanks to them, I not only gained valuable knowledge throughout the year, but also had fun. Also, I must express my very profound gratitude to the Republic of Turkey, especially the Ministry of National Education for their excellence both material and spiritual supports.

Finally, my sincere thanks and, also apologies, go to my wife for all her unwavering support and patience throughout this challenging process. Without your extraordinary understanding and encouragement, I would not have been able to complete my study.

Chapter 1 – Introduction

1.1 Context

With the influence of the industrial revolution, there has been an increasing movement of people from rural areas to urban spaces (United Nations, 2018). The main reason for this movement is the idea of people reaching a better quality of life and better economic opportunities in cities thanks to the developing circumstances (Grimm et al., 2008). As a result, there has been negative stress on the environment due to the over-population in cities and accelerated urbanization (United Nations, 1987). A significant amount of carbon footprint, excessive consumption of natural resources, air pollution affecting the life quality, and excessive waste production are the main ingredients of this stress (ibid.). It is obvious that this situation is not sustainable for the environment and the living creatures in the environment, so various solutions that can be applied at different scales were necessary to reduce the pressure on the environment (Corredor-Ochoa et al., 2020). Immediately after the industrial revolution, urban design solutions such as green cities and organic forms in the ecological context were put forward (Rees & Wackernagel, 2008). However, these solutions did not create global awareness as they only focused on certain aspects of sustainability.

Perhaps, the first study that approaches sustainability in a holistic way is the work of the United Nations called the Brundtland Report focused on the term sustainable development (Corredor-Ochoa et al., 2020). Although the definition of sustainability has changed throughout history, the definitions in this report constitute the origin of urban sustainability, and the global call made in this report has increased awareness and accelerated the work in the field of urban sustainability (Daly, 1990). Many organizations and institutions have set strategies and targets for sustainable development. However, as stated in the Brundtland report, assessment methodologies, which monitor the current situation regarding to different aspects of urban sustainability, are needed to evaluate strategies and to set new targets (Corredor-Ochoa et al., 2020). There are several sustainability assessment methodologies, for example, Environmental Sustainability Index (ESI) evaluates countries on a global scale in the context of the environmental pillar of sustainability (Mori & Christodoulou, 2012).

This dissertation will consist of two stages: Theoretical and methodological. At the theoretical stage, urban sustainability and its environmental pillar will be reviewed from a historical perspective. Environmental sustainability assessment methodologies will then be examined and associated with the relevant policy. At the methodological stage, the environmental sustainability performance of a selected region will be measure and evaluated by using the composite indicator method. In other words, the aim to construct composite framework that can measure the environmental sustainability performance of selected urban regions with the help of relevant datasets. The following questions will be answered:

- What factors promote environmental sustainability and how can environmental indicators affect sustainability?
- In order to measure environmental sustainability performance of cities, is it possible to get benefit from open data, and is it possible to create a composite framework which shows environmental performance effectively?
- Can such a framework assist decision-makers to provide better policy?

1.2 Research Outline

In this research, the issue of environmental sustainability will be approached both theoretically and methodologically. While chapter 2, 3 and 4 are more related to literature review as can be considered as theoretical part, chapter 5 and 6 contains methodological approaches.

Chapter 2 – while focusing on the definition of urban sustainability, it will also be discussed how the definition and content of urban sustainability have altered over the course of history. This section will also contain theoretical information on why and where urban sustainability emerged. Environmental, social, and economic pillars of sustainability will be also under discussion.

Chapter 3 – The concept of the environmental sustainability will be focused and, in this context, different assessment methodologies such as *environment in general, LCA and composite indicators* will be reviewed from a historical perspective. In addition, the composite indicator will be described in detail and assessment examples applied using the composite indicator method at different scales such as international, national and regional will be examined.

Chapter 4 – The relationship between assessment methodologies and policy regarding to sustainability will be discussed. The gaps in the literature and the limitations in the methodologies will also be discussed and the motivation of the study will be described.

Chapter 5 – The composite indicator methodology will be implemented for the selected region. 10 steps described in the OECD's handbook for constructing composite indicator will be followed.

Chapter 6 – The results obtained as a result of the applied methodology will be reviewed and visualised.

Chapter 7 – The research will be summarized with the overall findings.

1.3 Ethical Consideration

Since the purpose of this study is to measure environmental sustainability by using already publicly available data, all data used in the research are publicly available and can be found in London Datastore. In addition, no data contains information that can harm any institution or person, therefore there is no vulnerability issue in this research in terms of data usage. On the other hand, it should be noted that the index, which emerges as a result of the research and indicates environmental sustainability performance of selected region, may contain inaccuracies regarding to methodology and data used. For this reason, using the result alone may lead to misdirection.

Chapter 2 - Urban Sustainability Term and Its History

The straightforward and general explanation of the sustainability is the ability of any defined phenomenon or behaviour to maintain indefinitely (Glavič & Lukman, 2007). Unlike the concept of sustainability, urban sustainability which is a branch of sustainability aims to increase well-being of humankind and the planet in the long run, rather than continuing it indefinitely as it is impossible (National Academies of Sciences, Engineering, 2016). As can be understood from its basic definition and its own contradictions, the concept of sustainability is a wide, multifaceted, and multiscale topic. For this reason, it should be noted to what extent the concept of urban sustainability is covered in this research. To define the extent and to understand the definition in a clear way, it is crucial to review the roots of urban sustainability from a historical perspective. In this chapter, urban sustainability and its components will be defined by examining the historical evolution of the concept of sustainability.

As a result of industrial revolution, the world entered a new era. Expanding population, mass production, urbanization, and the exceeding consumption and depleting natural resources are the key elements of this new era (More & More, 2002). Cities had been an important part of socio-economic development, but there is no doubt that they also brought with them various environmental problems which could lead to the end of human civilization or even the ecosystem (Wu, 2014). The fact that this situation ("industrial way of life") is not sustainable first mentioned in a report called *A Blueprint for Survival* in 1972 published before the UN Conference on the Human Environment (Basiago, 1996). In UN conference, first global action plan was created to solve "the present problems for the preservation of the environment" with adequate policies and measures (Vasseur, 1973). Framework for environmental action plan also consisted of several recommendations under three types of action showed in diagram: Environmental assessment, environmental management, and supporting measures (Vasseur, 1973). Since it is the first comprehensive action plan on the urban sustainability as global issue, it has guided following research (Corredor-Ochoa et al., 2020). Although this action plan addressed environmental problems and makes recommendations to solve these problems, it did not clearly frame urban sustainability and sustainable development. This terminology gap has been successfully filled in the Brundtland report in 1987 (Kaur & Garg, 2019).

Before mentioning the Brundtland report, it would be useful to mention some studies that unintentionally form the basis of the concept of urban sustainability in terms of theory and methodology. These are studies that reflect on how sustainable cities can be without a holistic approach to urban sustainability. In terms of urban planning and urban form, Howard mentioned about a *garden city* which harmonized town and country, Wright and Geddes created an *organic form* intertwined with nature for the sustainable city, Canfield attempted to design city that respects the nature's carrying capacity, Lyle and Corbett tried to benefit from *green infrastructure, solar power* in a passive way*, natural drainage* (Howard, 1902) (Wright, 1935) (McHarg, 1969) (Canfield, 1993) (Corbett and Corbett, 1984). These are the attempts to achieve sustainability regarding to urban form (Basiago, 1996). In terms of academic literature, in Silent Spring, Carson mentioned about environment and its conservation and highlighted the danger of urbanization. As a result, she increased public awareness of sustainability issue and contributed to the academic literature by attracting the attention of global institutions in this field (Paull, 2013). In addition, in *The Limits of Growth* (1972), sustainability is implied by questioning inhabitant's presence on the world that is defined as "simply not ample enough nor generous enough to accommodate much linger such egocentric and conflictive behaviour by its inhabitants" (Meadows, Meadows, Randers, & Behrens, 1972, p.192) (Vojnovic, 2014). And, it is suggested that sustainability could be provided by achieving socio-ecological equilibrium (Vojnovic, 2014). But neither these studies nor the UN conference in 1973 approached urban sustainability as holistically and globally as in the Brundtland report.

In 1987, The UN's WCED (World Commission on Environment and Development) published the report called *Our Common Future* which is also known as Brundtland report. It contributed "a political opening for the proper concept of sustainable development to evolve" (Daly, 1990). The report developed series of clear definition for the terms such as urban sustainability, urban ecosystem, different aspects of sustainability and more importantly sustainable development (Basiago, 1996). Sustainable development introduced in the

WCED's report: "Humanity has the ability to make development sustainable to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs" (UN, 1987, 15). In addition, it provided the adequate link between three aspects of the sustainability to support urban sustainability: Economic, social and environmental (Corredor-Ochoa et al., 2020). The report organised a global action call with solutions for the need to change the behaviour of institution and policies to solve mentioned environmental problems (Vojnovic, 2014). Action call and detailed policy framework provided by WCED has triggered raising urban sustainability to the top of the agenda of politicians and scientist (Sharifi, 2021). Therefore, Brundtland report is seen as an origin point in sustainable development as it provides background information and framework to many studies. The most important of these studies is the largest environmental conference ever held by UN (UNCED) in Rio de Janeiro (Vojnovic, 2014).

In the light of historical information, the terms to be used in this research can be properly defined. From the existing literature mentioned above, the most appropriate definition of urban sustainability, discussed by Newman (1999), could be:

> *promote and enable the long-term well-being of people and the planet, through efficient use of natural resources and production of wastes within a city region while simultaneously improving its livability, through social amenities, economic opportunity, and health, so that it can better fit within the capacities of local, regional, and global ecosystems,*
>
> (National Academies of Sciences, Engineering, 2016)

The definition is derived from the main principles to promote urban sustainability mentioned in both UN Conference on the Human Environment and Brundtland report (Vasseur, 1973) (UN, 1987). According to this definition, urban sustainability is a desired state that refer to series of dynamic conditions which meet the requirements of current and future (Wu, 2014). To achieve urban sustainability, Brundtland report focuses on the terms sustainable development and sustainable growth. As mentioned in the report:

> *"sustainable development is a process of change in which the exploitation of resources, the direction of investments, the orientation of technological devel- opment; and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspi- rations"*
>
> (UN, 1987)

In other words, sustainable development is a process that sustainability can be achieved (Wu, 2014). While sustainable development is qualitative enhancement, sustainable growth is quantitative increase regarding to physical state (Daly, 1990). As there is strong relationship between these terms, without sustainable development, it is not possible to achieve urban sustainability and growth (Daly, 1990).

While Brundtland report focuses on six elements to review in order to accomplish urban sustainability (population and human resources, food security, species and ecosystems, energy, industry), urban sustainability has been classified into three pillars as the scope of sustainability has expanded over time (Corredor-Ochoa et al., 2020) (Purvis et al., 2019). Another reason for this classification is that Brundtland's definition of sustainability is too abstract for decision-makers in terms of planning and management (Fiksel et al., 2012). Therefore, the need for functional definition of sustainability for decision-makers have resulted in classifying sustainability into three pillars (Figure 2.1): Environmental, economic and social (Fiksel et al., 2012). To define it simply, social sustainability is the ability of the social systems to remain at a threshold level of well-being. Wars, injustice, and poverties can be examples of social sustainability indicators. Economic sustainability is the ability of an economy to remain a threshold level of production. Similarly, environmental sustainability is the ability of the environment to stand on a defined level of environmental balance and natural resources (Purvis et al., 2019). There is a clearly dialectical relationship between these pillars. The fact that one of them is not at the defined level negatively affects the others (Purvis et al., 2019). Therefore, organizations must consider all of them when mentioning or assessing urban sustainability, however, they can put more emphasis -trade-offs- on one. For example, while WTO and UN focuses mostly on the economic pillar due to the request of its members, OECD focuses both economic and social pillars (Purvis et al., 2019). Since the economic pillar is prioritized by the institutions and followed by the social pillar, the fact that environmental pillar, which is relatively more affected by them, is not prioritized creates a gap (Purvis et al., 2019). Therefore, in this research, the environmental aspect will be emphasized more.

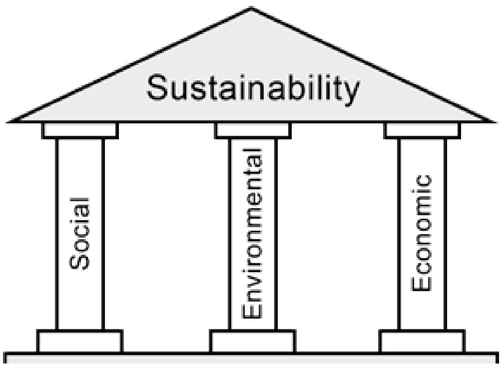


Figure 2.1: Three Pillars of Sustainability. Adapted from (Purvis et al., 2019).

Appropriate policies must be made to achieve sustainability in all three pillars. As the UN Conference in 1972 have emphasized, proper policies can only be made when the current situation is followed and monitored properly (Figure 2.2.) (Vasseur, 1973). For this reason, methodologies have been developed to monitor and assess the current situation in terms of sustainability manner in the historical process.

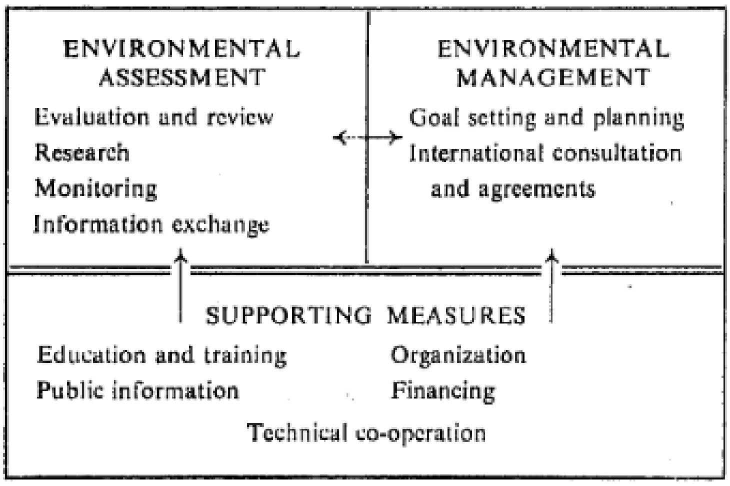


Figure 2.2: The Framework of the Action Plan. Adapted from (Vasseur, 1973).

Chapter 3 - Environmental Sustainability Assessment Methodologies

As explained in Chapter 2, the definition of sustainability developed cumulatively over time. As this definition evolved, evaluating sustainability emerged spontaneously in the context of the literature. Various approaches have emerged on the sustainability assessment in the international arena, with the influence of the Brundtland report. In this context, the sustainability assessment has been seen in two different contexts both in methodological and theoretical fields. The first is to control how sustainable the city is, and the second is to measure the impact of the actions or policies to be taken on urban sustainability (Adinyira et al., 2007). In this chapter, different types of sustainability assessments throughout history will be examined in terms of methodological foundation and scale.

3.1 Assessment Methodologies

As mentioned in Chapter 2, prior to the publication of the Brundtland report, there were studies that indicated that factors such as population explosion, mass production and overuse of natural resources cause several environmental problems such as pollution and adverse effects on biodiversity. As a result of these studies, environmental aspect of urban sustainability is measured generally using traditional ways such as predefined checklists or matrices and several methods such as logical framework, cost-effectiveness analysis and multi-criteria evaluations (Adinyira et al., 2007). This kind of measurement methods are called “environment in general” (Adinyira et al., 2007). With the effect of UN Conference in 1973, values about resource consumption and pollution are mostly considered in this methodology. In addition, due to the studies that attempt to achieve ecological development by changing urban form such as *garden city* are also given importance. Due to its limited capacity and the availability of more advanced methods, environment in general methodology is not used today (Adinyira et al., 2007).

With the influence of the UN Conference in Rio de Janeiro, where the Brundtland report was also discussed, we can see that a new sustainability assessment method called life cycle assessment (LCA), which is more advanced than traditional “environment in general” method, has emerged (Bond et al., 2012). LCA is basically a tool evaluate the potential environmental impacts of products or services at all stages in their life cycle. Compared to traditional method, LCA considers not only environmental issues but also social and economic issues (Adinyira et al., 2007). However, this method also has some weakness in processing complex and large data about sustainability issues. Therefore, although it can evaluate different aspects of sustainability, it fails when it comes to integrating them (Adinyira et al., 2007).

Since LCA and environment in general methods could not assess urban sustainability in a holistic and cumulative way regarding to three pillars of environment, a new method was needed. While method called environment in general focuses only environmental aspect of sustainability due to the lack of term sustainability in the academic literature, LCA method tries to evaluate three pillars separately. However, decision-makers need to evaluate all aspects of sustainability in an integrated way to make more effective decisions. Therefore, new methodology called sustainability indicators that benefit from wide variety of indicators have been emerged (Adinyira et al., 2007). With significant indicators and as a result an effective feedback mechanism, policy makers can easily understand and quantify and monitor the cumulative effects of policies. This method also gives flexibility to institutions to focus on desired pillar of urban sustainability, as it allows freedom in the selection of extensive list of indicators (Adinyira et al., 2007). There are many studies indicating that this method has been applied at different scales such as international, urban and regional (Alberti, 1996). But before diving into these studies, it would be useful to look at composite indicator terminology which is a popular way of integrating different types of sustainable indicator.

3.2 Composite Indicator (CI)

Composite indicators, also called composite indices, has become popular recently, with some institutions adopting it in their research. This has attracted the attention of many researchers and policy makers, as has led to more use of this methodology. In terms of conceptual definition, composite indicators are "[…] based on sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting these sub-indicators" (Saisana and Tarantola 2002, p. 5). Technical definition could be a mathematical aggregation of a group of sub-indicators (Greco et al., 2019). CI method helps researchers and decision-makers compile and summarize complicated and multifaceted problems, it also helps users such as citizens to understand the problem in a clear way (El Gibari et al., 2019).

If we look at the practical examples of this field, CI is used to measure and evaluate "county's competitiveness (World Economic Forum), the quality of its governance (World Justice Project), the freedom of its press (Freedom House), the global, regional and national Human Development (The United Nations Development Programme) and […] etc" (El Gibari et al., 2019). There are also several limitations of this methodology. For example, it may misguide and non-robust policy actions, however, this problem can be solved by evaluating sensitivity analysis for robustness (Greco et al., 2019). In addition, it may force politicians to get simple and casual conclusions (Greco et al., 2019). Since it excludes qualitative data, it can only examine quantitative data (Greco et al., 2019). In order for the methodology not to give misleading results, sub-indicators should be carefully selected and treated.

3.3 Methodologies in terms of scale

As mentioned above, there are studies in which environmental sustainability is evaluated at different scales. Examining evaluation studies at different scales using different methods is important in terms of understanding the gaps and limits of sustainability assessment methodologies. On an international scale, there are two effective studies that are both intertwined with policy and focusing on environmental pillar of sustainability: Environmental Sustainability Index (ESI) and Environmental Performance Index (EPI). (Singh et al., 2009). There are indeed more studies that focus on sustainability assessment on a global scale such as Innovation Index, Index of sustainable and economic welfare and Living Planet Index etc (Singh et al., 2009). However, as the scope of this research is limited to the environmental pillar of sustainability, there is no need to mention the others.

When the first version of the ESI was published in 2000, there were no other sources assessing environmental sustainability on a global scale, therefore, it can be accepted as an origin for an environmental assessment on a global scale (National & Stewardship, 2005). ESI integrate 76 different datasets into 21 indicators of environmental sustainability which help to make comparison across five broad categories. These are "environmental systems, reducing environmental stresses, reducing human vulnerability to environmental stresses, societal and institutional capacity to respond to environmental challenges and global stewardship" (National & Stewardship, 2005). Although it is aimed to evaluate the countries under these categories and make comparisons between them, it is obvious that the results obtained from this study are too broad and abstract for policies as declared in EPI (Yale Center for Environmental Law & Policy, 2006). In response's to ESI's weak link with policy, EPI aimed to fill this gap by organizing the indicators into 6 policy categories (Yale Center for Environmental Law & Policy, 2006). As a result, each country is indexed with a certain score, which give better sustainability insights to policy makers in international and national context, with the help of selected environmental sustainability indicators as described in the methodology of EPI (Yale Center for Environmental Law & Policy, 2006).

On the building and its environment scale, Building Research Establishment Environmental Assessment Method (BREAAM) and Leadership in Energy and Environmental Design (LEED) could be a good and effective example (Kaur & Garg, 2019). These studies focus on the environmental assessment of existing or under construction urban projects (Ameen et al., 2015). BREEAM published in 1990 uses composite indicator method to evaluate sustainability values of the building that cover a range of environmental issues at every stage. These environmental issues are categorised such as energy consumption, water usage, waste management, land use, transportation, health, and well-being. The assessment process of BREEAM results in certification which may also help to raise public awareness of environmental problem (Kaur & Garg, 2019). Although LEED's methodological approach is similar to BREEAM, the main differences are weight of environmental indicators used and certification systems (Ameen et al., 2015). While LEED bases its own threshold on percentages, BREEAM has its own quantitative standards and puts more weight on the construction stage of building (Ameen et al., 2015).

While important studies such as EPI and ESI contribute to environmental assessment on a global scale, certification methods such as LEED and BREEAM contribute to the building environment scale. However, when we look at the regional scale, it is obvious that there is a systematic research gap and confusion (Cohen, 2017). As declared in Cohen's systematic literature review of urban sustainability assessment, environmental assessment studies on global scale could be insufficient for actions and policies to be taken on regional scale (Cohen, 2017). Cohen examined 69 studies aimed to fill this deficiency. From the literature, various methods such as sustainability indicator, urban carrying capacity, asset-based framework, urban form and etc are used in these studies (Cohen, 2017). However, the majority of these studies are limited to the socio-environmental context which could be considered as abstract regarding to decision-making process rather than environmental sustainability itself. In addition, most of these studies are limited in the diversity of selected region (Cohen, 2017). Therefore, the need for a holistic approach to environmental sustainability assessment at the region scale is inevitable.

Chapter 4 - Policy View

4.1 Link to Policy

So far, evolution of sustainability definitions has been discussed in Chapter 2 and sustainability assessment in terms of methodology and scale has been reviewed in Chapter 3.

At this point, it is crucial to explain how assessments and indicators can be linked to policy and its implementations. The first link to policy within the scope of sustainability was made in the Brundtlandt report with the term sustainable development which refers common issues and challenges discussed in the “Our Common Future” (Shields et al., 2002). After the UN’s approach to sustainable policy, sustainability goals, strategies and principles are enhanced (Vojnovic, 2014). As sustainable development is a dynamic and cyclical process, objectives and strategies have changed and improved over time with international studies. Therefore, instead of examining objectives, it is more important to examine how the ever-changing systems connects with the methodology.

Policy makers will be tended to make more efficient and meaningful strategies-goals on the off chance that they comprehend the consequence of selecting strategies. In addition to this fact, sustainability development is data driven issue. Therefore, it is clear to realize the importance of indices and indicators as they give comprehensive insights to decision makers while defining sustainability policies. Resource management which is the model frequently referred by policy makers may explain hierarchy behind data-driven systems. Values are put at the top of the hierarchical triangle. Those values impact and are affected by society. Dialectic relationship between society and values results in defining objectives which also influence to make actions and their impact. As a result, a feedback loop between values and systems through control and information flow is established (Shields and Mitchell, 1997). In this hierarchy (Figure 4.1), assessment methodologies play the role at the values level by processing primary data and presenting it as an indicator or indices. In other words, assessment methods help not only to determine objectives and acts but also review and adaptation as a feedback process.

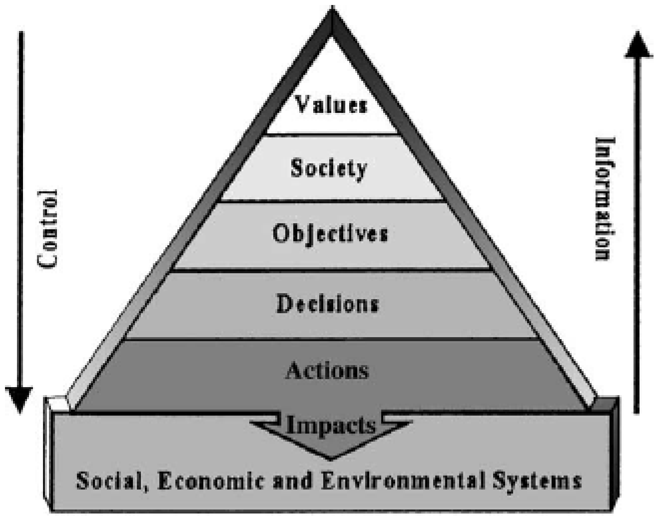


Figure 4.1 Control and information flow – Hierarchical Model of Resource Management. Adapted from (Shields et al., 2002).

4.2 Motivation of The Study

Although international assessment studies such as EPI and ESI are successful and innovative in terms of the quality of work and policy side, there are also several limitations. Firstly, as mentioned in the EPI report, lack of data and poor quality of data are one of the problems regarding to assessment framework (Mori & Christodoulou, 2012). Since the resulting composite framework is naturally related to create competitive indices among countries, this poor-quality data problem may affect all countries in the study as threshold level is decided according to all data (Yale Center for Environmental Law & Policy, 2006). Therefore, this limitation may cause to mislead the decision-makers and public while taking actions and codifying laws. Secondly, each country is characteristic as they have different properties. Therefore, indicator selection and weighting approach for indicator should be carefully designed. However, ESI gives equal weight to each indicator. On the contrary, EPI contributes more detailed weighting approach to indicators (Yale Center for Environmental Law & Policy, 2006). However, this contribution remains still vague and abstract in terms of different scales for sustainability. In addition, for example, equal weighted overfishing data used in the ESI may give undesirable result as some countries have no problem with it (Yale Center for Environmental Law & Policy, 2006).

Thirdly, although UN's and EU's approaches to define objectives and goals are effective to achieve environmental sustainability and these objectives are adopted to different scales such as national and regional level, there is clearly research gaps in terms of policy adjustment when it comes to measurement of actions taken on the regional scale. When environmental sustainability can be achieved on an international scale, it may not be achieved at small scales (Yigitcanlar et al., 2015). The opposite of this situation could also work. Because, for example, although natural carrying capacity data give good results on global scale, some cities may exceed the carrying capacity limit and some may be far below the limit. Therefore, it is inevitable that there is need to assess environmental sustainability on the lower scales. While there are effective evaluation methods at the building environment scale such as BREEAM and LEED, assessment issue is problematic at the urban scale (city itself and its regions). In recent studies of UN and OECD, it is seen that big cities are also taken into consideration of assessment (Morelli, 2011). However, there is still assessment gap regarding to cities regions. Most of city councils consult several NGOs and consultation firms for different aspects of environmental sustainability. However, there is need for holistic approach to all aspect of environmental sustainability in order to give comprehensive guide to policy makers at the urban level.

When we look at the environmental sustainability issue at the urban level with the principle of "think globally, act locally" and consider research gaps in terms of policy view, composite indices methodology to assess the actions and redefine objectives for environmental sustainability at the urban level is crucial. Therefore, in this study, selected urban region and its components will be examined by compositing indicator selected and weighted carefully for that region.

Chapter 5 - Methodology

As defined in Chapter 3, composite indicator is a value aggregated from individual datasets. In other words, it is useful and informative summary of multiple dataset that can decrease information flow to policy-makers and public (Jacobs et al., 2004). When it comes to environmental sustainability development, composite indicator has many advantages regarding to policy view as it is providing a more simplistic results of assessment (Marazza et al., 2011). In this research, a single index presenting environmental sustainability score will be provided at the urban scale with the help of composite indicator method. Comprehensive guide of OECD's *handbook on construction composite indicators* will help to construct the methodology step by step. The methodology will be like ESI and EPI, however, indicator selection and aggregation approach will be reviewed as scale of research is different. In addition, the construction of aggregated index is complicated issue and if not built properly, it can misguide the policy and public (Jacobs et al., 2004). Therefore, while following OECD's steps, different pathways will be applied and reviewed.

5.1 Urban Region Selection for Analysis

There are two main factors for urban region selection. Firstly, selected region should include development plan and strategy for environmental sustainability so that the composite indicator score provided can assess the strategy and its actions. As discussed in Chapter 4, assessment of environmental sustainability can help policy makers in two way: Assessing existing strategies and defining new objectives according to the composite results (Shields et al., 2002)

In addition, existing strategies and objectives can guide the indicator selection process for the composite indicator. Second factor is that the selected area should have good quality of information flows, in other words the data availability is necessary. Nowadays, many developed cities have already good quality data about the context. However, it is necessary that these data can be integrated to lower urban scales so that lower scales in the selected region can be assessed as it is the aim of this research. Therefore, Greater London, which provides data up to lower scales and has detailed strategy plan within the scope of environmental sustainability, is preferred.

5.2 Developing a Theoretical Framework

Theoretical framework is an essential part of the methodology as it provides clear definition of the phenomenon to be assessed on a specific dimension (Nardo et al., 2005). As aim of this research mentioned above, environmental sustainability of Greater London's regions in a competitive manner is what is desirable to measure. To measure, it is important to define sub-components and indicators. While ESI's categorized data is referred for providing guideline to construct domains, *London Environmental Strategy and Implementation Plan* (LESIP) published in 2018 assists to find which data is necessary for measurement. LESIP's approach gathers all aspects of environment and categorised the following sub-components:

- *"Air quality"*
- *"Green Infrastructure"*
- *"Climate Change Mitigation and Energy"*
- *"Waste"*
- *"Adapting to Climate Change"*

- *"Ambient Noise"*
- *"Low Carbon Circular Economy"*

  (Greater London Authority, 2018)

Under these categories, objectives such as clean energy, zero waste city, more green infrastructure, better air quality etc. are set. Data will be selecting according to goals defined in this policy document. On the other hand, ESI's 21 indicators and 5 sub-components are also useful, however, as it measures each nation's environmental sustainability some indicators may not be effective. For example, several variables in global stewardship and social-institutional capacity components will not work at the regional scale if these variables are same for all nation or London. Therefore, components for theoretical framework will be defined in the data selection stage.

5.3 Data Selection

As mentioned in OECD's handbook: "The strengths and weaknesses of composite indicators largely derive from the quality of the underlying variables" (Nardo et al., 2005). For this reason, data selection is one of the building blocks of the methodology. Although data selection is left to the researcher's perspective on the subject, it should follow the theoretical structure (Jacobs et al., 2004). The intended theoretical structure is to follow LESIP as much as possible. However, in this policy document, since each component stated above is evaluated separately, ESI's assistance is sought in finding the domains needed for a comprehensive measurement. At this stage, theoretical structure remains unclear without testing the availability of data. Therefore, domains will take shape after examining the available publicly open data. Proxy variables can be also reviewed when the requested data is unavailable (Nardo et al., 2005).

5.3.1 Air Quality

It is necessary to increase air quality by minimizing air pollution concentrations in order to protect human and environmental health (Greater London Authority, 2018). Air concentrations datasets are derived from London Atmospheric Emissions by GLA and TFL. The publicly open dataset includes ground level concentrations of annual mean NO_x, NO_2, PM_{10} and $PM_{2.5}$ in $\mu g/m^3$ lastly updated in 2016 (GLA, 2016). 20m grid resolution data provided in GIS format has been converted to ward scale by taking mean concentrations within each ward with the help of QGIS Application. High concentrations value indicates poor air quality and lower air quality means lower value for environmental sustainability (Button, 2002). While objectives and strategies about air quality as describe in separate chapter in LEPIS, ESI has placed the air quality indicator under the environmental systems domain.

5.3.2 Land Usage

In ESI document, land usage indicator which is under the environmental system domain are divided into separate variables such as wilderness area, developed area and water quantity (Yale Center for Environmental Law & Policy, 2006). On the other hand, in LEPIS, only green infrastructure issue is mentioned in an objective manner as using the full potential of green spaces to enhance people's quality of life. Available data found in London Green and Blue Cover dataset which provide the percentage of green cover value including the city's park, gardens, trees, green spaces and green roofs, and the percentage of blue cover value including rivers and wetlands (GLA, 2019). As it is not appropriate to talk about wilderness in urban area, this variable in ESI is neglected. In an environmental sustainability manner, more percentage values of both green and blue cover mean more sustainability value and this land usage indicator could be placed under environmental system domain.

5.3.3 Ambient Temperature

Due to the urban heat island effect, air and surface temperature increase in urban areas more than rural parts. (Wu, 2014). High temperatures due to probably relatively large urban spaces and anthropogenic heat sources maybe uncomfortable in terms of human health and environmental life as describe in the adapting to climate change chapter of LESIP (Greater London Authority, 2018). Therefore, more ground and air temperature values mean less environmental sustainability score. Relevant data found in London's Urban Heat Island datasets (GLA, 2013). Average temperature across the summer period value provided in a high-resolution shapefile format is converted into ward scale by calculating mean values within each ward with the help of QGIS tools. This variable could be also considered as a part of environmental system and therefore placed under this domain.

5.3.4 Population Density and Fertility Rate

Although human population value is more related to social pillar, it is inevitable that urbanization variables such as population density and fertility rate can dominantly affect environmental side of sustainability (Rees & Wackernagel, 2008). ESI defines the impact of human factors like population growth on the environment as environmental stresses and places these factors under reducing environmental stress domain (Yale Center for Environmental Law & Policy, 2006). Ward Profiles and Atlas datasets include population density per sqm and general fertility rate (GLA, 2014).

5.3.5 Waste and Recycle Rate

In LESIP's chapter 6, waste management strategy is issued as a big impact on the environment, and it aims to reduce the amount of waste year by year in order to transform London into a zero-waste city (Greater London Authority, 2018). Additionally, ESI use waste recycling rates and waste generation variables to measure waste and consumption pressures indicator which is under the reducing environmental stresses domain (Yale Center for Environmental Law & Policy, 2006). Unfortunately, relevant datasets are not available at the ward scale. Instead, as proxy measures, waste generated per head and household recycled rate datasets at borough scale are derived from the waste reduction and recycling plans for each borough by scanning them separately.

5.3.6 Ecological and Carbon Footprint

Ecological footprint and carbon footprint are an indices that attempt to measure external impacts on environment such as the total consumption of goods and services (Mori & Christodoulou, 2012). Therefore, it is an essential part of environmental sustainability. In ESI, ecological footprint is categorised under the same heading as waste management (Yale Center for Environmental Law & Policy, 2006). In LESIP, this issue is mentioned in the low carbon circular economy section. In London Datastore, environmental footprint datasets include ecological and carbon footprint values provided ktoe per capita at borough scale. Although the data is not available at ward scale, it can still provide better insight to compare boroughs regarding to environmental sustainability

5.3.7 Vehicles in Use

Composite indicators can have both output and input variables simultaneously (Nardo et al., 2005). These variables can correlate with each other and therefore this situation may result in the problem called "double counting of element" (ibid.). Without having multivariate analysis, one can easily understand that vehicles in use and relevant datasets such as air quality and footprint data are related to each other. In order to eliminate the double counting problem, several weighting approaches are developed which will be reviewed in next sections. The main reason behind the usage of this data from policy document. As mentioned in chapter 9, low and ultra-low carbon zones strategy, which basically limits fuel vehicle traffic in defined areas, is implemented in order to minimise the negative effect of vehicles usage (Greater London Authority, 2018). Correlated datasets with vehicles in use include other factors and therefore to monitor the consequences of the strategy vehicles in use datasets intentionally will be used. In ESI, this indicator is associated with the reducing environmental stress domain. Ward Profiles and Atlas datasets have relevant information flow about vehicles in use per household (GLA, 2014).

5.4 Theoretical Framework

The theoretical framework has emerged more clearly when compliance of publicly available data is scanned and checked by referring to the objectives in the London environmental strategy policy document. In addition, the ESI methodology has been also reviewed to decide on the main domains of the composite indicator, and considering the publicly available data, it is appropriate to have two main domains under the name of environmental systems and environmental stresses. Environmental systems consist of air quality variables- PM25, PM10, NOx and NO2, and land usage variables- the percentage of water and green spaces and ambient temperature which can be considered as proxy measurement of urban heat island effect. Environmental stresses include population growth variables- population density and fertility rate, waste and recycle rate, ecological and carbon footprint and finally vehicles in use. However, it is necessary to have multivariate analysis in order to decide the way of combination of these variables.

		Relevant Data Source	Explanation	Policy Reference	Code	Scale	Effect
Environmental Systems Domain	Air Quality (AIR)	NO2 Concentrations from the London Atmospheric Emissions Inventory 2016	ground level concentrations of annual mean NOx, NO2, PM10 and PM2.5 in µg/m3 lastly updated in 2016	London Environment Strategy 2018 (LESIP) – Chapter 3	NO2	20m grid resolution data	Negative
		NOX Concentrations from the London Atmospheric Emissions Inventory 2016			NOX		
		PM25 Concentrations from the London Atmospheric Emissions Inventory 2016			PM25		
		PM10 Concentrations from the London Atmospheric Emissions Inventory 2016			PM10		
	Urban Heat Island	London's Urban Heat Island datasets 2013	Average ambient temperature across the summer period	LESIP 2018 – Chapter 5	TEMP	High Resolution Polygon	Negative
	Land Usage	London Green and Blue Cover dataset	the percentage of green cover value including the city's park, gardens, trees, green spaces and green roofs	LESIP 2018 – Chapter 4	GREEN%	Ward Level Csv Data	Positive
			the percentage of blue cover value including rivers and wetlands	ESI	WATER%		
Environmental Stress Domain	Vehicles in Use	Ward Profiles and Atlas datasets 2013	Vehicles in Use per household	LESIP 2018 – Chapter 5	CAR_USE	Ward Level Csv Data	Negative
	Population Density	Ward Profiles and Atlas datasets 2013	Population Density per sqm	ESI	POPD	Ward Level Csv Data	Negative
	Waste Production	Waste Reduction and Recycling Plans 2018	waste generated per head	LESIP 2018 – Chapter 6	WASTE	Borough Level Csv Data	Negative
	Carbon Footprint	environmental footprint datasets 2018	carbon footprint values provided ktoe per capita	ESI	CARBONF	Borough Level Csv Data	Negative
	Ecological Footprint	environmental footprint datasets 2018	ecological footprint values provided ktoe per capita	ESI	ECOF	Borough Level Csv Data	Negative
	Recycle Rate	Waste Reduction and Recycling Plans 2018	Household Recycle Rate	LESIP 2018 - Chapter 6	RECYR	Borough Level Csv Data	Positive
	Fertility Rate	Ward Profiles and Atlas datasets 2013	General Fertility Rate	ESI	FERTR	Ward Level Csv Data	Negative

Table 5.1: Details of final datasets

5.5 Multivariate Analysis – Initial Data Exploration

It is essential to understand each dataset that will be used in composite indicator before doing deeper analysis. Analyses without looking at the interrelationship between each individual indicator can mislead the decision makers (Nardo et al., 2005). Therefore, initial data exploration may give better insight for normalisation, weighting, and aggregation processes. In addition, in the data exploration part, correlation analysis between individual datasets will be applied in order to avoid redundancy effect.

As demonstrated in the Table 5.1, each individual dataset has different unit. For example, while carbon footprint values are provided in ktoe per capita unit, ambient temperature values are provided in centigrade format. In order to combine individual datasets in a meaningful way, it is necessary to standardize them (Singh et al., 2009). In the normalisation and weighting stages, standardize methods will be discussed.

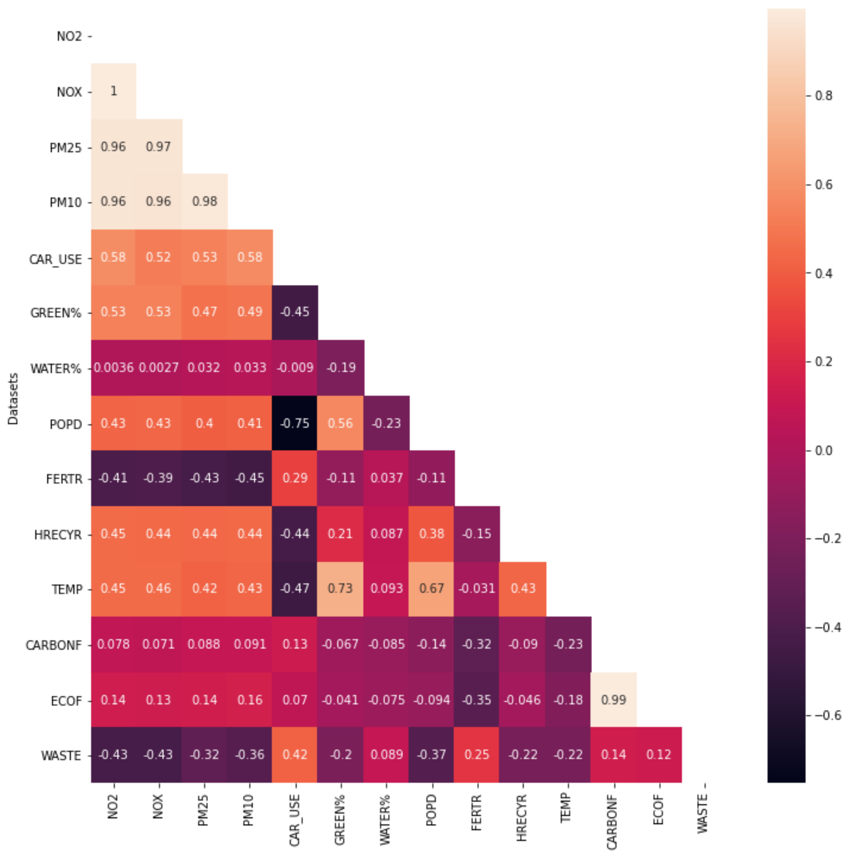


Figure 5.1: Correlation plot for selected datasets

The correlation plot (Figure 5.1) may help to get rid of double counting problem if there is high degree of correlation between variables (Nardo et al., 2005). It can also help to understand the relationship between individual datasets in an environmental sustainability manner. However, before discussing correlation plot, it is important to note that there is always a certain correlation between the variables since environmental sustainability variables include both input and output measures (Nardo et al., 2005). The aim is minimising redundancy by defining threshold correlation level since it could be dangerous to aggregate strongly correlated variables. From the plot, most obvious correlation is among the air quality datasets. For instances, there is a 97 percent correlation between PM10 and NO2. Therefore, PM25, PM10, NO2 and NOX datasets should be combined with each other under the name of air quality indicator without affecting overall result. On the other hand, ecological footprint and carbon footprint datasets are also strongly correlated therefore same procedure should be applied to them. As it is expected, percentage of green space and ambient temperature datasets are negatively correlated. There is also manageable correlation between vehicles in use and air quality datasets, however, as mentioned above vehicles in use datasets intentionally are added to composite indicator in order to see the effect of low carbon zone policies. As a result, as it is expected there is generally a 10 to 40 percent correlation between each dataset excluding air quality and footprints datasets. This expected amount of correlation is manageable.

	NO2	NOX	PM25	PM10	CAR_USE	GREEN	WATER
count	649	649	649	649	649	649	649
mean	33.22	52.2107	11.0704	18.4952	0.842768	26.24256	1.98384
std	4.83177	11.7415	1.275322	0.88111	0.3260056	16.61125	6.0588628
min	23.6	30.6	9	17	0.23	1.5	0.1
25%	29.7	43.7	10	17.9	0.56	14	0.1
50%	32.6	49.9	11	18.3	0.82	22	0.3
75%	35.9	58.4	12	18.9	1.11	34.5	1.2
max	51.2	99.7	16	22.6	1.71	90.1	74.2

	POPD	FERTR	HRECYR	TEMP	CARBONF	ECOF	WASTE
count	649	649	649	649	649	649	649
mean	8703.15	64.3286	34.696	18.2543	12.10008	5.47384	323.515
std	5107.98	15.1846	9.544115	0.3769	0.8900483	0.43398	50.2437
min	189.7	16.5	20	16.4161	10.03	4.51	166.3
25%	4893.9	54	27	18.089	11.66	5.26	305.23
50%	7433.3	64	32	18.3485	12.02	5.39	328.71
75%	11875	74.3	42	18.4977	12.55	5.69	351.2
max	30125	127.5	54	18.702	15.25	7.18	482.65

Table 5.2: Summary Statistics for final datasets

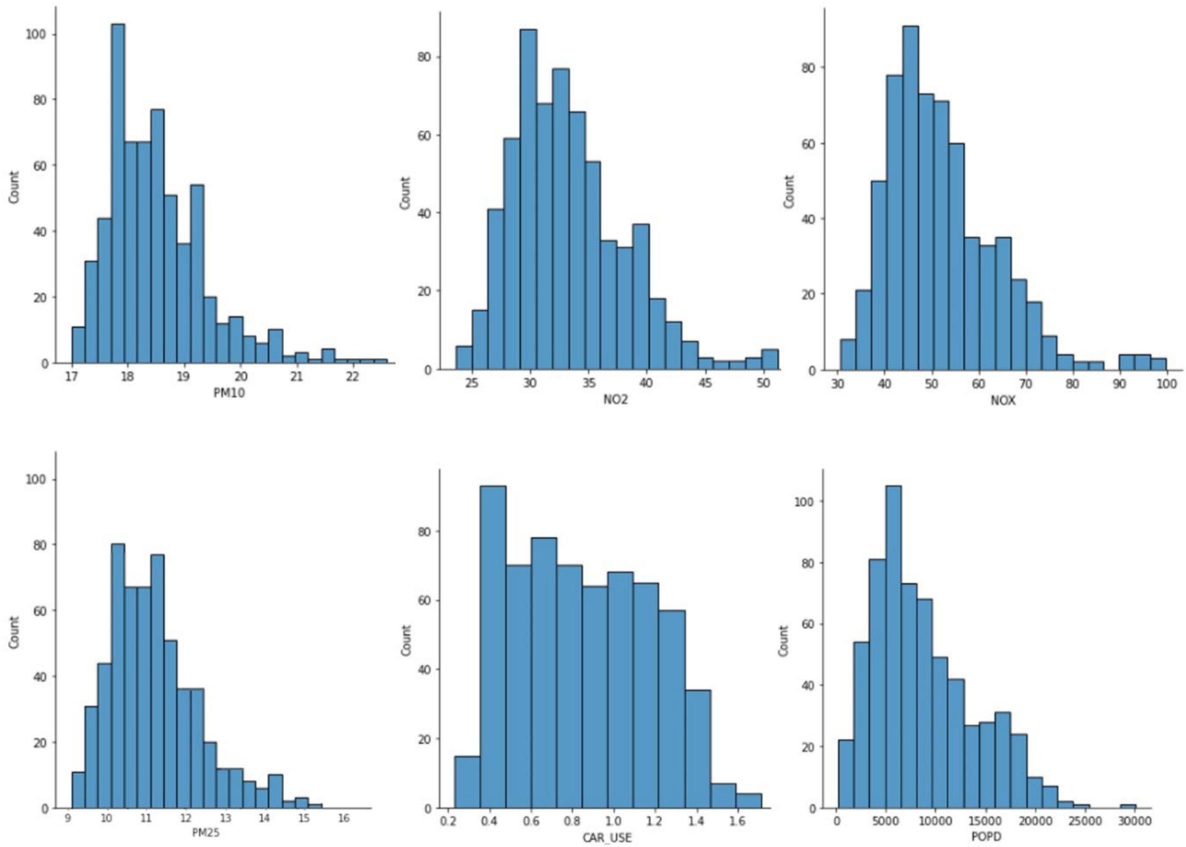


Figure 5.2: Data Distribution Plot for PM10, PM25, NO2, NOX, CAR_USE and POPD

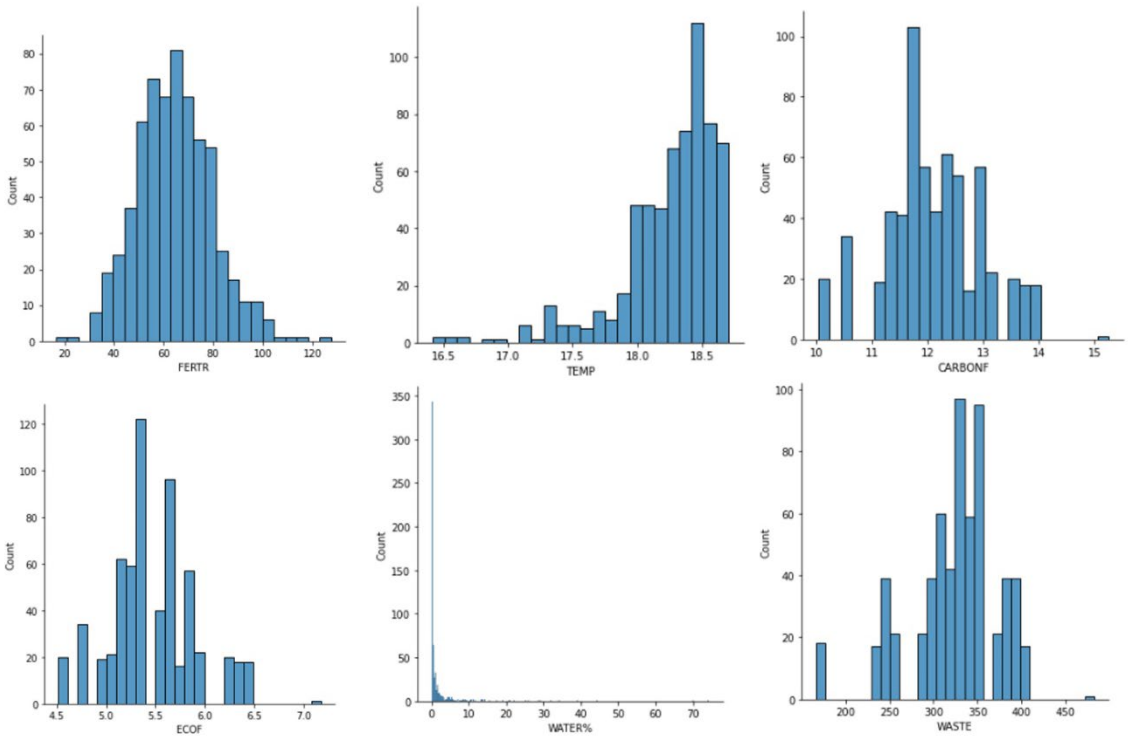


Figure 5.3: Data distribution plot for FERTR, TEMP, CARBONF, ECOF, WATER AND WASTE

5.6 Normalisation

Normalisation process is recommended before any aggregation since each indicator have different unit (Cherchye et al., 2008). There are many normalisation methods when creating a composite indicator, however, in order to select appropriate method it is useful to look at the distribution of datasets. (Nardo et al., 2005). As can be clearly seen from the Figure 5.2 and Figure 5.3, the most problematic data is the percentage of water. This dataset is extremely skewed due to the large amount of zero values. More than 50 percent of the wards have zero value in the blue cover percentage. Therefore, this dataset is excluded since it is not manageable in the normalisation process. Except for water percentage data, there is no extremely skewed distribution and also no extreme outlier. The air quality datasets including PM10, PM25, NOX and NO2 have positively skewed distribution. On the other hand, carbon footprint, ecological footprint, fertility rate and waste datasets are normally distributions with small noises. Ambient temperature is the only data with a negatively skewed distribution.

Although there is no extreme outlier in selected datasets, there are some minor outliers that may affect the composite indicator results. Therefore, in order to de-emphasize these outliers and to fix skewed distribution problems, logarithmic transformation is applied to each dataset. The idea of logarithmic transformation is that "taking the log of the data can restore symmetry to the data" (Metcalf & Casey, 2016).

After reducing the skewness of raw data with logarithmic transformation, standardization process is still necessary as log transformed values have still different measurement units. There are several standardization methods such as ranking, Min-Max and z-scores. In addition, distance to a reference method, which is basically measures relative position of individual indicator according to pivot point, is widely used in composite indicators (Nardo et al., 2005). However, in this case, defining reference point is not applicable as it requires more data in terms of environmental characteristics of each ward. Therefore, the standard scaler is used from the sklearn package which standardizes values by scaling entire dataset with a zero mean and unit variance (Scikit, 2018). This process is also necessary for principal component analysis stage which will be discussed in the next section.

5.7 Weighting and Aggregation

In the framework that make comparative assessment as in this study, the weight given to each individual indicator is of great importance as it significantly affects the overall result (Greco et al., 2019). There are several statistical weighting techniques such as factor analysis, principal component analysis, analytic hierarchy processes, conjoint analysis and most widely used equal weighting (Nardo et al., 2005). Apart from statistical methods, there are also weighting methods based on expert opinion (ibid.). At this stage, different scenarios will be attempted and the most suitable one will be selected for the following process. Prior to look at different scenarios, it is essential to define each method used. Equal weighting can be basically defined as giving equal weight to each variable, although it is generally perceived as no weighting (Gómez-Limón et al., 2020). In composite indicators related to assess sustainability like ESI, equal weighting technique is generally applied. If there is a high correlation between the variables, data reduction techniques such as principal component analysis(PCA) and factor analysis(FA) are used because equal weighting may not give suitable results (Greco et al., 2019). PCA creates one or more index components by using a linear combination of variables and therefore it is useful to gather individual indicators according to their correlation (Nardo et al., 2005).

As mentioned above, correlation between ecological and carbon footprint datasets and also correlation between air quality datasets are quite strong. Before aggregating process, these variables should be grouped to avoid double counting. As a result of PCA for air quality datasets, 4 principal components are created. While the first one has 96 percent explained variation ratio, others have relatively small values considered as noise. Therefore, first component is selected to represent air quality indicator. This situation is observed similarly in carbon and ecological footprint datasets. Two principal components are created and larger one has 99 percent explained variation ratio; therefore, these datasets are grouped called environmental footprint. However, PCA for the green space percentage and ambient temperature is slightly different. Two components are formed, one has 76 percentage of explained variance ratio and other has 24 percentage which is large enough to be underestimated. Therefore, principal components are multiplied by the square root of eigenvalues and summed to generate one indicator called ecological assets. It should be noted that if any individual indicator has negative effect on environmental sustainability, all values of the data are converted to negative. Then values are aligned between 0 and 1 with the help of Min-Max scaler to avoid that negative values may create cancelling issue. At the next stage, the results of PCA and other individual indicators are arithmetically combined to create two main domains: environmental systems and environmental stresses. At the final stage, two main domains are also arithmetically combined to construct composite indicator.

PC: Principal Component	Explained Variation Ratio	Explained Variance
PC1 for Environmental Footprint	0.997	1.997
PC2 for Environmental Footprint	0.003	0.003
PC1 for Air Quality	0.963	3.852
PC2 for Air Quality	0.024	0.096
PC3 for Air Quality	0.008	0.032
PC4 for Air Quality	0.005	0.020
PC1 for Ecological Assets	0.761	1.524
PC2 for Ecological Assets	0.239	0.479

Table 5.3: PCA results for different indicators

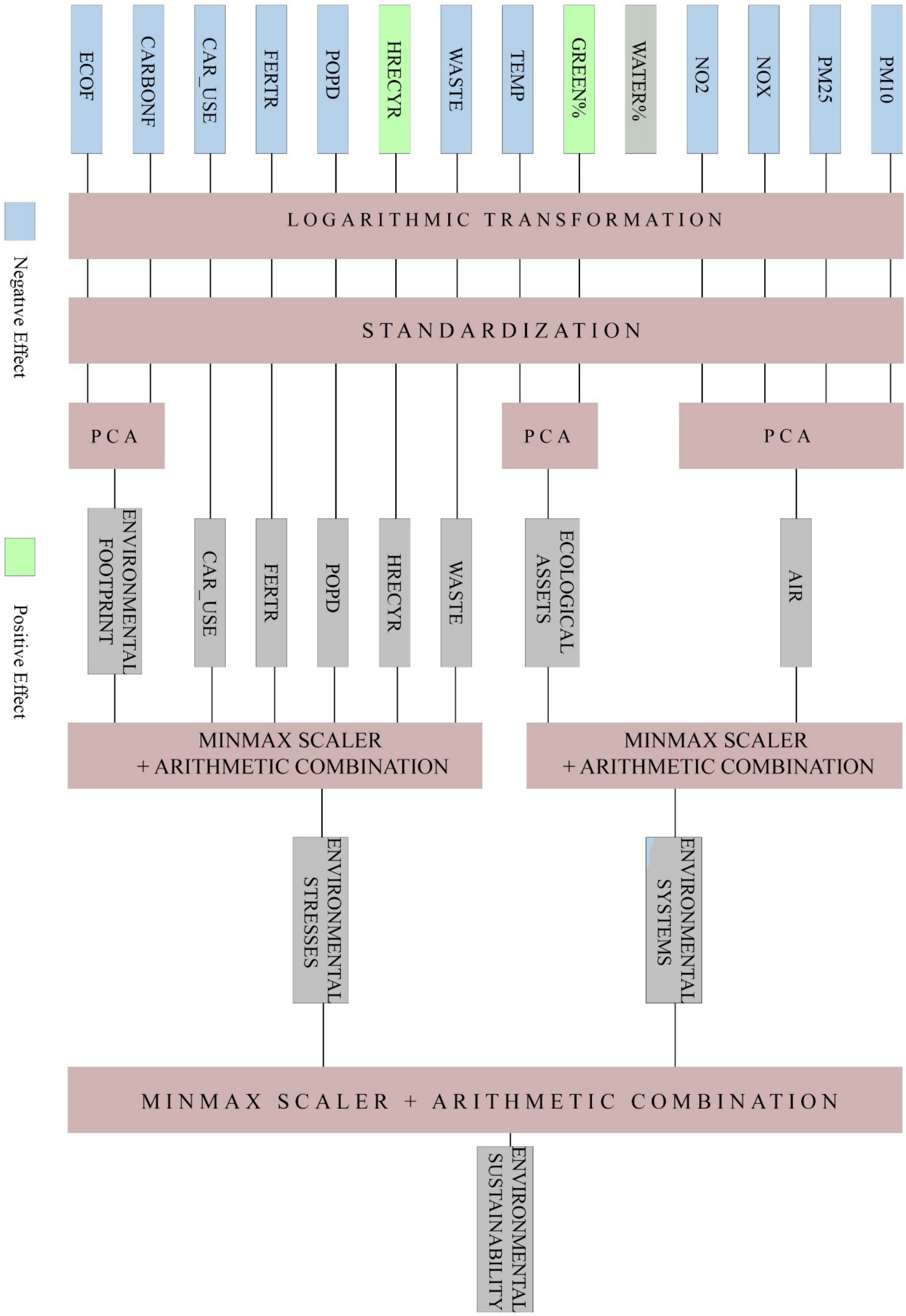


Figure 5.4: Framework diagram of composite indicator

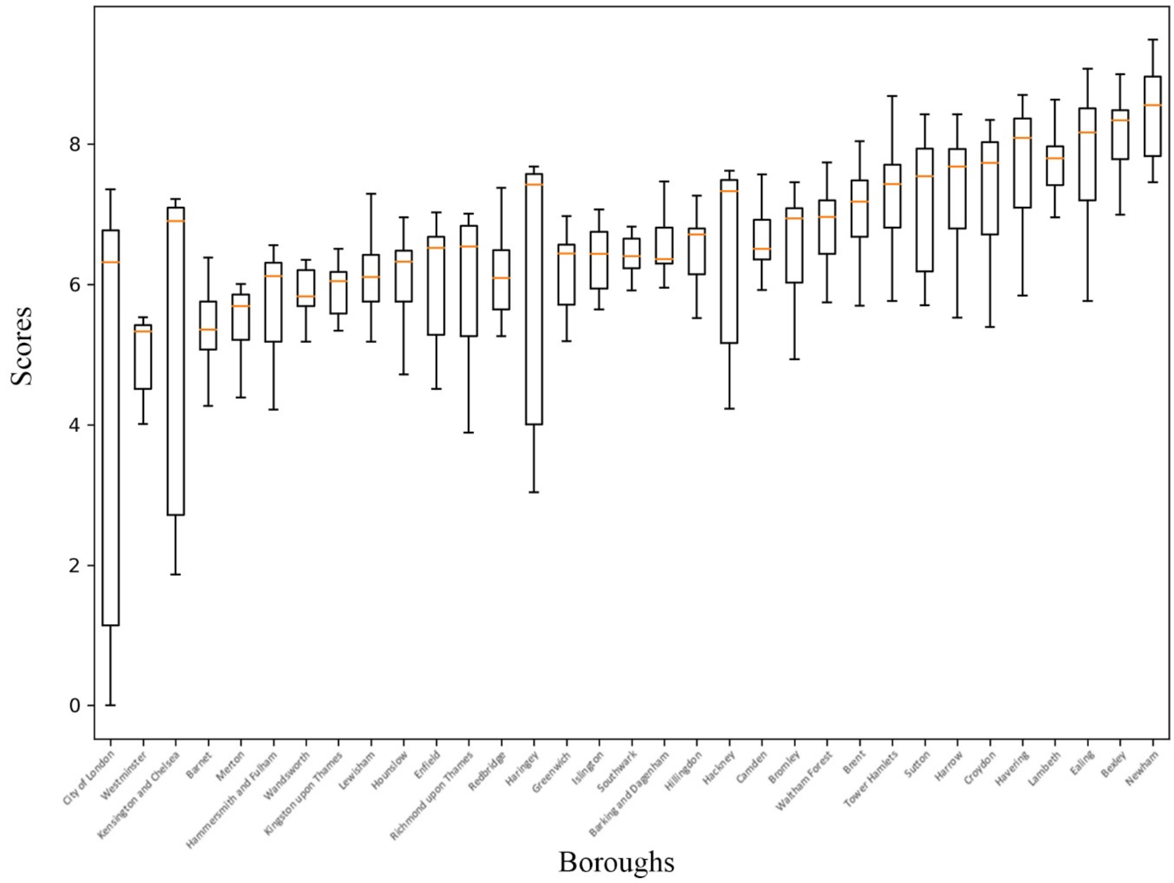
5.8 Uncertainty Analysis

As described in OECD's handbook on constructing composite index, "Uncertainty analysis focuses on how uncertainty in the input factors propagates through the structure of the composite indicator and affects the composite indicator values" (Nardo et al., 2005). Therefore, a set of simulations is implemented to analyse alterations of the composite indicator result by changing the selection of individual indicators. As shown in Table, four input factors are defined and as a result, 189 samples are generated. While first two factors are more related to PCA components, last two factors are based on the selection of individual indicators. In order to further simplify the results of the test, the results of 33 boroughs in total are reflected by taking the average of the wards in each borough instead of presenting each ward.

	Explanation	**Samples**
Factor 1	Various subtractions in air quality indicator	1.1 : air quality dataset with PCA 1.2 : air quality dataset with EW method 1.3 : air quality dataset PCA weighted without PM25
Factor 2	Various subtractions in ecological footprint indicator	2.1: ecological footprint dataset with PCA 2.2: ecological footprint dataset with EW method 2.3: only carbon footprint
Factor 3	Various subtractions in environmental stresses domain	3.1: All environmental stress 3.2: POPD excluded 3.3: FERTR excluded 3.4: CAR_USE excluded 3.5: footprint excluded 3.6: HRECYR excluded 3.7: WASTE excluded
Factor 4	Various subtractions in environmental systems domain	4.1: all environmental systems 4.2: air quality excluded 4.3: ecological footprint excluded

Table 5.4: Description of 4 factor for Uncertainty Test

Looking at the results (Figure 5.5), it is clear that there is a substantial variance. The possible reason for this expected result is hidden in the last two factors based on subtracting the indicators from each domain. As described in the Table 5.4, the third factor is based on the subtraction of each of the environmental stress indicator, while the fourth factor is based on the subtraction of the environmental system indicators. To make sure of this possible argument, the first two factors that depend on the PCA analysis have been decomposed and presented in the Figure 5.7. As can be seen from this graph, the first two factors do not affect the variance in the results as last two factors. When the third factor is added to this decomposed test, the variance increases relatively, as indicated in the Figure 5.6. The conclusion that can be drawn from this is that each indicator simulated by subtracting them in the third and fourth factors has a great effect on the results. In addition, the reason for the accumulation in the middle region is probably the logarithmic transformation applied at the



initial stages.

Figure 5.5: Uncertainty Test Results including 189 samples with 4 factors applied

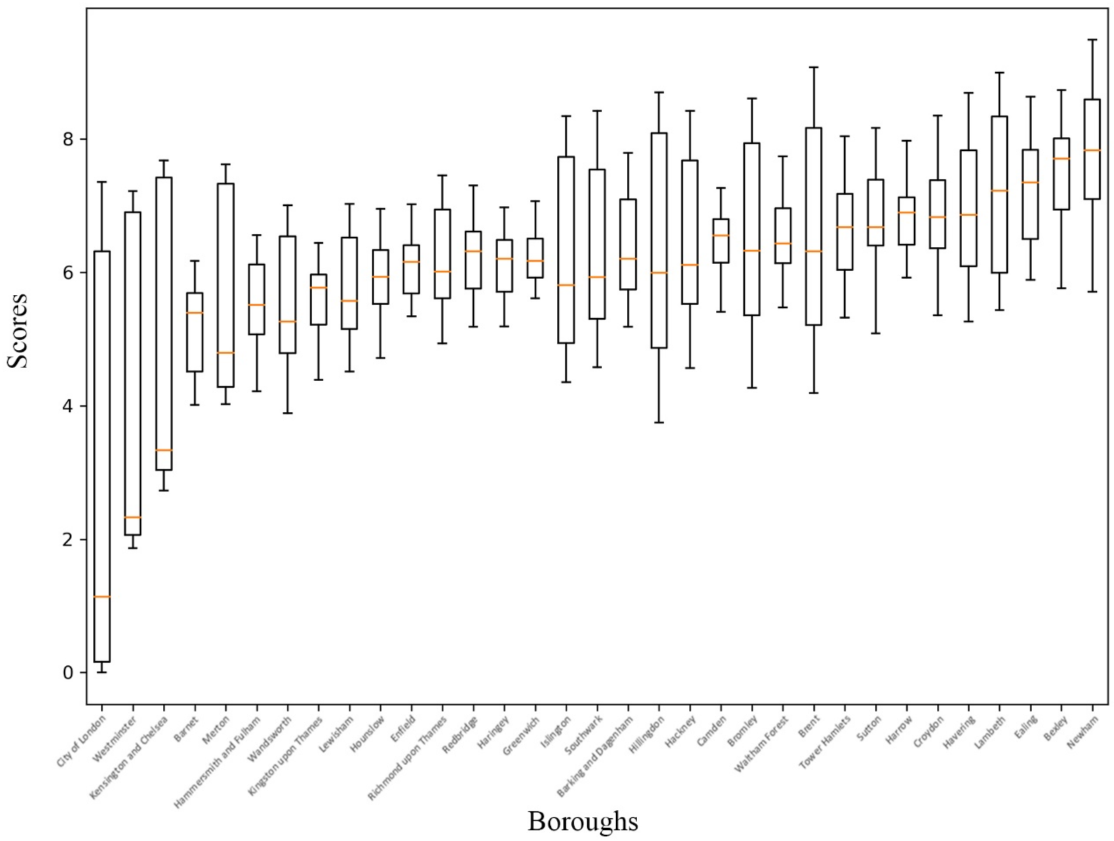


Figure 5.6: Uncertainty Test Results including 63 samples with first three factor applied

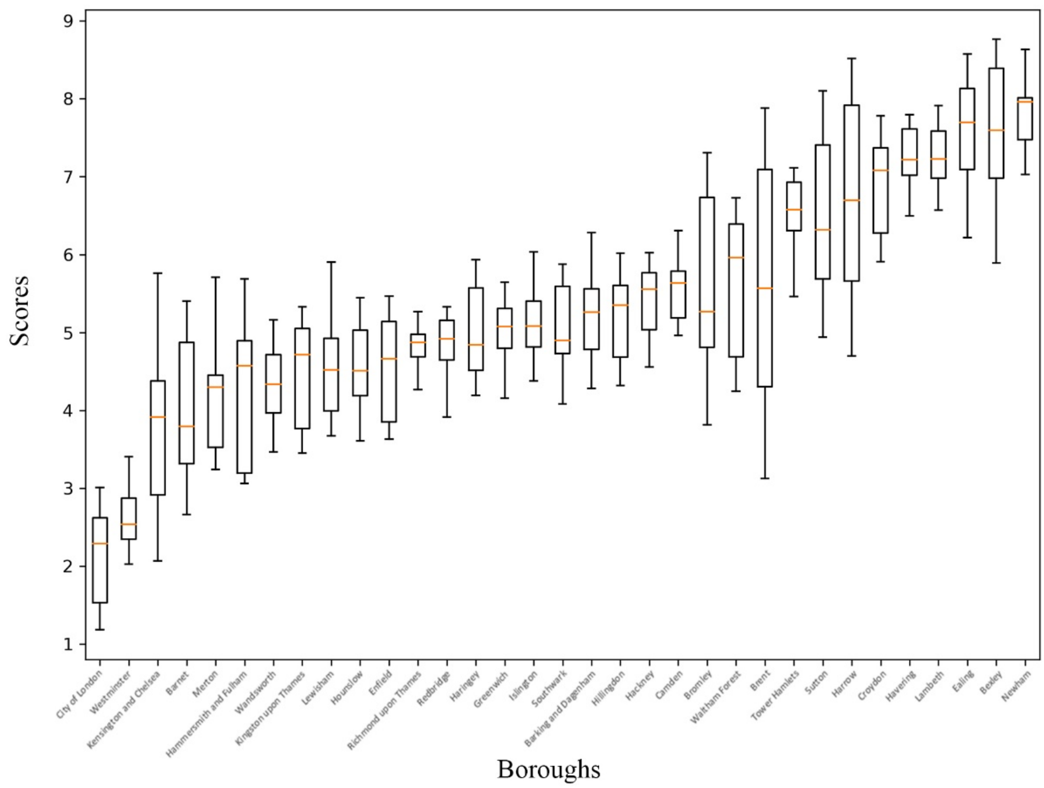


Figure 5.7: Uncertainty Test Results including 21 samples with first 2 factors applied

Chapter 6 – Results

6.1 Overall Results

The environmental sustainability index, which is the result of the arithmetic combination of the two domains in the framework, is scaled between 0 and 10 with the help of Min-Max Scaler for ease of representation. In this context, the results will be represented in 3 different stages. Firstly, the index created will be displayed on the map at the ward scale, then the arithmetic average of the wards in each borough will be taken to display at the borough scale. Secondly, the index will be decomposed, and individual indicators will be presented both at the ward and borough scale since decomposition may give better insight for overall performance of regions. At the last stage, cluster analysis will be applied to classify wards.

From the Figure 6.1, the first notable area is the inner part of the Greater London. These regions with the lowest environmental sustainability scores mostly belong to the Westminster and City of London Boroughs. In fact, this was an expected result. Relatively high individual indicators' values such as carbon footprint and air quality datasets have been already indicated that these regions would have lower results compared to other parts. Also, when we look at the maps of the 2 domains, we can see that they indicate low values in the inner part of the city. On the other hand, the index score is homogeneously distributed in the Greater London except some noises. South western and northern parts of the city have higher values compared their neighbours. Probably, the effect of the environmental stress domain produced lower sustainability index at these regions. It should also be noted that it is obvious that borough-scale datasets such as environmental footprint and waste production manipulate the results since it indicates the boundaries of the boroughs. In the environmental systems domain figure, the difference between the inner and outer regions appears clearer than the composite index result. The main reason for this is probably that the air quality value and the green area percentage are low in the central regions. On the contrary, a chaotic result has emerged in the environmental stress domain figure. Interestingly, in the central areas of the city, there are both high and low score areas. Low fertility rates and high recycling rates in some of these regions may have contributed to this situation. When we look at the environmental sustainability score on the borough scale (Figure 6.3), which is produced by simply taking arithmetic mean of wards in each borough, an image is close to the results mentioned above. However, instead of homogenous distribution with some noises, the results in here sharply reveal the differences in the borough.

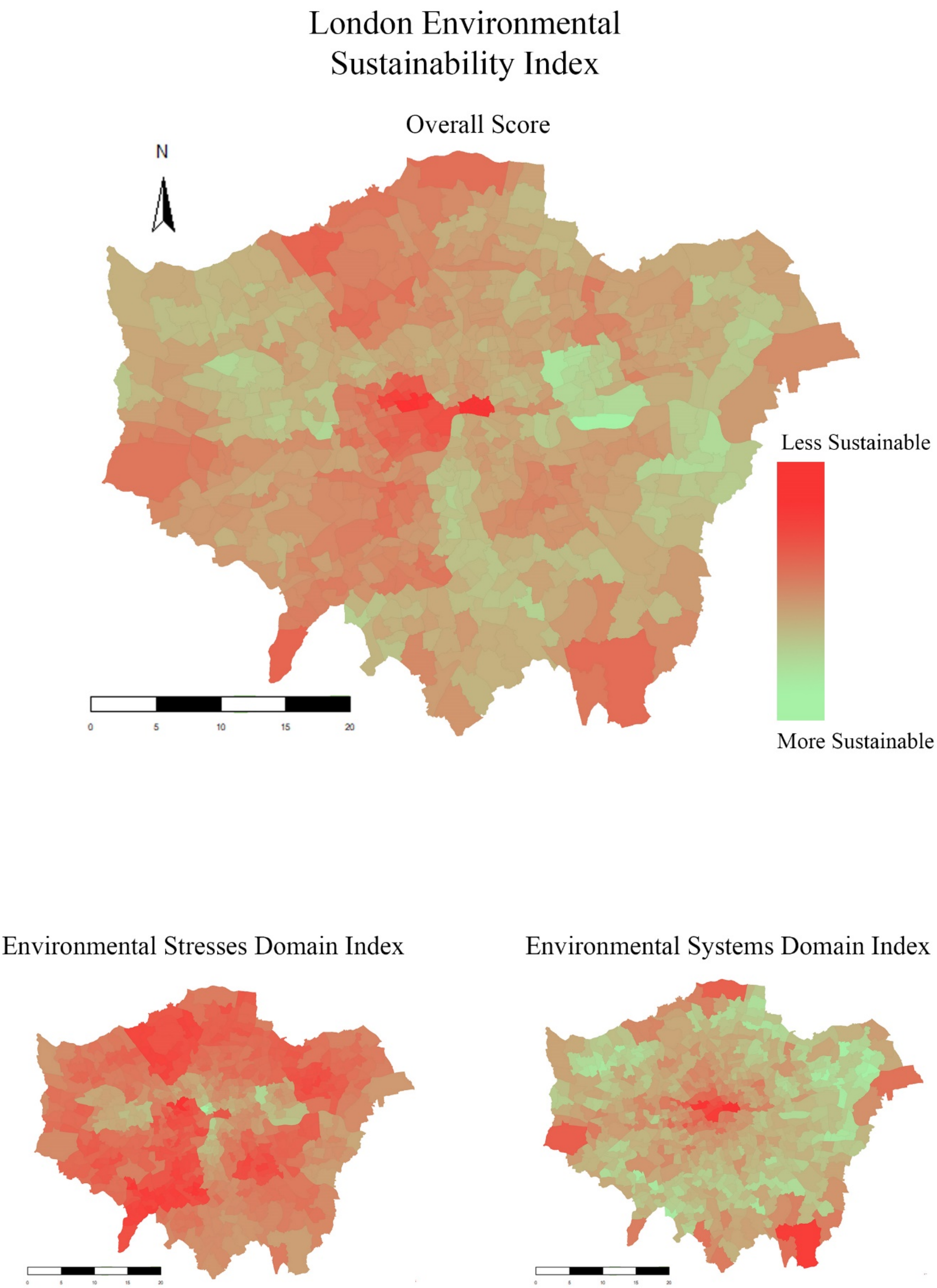


Figure 6.1: Overall Composite Index and Domains at ward scale

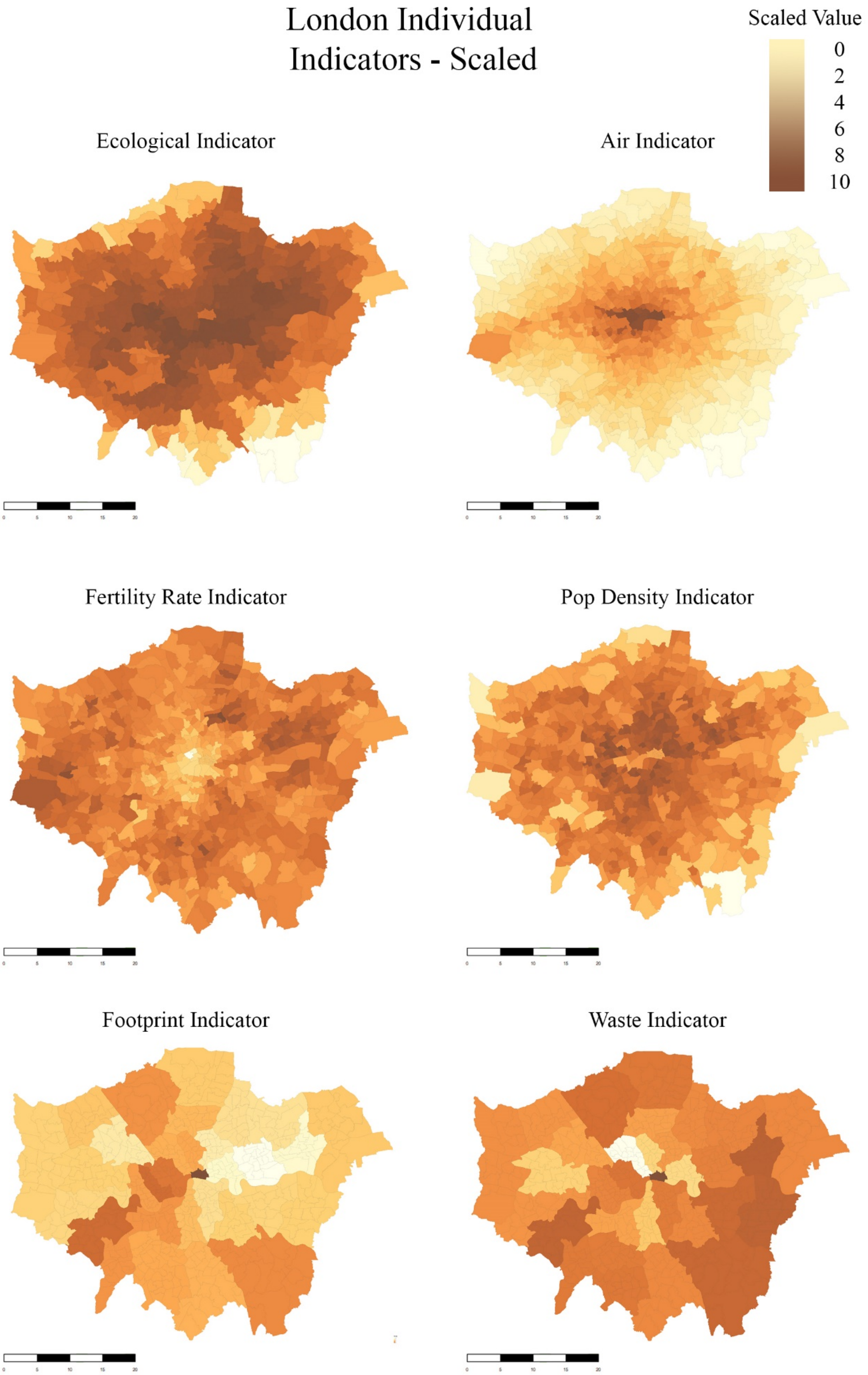


Figure 6.2: Individual Indicators Score Scaled between 0 and 10

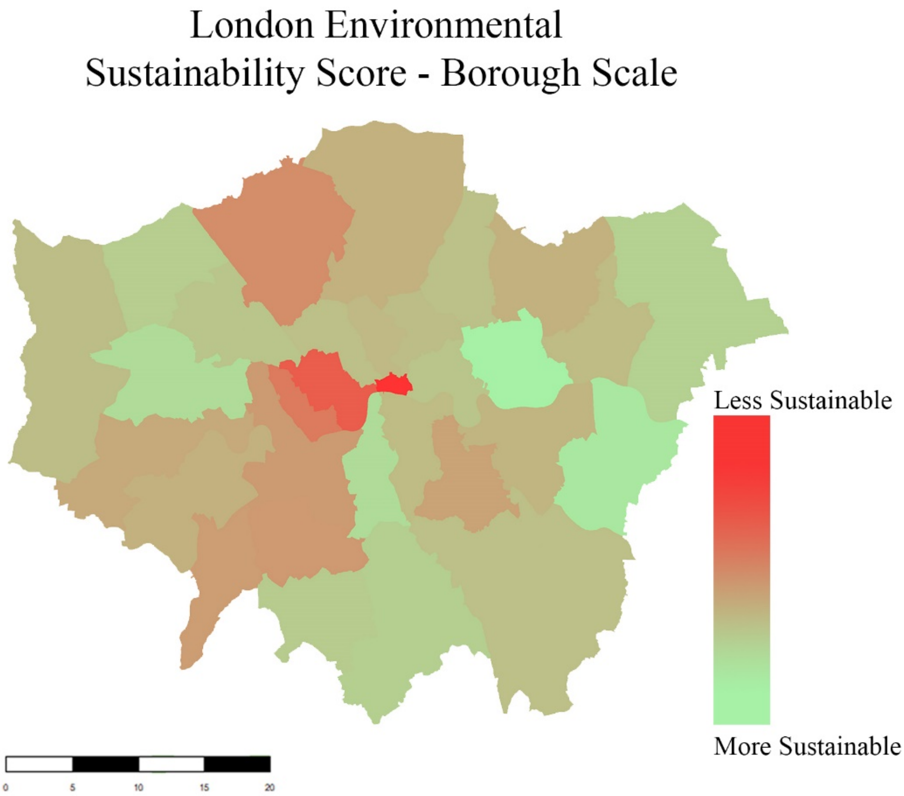


Figure 6.3: Overall Composite Index at borough scale

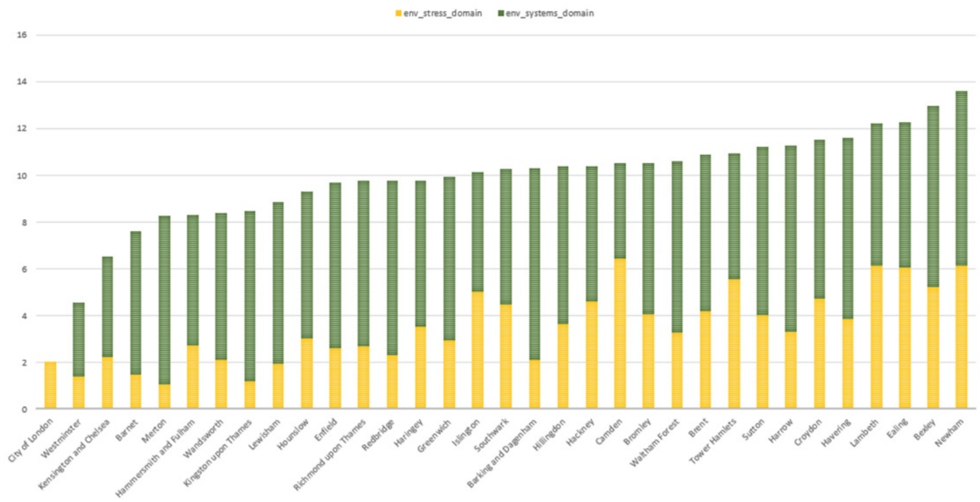


Figure 6.4: Domains Score for all boroughs

6.2 Cluster Analysis

Cluster analysis is applied as one of the result analysis methods. It is considered that it would be appropriate for decision makers to categorize the regions of the city according to the composite score and the indicators used in this structure. It may be effective for sustainable development to create plans on a group basis instead of applying a separate strategy to each region and evaluating each one separately. The cluster model is established with the help of K-means algorithm. As seen in the figure 6.5, there are three notable peaks: first one is when k=2 and second one is when k=7 and last one is when k=15. Although the silhouette scores are low, grouping wards according to their characteristics indicates that it is still essential when looking at the SSE results (Figure 6.6). As a results of cluster analysis, 15 different clusters are created.

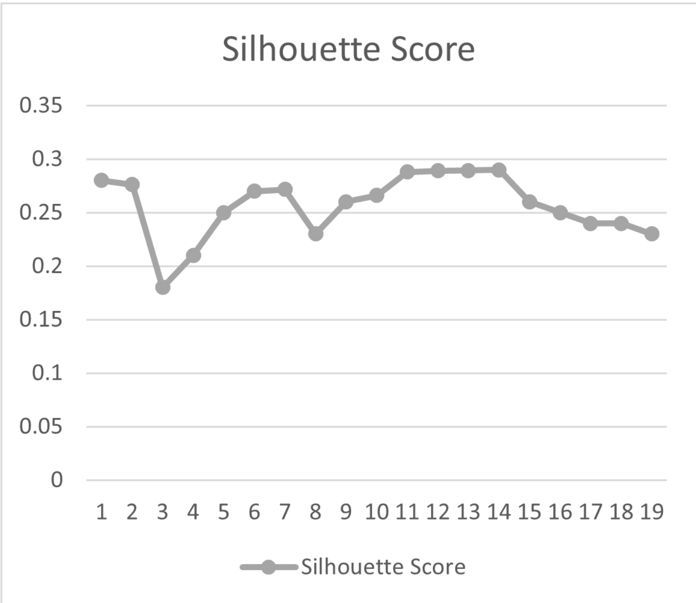


Figure 6.5: Cluster Analysis Silhouette Score for k=[1,20]

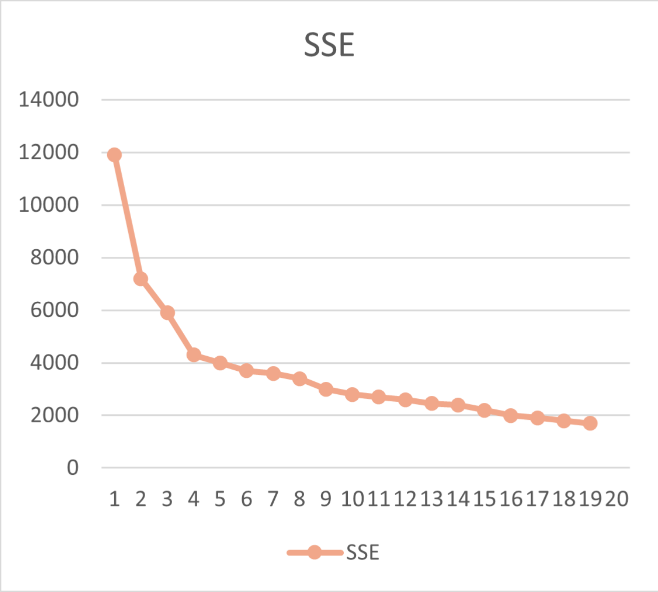


Figure 6.6: Cluster Analysis SSE results for k = [1,20]

As a result of cluster analysis, the lowest scoring areas are divided into 2 groups: cluster 1 and 4. Indeed, cluster 4 only consists of City of London since it's wards have the same value in most datasets which leaves a question mark in terms of data quality. These two groups are characteristics in terms of higher value in footprint indicator and lower value in air quality indicator. While cluster 6 and 8 consists of wards that have higher score in the environmental systems domain, cluster 9 and 10 includes wards that have lower score in same domain. Cluster 2 and 3 are more characteristic in terms of relatively lower air pollution value and lower footprint value, therefore they have higher environmental sustainability value. Cluster 5 and 7 have average values for all individual indicators. However, in some groups the analysis appears to act independently of the overall sustainability score. For example, cluster 10, 11 and 15 includes different characteristics of wards. While cluster analysis works effectively In cluster 13 which has lowest footprint value and cluster 14 which has lowest waste production value, it is difficult to distinguish characteristics easily in other clusters. The reason for this maybe the low silhouette score and the fact that some datasets provide values at different scales. However, when we look at the overall result of cluster analysis, categorization seems to work partially but effectively.

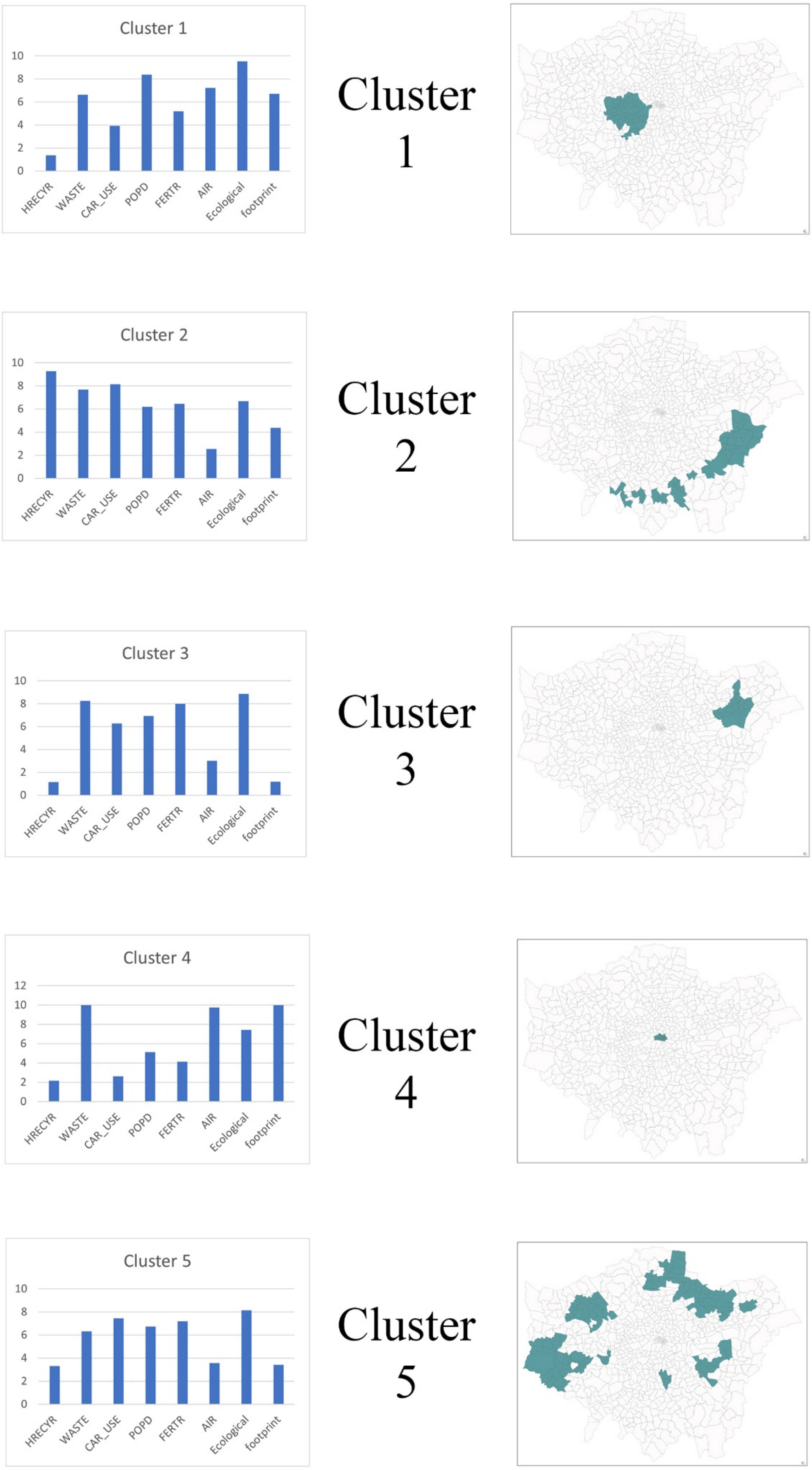


Figure 6.7: Cluster Analysis Results – 1 to 5

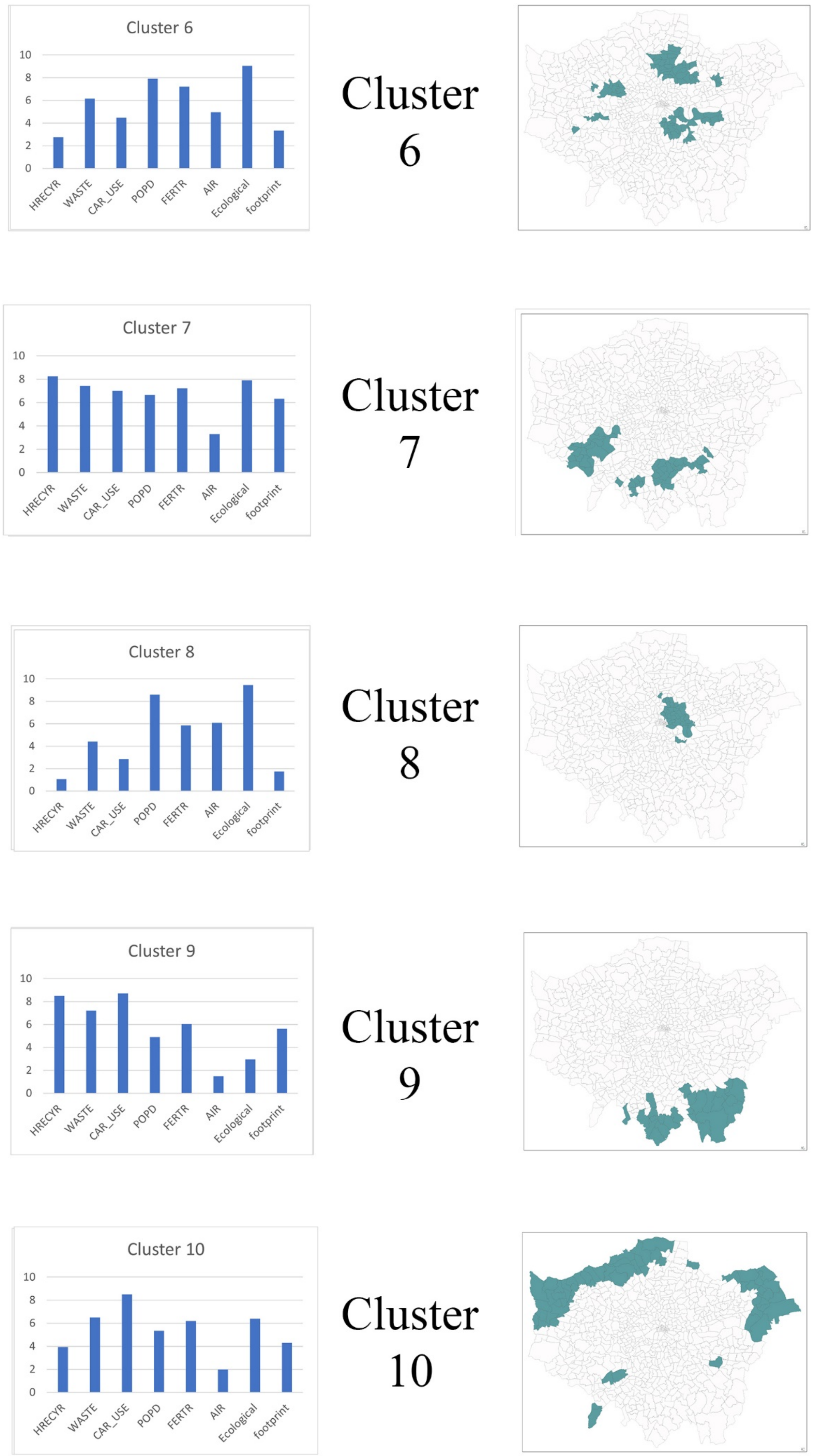


Figure 6.8: Cluster Analysis Results – 6 to 10

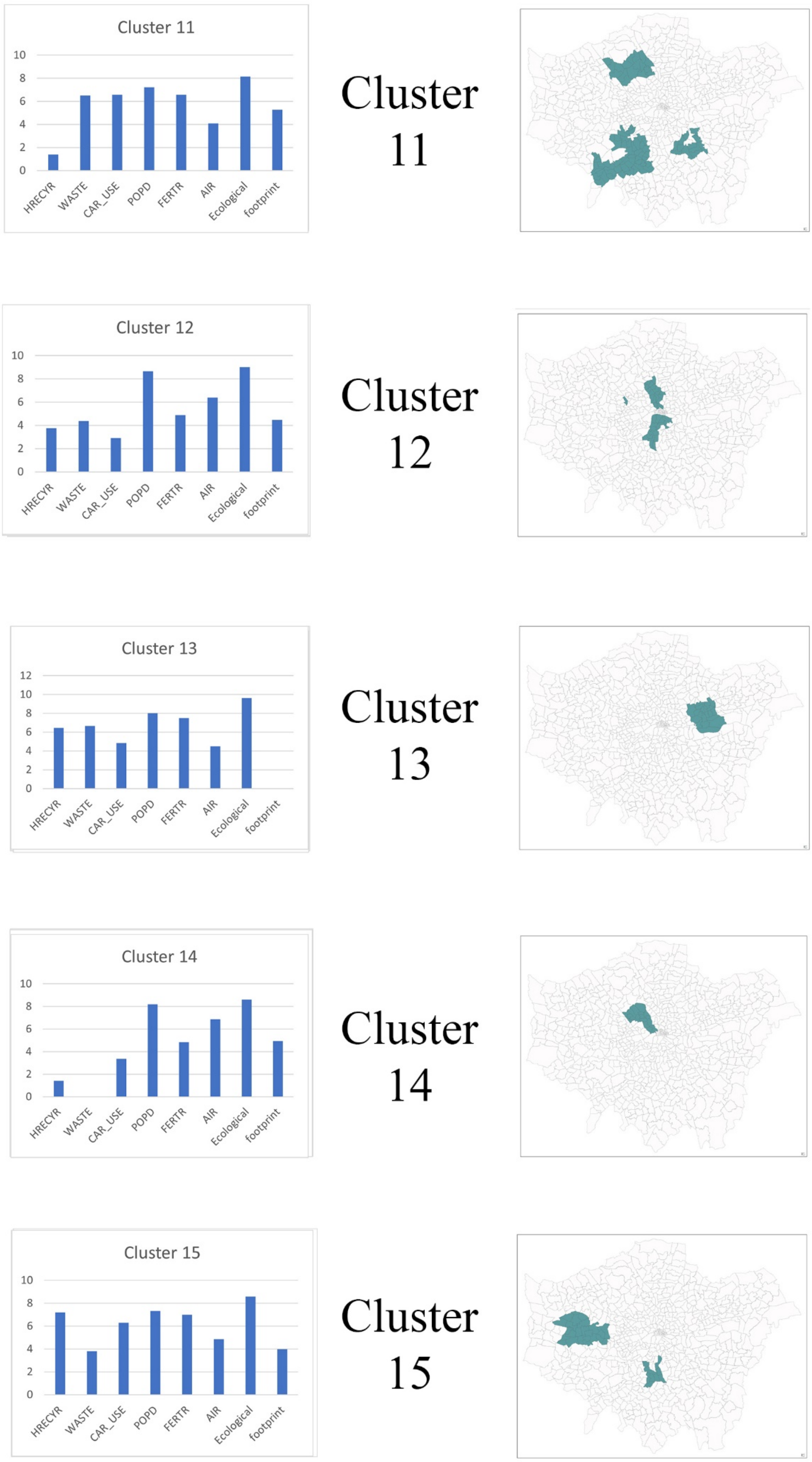


Figure 6.9: Cluster Analysis Results – 11 to 15

Chapter 7 - Discussion

Three research questions were mentioned at the beginning of the research. In this section, these three questions will be discussed considering the limitations, challenges and the possibilities for improvement.

- What factors promote environmental sustainability and how can environmental indicators affect sustainability?
- In order to measure environmental sustainability performance of cities, is it possible to get benefit from open data, and is it possible to create a composite framework which shows environmental performance effectively?
- Can such a framework assist decision-makers to provide better policy?

Before defining the factors, environmental sustainability was defined in the historical context, and thus, the factors emerged spontaneously. Then the factors affecting environmental sustainability were defined in the literature review sections with the help of international measurement methods such as ESI and region-oriented policy documents. Considering the lack of data on the ward scale, 14 different datasets provided from publicly available datastore were found. By the policy document's guidelines, these datasets are categorised under two main domains: environmental stresses and environmental systems. As a result, several factors affecting environmental sustainability were identified, and available datasets related to these factors were used to construct composite indicator. As stated in the previous sections, some factors may not work effectively in certain regions, in other words, factor selection for assessment depends on the characteristics of the region. Although reference is made to policy of selected region in this study, it is open to debate whether the selected indicators will work effectively at different scales. These challenges are mostly related to dataset selection. In addition to these challenges, proxy measures were used in this research due to the lack of available data. The extent to which these proxy variables reflect environmental sustainability is another challenge.

The framework created is a proof that the environmental sustainability of the city can be measured and analysed with publicly available data. Although the single index generated gives comprehensive information about the assessment of the regions in the city, the accuracy of the methods used is also open to discussion. For example, the index does not make sense on its own, on the contrary, it is efficient for comparison with other regions. In other words, instead of considering reference points for each indicator, competitiveness among wards and boroughs is under consideration. The correlation between selected datasets is another challenge in the research. This issue is fixed by applying PCA method. At the beginning of the research, it was thought to combine all indicators with the PCA method instead of giving equal weight. However, according to some researchers, using the PCA method in composite indicators may cause some data to be ignored because it can over-reduce the multidimensionality of the data (Grupp & Schubert, 2010). Therefore, PCA method was used only for the data with high correlation between them, and the created indicators were arithmetically combined by giving equal weight. Normalisation process is also another challenge as data

Composite indicators related to assessment of sustainability have 2 main purposes: to increase public awareness and to guide decision makers (Gómez-Limón et al., 2020). It is possible that the index generated in this research can guide decision makers to understand the current situation of environmental sustainability in a holistic way. In this way, they can review existing strategies and set new objectives and strategies according to composite index result. In addition, thanks to the cluster analysis, they can apply different strategies to regions in similar situations. As stated in Chapter 4, the information provided by the composite indicator allows to establish a dialectical structure in the policy triangle. In addition, it was intended at the beginning of the research to create more than one index with data from different time periods. Thus, the effects of strategies implemented in different time periods could be observed and the performance of regions in terms of improving environmental sustainability could be evaluated. However, this is not possible for now due to the lack of available data at different time. This issue can be considered as further development of the research.

Chapter 8 - Conclusion

In this study, firstly, the factors affecting environmental sustainability were found with the help of literature review. Thus, first question in the research was attempted to answered theoretically. Subsequently, the environmental sustainability performance of the Greater London area was measure using the composite indicator methodology, with the help of London's relevant policy documents on environmental sustainability. Also, inspired by existing methods such as EPI and ESI, the processes of the composite indices were appropriately applied. Although the Greater London region was chosen due to the data richness, the relatively small number of indicators used was one of the main limitations affecting the result. One of the aims of the research was that the data used should be publicly available data, so the selected data were obtained from the London datastore, but as mentioned above some proxy measures may not accurately reflect the results. Therefore, this issue could be another challenge for this research.

The values obtained from the composite framework indicate the competitive results between each ward and borough in the Greater London. Although the environmental sustainability reference point is not considered in this research, the competitive results among lower scale regions can be useful and informative for decision makers. When we look at the results, the 3 boroughs with the lowest values are City of London, Westminster, and Kensington and Chelsea. These results are not surprising because it can be easily obtained when we look at the indicators independently of the composite framework. But it should be noted again that these 3 boroughs are the results when compared with other boroughs' performance. On the other hand, the 3 most environmentally sustainable boroughs are Newham, Bexley and Ealing. Apart from these, there were relatively many boroughs with medium values, that is, with a sustainability score close to 5. The reason for this can be explained that the logarithmic transformation applied in the initial step influenced the results. However, this transformation was necessary due to some outliers and skewed distributions in most of the datasets.

While constructing the composite framework, the OECD's informative handbook successfully guided, however, actions at each step are still open to debate in terms of accuracy. For example, in the uncertainty test based on the subtraction of indicators, it has been observer that the value changes are relatively high. The reason for this situation can be seen as the limitation of the available data. In addition, double counting issue was attempted to be solved by applying PCA among datasets with relatively high correlation. Equal weighting method was applied to other datasets and thus domains referenced from ESI were valued. As a result, an index was created by assigning negative value to datasets that have negative impact on sustainability and vice versa. The composite index in this study is aimed to allow decision makers to approach environmental sustainability in an urban context in a holistic way. With the cluster analysis applied on the results, the regions with a similar situation were classified and it became a comprehensive guide for the strategies to be made in the future. However, due to the limitations stated in the previous section, it is not recommended to use the results alone, but rather as a supplementary element for another source. To further improvement, measuring the performance of urban regions in different time periods by analysing different datasets could be appropriate.

Bibliography

Adinyira, E., Oteng-seifah, S., & Adjei-kumi, T. (2007). A review of urban sustainability assessment methodologies. *Whole Life Urban Sustainability and Its Assessment*, 1–8.

Alberti, M. (1996). Measuring urban sustainabilityAlberti, M. (1996). Measuring urban sustainability. Environmental Impact Assessment Review, 16(4-6), 381-424. New York: Plenum Press, 1980-. doi:10.1016/S0195-9255(96)00083-2. *Environmental Impact Assessment Review*, *16*(4–6), 381–424. http://linkinghub.elsevier.com/retrieve/pii/S0195925596000832

Ameen, R. F. M., Mourshed, M., & Li, H. (2015). A critical review of environmental assessment tools for sustainable urban design. *Environmental Impact Assessment Review*, *55*, 110–125. https://doi.org/10.1016/j.eiar.2015.07.006

Basiago, A. D. (1996). The search for the sustainable city in 20th century urban planning. *Environmentalist*, *16*(2), 135–155. https://doi.org/10.1007/BF01325104

Bond, A., Morrison-Saunders, A., & Pope, J. (2012). Sustainability assessment: The state of the art. *Impact Assessment and Project Appraisal*, *30*(1), 53–62. https://doi.org/10.1080/14615517.2012.661974

Button, K. (2002). City management and urban environmental indicators. *Ecological Economics*, *40*(2), 217–233. https://doi.org/10.1016/S0921-8009(01)00255-5

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., & Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *Journal of the Operational Research Society*, *59*(2), 239–251. https://doi.org/10.1057/palgrave.jors.2602445

Cohen, M. (2017). A systematic review of urban sustainability assessment literature. *Sustainability (Switzerland)*, *9*(11). https://doi.org/10.3390/su9112048

Corredor-Ochoa, Á., Antuña-Rozado, C., Fariña-Tojo, J., & Rajaniemi, J. (2020). Challenges in assessing urban sustainability. *Urban Ecology*, 355–374. https://doi.org/10.1016/b978-0-12-820730-7.00019-7

Daly, H. E. (1990). Toward some operational principles of sustainable development. *Ecological Economics*, *2*(1), 1–6. https://doi.org/10.1016/0921-8009(90)90010-R

El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, *89*(1), 1–24. https://doi.org/10.1007/s11573-018-0902-z

Fiksel, J., Tarsha, E., Frederickson, & Herbert. (2012). A Framework for Sustainability Indicators at EPA A Framework for Sustainability Indicators at EPA Authors. *Livestock Research for Rural Development*.

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

Gómez-Limón, J. A., Arriaza, M., & Guerrero-Baena, M. D. (2020). Building a composite indicator to measure environmental sustainability using alternative weighting methods. *Sustainability (Switzerland)*, *12*(11). https://doi.org/10.3390/su12114398

GLA (2013). London's Urban Heat Island 2013 - London Datastore. URL: https://data.london.gov.uk/dataset/london-s-urban-heat-island (visited on July 25, 2021).

GLA (2014). Ward Profiles and Atlas 2014 - London Datastore. URL: https://data.london.gov.uk/dataset/ward-profiles-and-atlas (visited on July 25, 2021).

GLA (2016). London Atmospheric Emissions (LAEI) 2016 - London Datastore. URL: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2016 (visited on July 25, 2021).

GLA (2018). Waste Plans 2018 - London Datastore. URL: https://data.london.gov.uk/dataset/waste-plans (visited on July 25, 2021).

GLA (2019). Green and Blue Cover 2019 - London Datastore. URL: https://data.london.gov.uk/dataset/green-and-blue-cover (visited on July 25, 2021).

Greater London Authority. (2018). *London Environment Strategy*. www.london.gov.uk

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research*, *141*(1), 61–94. https://doi.org/10.1007/s11205-017-1832-9

Grupp, H., & Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, *39*(1), 67–78. https://doi.org/10.1016/j.respol.2009.10.002

Jacobs, R., Smith, P., & Goddard, M. (2004). Measuring performance: An examination of composite performance indicators. *Technical Paper Series*, *29*, 112.

Kaur, H., & Garg, P. (2019). Urban sustainability assessment tools: A review. *Journal of Cleaner Production*, *210*, 146–158. https://doi.org/10.1016/j.jclepro.2018.11.009

Marazza, D., Bandini, V., & Contin, A. (2011). A composite indicator on environmental sustainability performance at urban scale: the K8 approach. *Esee2011.Org*, *July 2015*, 13–17. http://www.esee2011.org/registration/fullpapers/esee2011_a11f9e_1_1304865006_2441_2319.pdf

Metcalf, L., & Casey, W. (2016). Introduction to data analysis. *Cybersecurity and Applied Mathematics*, 43–65. https://doi.org/10.1016/b978-0-12-804452-0.00004-x

More, D. C., & More, C. (2002). Understanding the Industrial Revolution. *Understanding the Industrial Revolution*. https://doi.org/10.4324/9780203136973

Morelli, J. (2011). Environmental Sustainability: A Definition for Environmental Professionals. *Journal of Environmental Sustainability*, *1*(1), 1–10. https://doi.org/10.14448/jes.01.0002

Mori, K., & Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review*, *32*(1), 94–106. https://doi.org/10.1016/j.eiar.2011.06.001

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators. In *OECD Statistics Working Papers* (Issue 03). http://www.oecd-ilibrary.org/docserver/download/5lgmz9dkcdg4.pdf?expires=1471336777&id=id&accna

me=guest&checksum=158391DADFA324416BB9015F3E4109AF

National Academies of Sciences, Engineering, and M. 2016. (2016). Pathways to urban sustainability: Challenges and opportunities for the United States. *Pathways to Urban Sustainability: Challenges and Opportunities for the United States*, 1–176. https://doi.org/10.17226/23551

National, B., & Stewardship, E. (2005). *2005 Environmental Sustainability Index*.

Paull, J. (2013). The Rachel Carson letters and the making of silent Spring. *SAGE Open*, *3*(3), 1–13. https://doi.org/10.1177/2158244013494861

Purvis, B., Mao, Y., & Robinson, D. (2019). Three pillars of sustainability: in search of conceptual origins. *Sustainability Science*, *14*(3), 681–695. https://doi.org/10.1007/s11625-018-0627-5

Rees, W., & Wackernagel, M. (2008). Urban ecological footprints: Why cities cannot be sustainable-and why they are a key to sustainability. In *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*. https://doi.org/10.1007/978-0-387-73412-5_35

Saisana, M., & Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for com- posite indicator development. European Commission, Joint Research Centre, Institute for the Protec- tion and the Security of the Citizen, Technological and Economic Risk Management Unit, Ispra, Italy.

Scikit (2018) *Sklearn Preprocessing Standard Scaler* [Online] Available at:https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html [Accessed: 28 July 2021].

Sharifi, A. (2021). Urban sustainability assessment: An overview and bibliometric analysis. *Ecological Indicators*, *121*, 107102. https://doi.org/10.1016/j.ecolind.2020.107102

Shields, D. J., Šolar, S. V., & Martin, W. E. (2002). The role of values and objectives in communicating indicators of sustainability. *Ecological Indicators*, *2*(1–2), 149–160. https://doi.org/10.1016/S1470-160X(02)00042-0

Shields, D.J., Mitchell, J.E., 1997. A Hierarchical Systems Model of Ecosystem Management. Working Paper, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, *9*(2), 189–212. https://doi.org/10.1016/j.ecolind.2008.05.011

United Nations, 1987. World commission of environment and development report. Our Common Future,.

United Nations, 2018. World Urbanization Prospects - the 2018 Revision.

Vasseur, E. (1973). United Nations Conference on the Human Environment. *Water Research*, *7*(8), 1227–1233. https://doi.org/10.1016/0043-1354(73)90077-8

Vojnovic, I. (2014). Urban sustainability: Research, politics, policy and practice. *Cities*, *41*, S30–S44. https://doi.org/10.1016/j.cities.2014.06.002

Wu, J. (2014). Urban ecology and sustainability: The state-of-the-science and future directions. *Landscape and Urban Planning*, *125*, 209–221. https://doi.org/10.1016/j.landurbplan.2014.01.018

Yale Center for Environmental Law & Policy. (2006). *Pilot 2006 Environmental Performance Index Main Report*. http://www.yale.edu/epi/

Yigitcanlar, T., Dur, F., & Dizdaroglu, D. (2015). Towards prosperous sustainable cities: A multiscalar urban sustainability assessment approach. *Habitat International*, *45*(P1), 36–46. https://doi.org/10.1016/j.habitatint.2014.06.033

Appendix

Scripts

```
#Import Necessary Library
import geopandas as gpd
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler

#DATA
# Temperature data processed in QGIS application and converted to GeoPackage file
# Data
temperature_gpkg = gpd.read_file("./Data/temperature.gpkg")
temperature_raw = pd.DataFrame()
temperature_raw[["GSS_CODE", "TEMP"]] =  temperature_gpkg[["GSS_CODE", 
"WSAVG"]]
temperature_raw.describe()
# Air quality datas processed in QGIS application(points to wards)
pm10_raw = pd.read_csv("./Data/annual_mean_of_particulate_matter_pm10_2016.csv")
pm25_raw = pd.read_csv("./Data/annual_mean_of_particulate_matter_pm25_2016.csv")
nox_raw  = pd.read_csv("./Data/annual_mean_of_nitrogen_oxide_nox_2016.csv")
no2_raw  = pd.read_csv("./Data/annual_mean_of_nitrogen_dioxide_no2_2016.csv")
# Vehicle use and land use data
land_use_raw = pd.read_csv("./Data/land_use_ward.csv")
vehicle_use_raw = pd.read_csv("./Data/licenced_vehicle.csv")
ward_atlas_raw = pd.read_csv("./Data/ward_atlas_data.csv")
london_all = ward_atlas_raw[["Codes", "Borough"]]
```

```
# Borough level data

environmental_footprint_borough_raw =
pd.read_csv("./Data/environmental_footprint_borough.csv")

household_recycle_rate_borough_raw =
pd.read_csv("./Data/household_recycle_rate_borough.csv")

waste_per_head_borough_raw =
pd.read_csv("./Data/annual_household_waste_per_head.csv")

# integrating the temp data into the same index as the others

temperature_raw =temperature_raw.set_index("GSS_CODE")

temperature_raw = temperature_raw.reindex(index=ward_atlas_raw["Codes"])

temperature_raw = temperature_raw.reset_index()

# Null check

temperature_raw.isnull()

# Manuel entry for null variable (City of London = 18.0872)

temperature_raw["TEMP"][0] = 18.0872

# Changing column name

environmental_footprint_borough_raw =
environmental_footprint_borough_raw.rename(columns={"Area":"Borough"})

environmental_footprint_ward = pd.DataFrame()

environmental_footprint_ward = london_all.merge(environmental_footprint_borough_raw,
on="Borough")


household_recycle_rate_borough_raw =
household_recycle_rate_borough_raw.rename(columns={"Area":"Borough"})

household_recycle_rate_ward = pd.DataFrame()

household_recycle_rate_ward = london_all.merge(household_recycle_rate_borough_raw,
on="Borough")


waste_per_head_borough_raw =
waste_per_head_borough_raw.rename(columns={"Area":"Borough"})

waste_per_head_ward = pd.DataFrame()

waste_per_head_ward = london_all.merge(waste_per_head_borough_raw, on="Borough")

# Null check

waste_per_head_ward.isnull()

household_recycle_rate_ward.isnull()
```

```
environmental_footprint_ward.isnull()
#Datasets from different sources are collected to one dataframe

ecological_footprint = environmental_footprint_ward["Ecological_Footprint_(gha/capita)"]
carbon_footprint = environmental_footprint_ward["Carbon_Footprint_(tonnes CO2/capita)"]
temperature = temperature_raw["TEMP"]
household_recycle_rate = household_recycle_rate_ward["household_recycle_rate"]
fertility_rate = ward_atlas_raw["fertility_rate"]
population_density = ward_atlas_raw["population_density_per_sqkm"]
greenspace_area_percentage = land_use_raw["%Area_of_Greenspace"]
vehicle_use = vehicle_use_raw["licenced_vehicle_per_household"]
waste_per_head = waste_per_head_ward["annual_household_waste_per_head_kgs/head"]
water_area_percentage = land_use_raw["%Area_of_Water"]
no2 = no2_raw["NO2"]
nox = nox_raw ["Nox"]
pm25 = pm25_raw["pm25"]
pm10 = pm10_raw["pm10"]
GSS_CODE = ward_atlas_raw["Codes"]
london_all_data = {"GSS_CODE": GSS_CODE,
        "NO2" : no2,
        "NOX" : nox,
        "PM25" : pm25,
        "PM10" : pm10,
        "CAR_USE" : vehicle_use,
        "GREEN%" : greenspace_area_percentage,
         "WATER%": water_area_percentage,
        "POPD" : population_density,
        "FERTR" : fertility_rate,
        "HRECYR" : household_recycle_rate,
        "TEMP" : temperature,
        "CARBONF" : carbon_footprint,
        "ECOF" : ecological_footprint,
```

```
                "WASTE":waste_per_head}
london_dataframe_raw = pd.DataFrame(london_all_data)
#Correlation Plot
corr_data = london_dataframe_raw.drop(["GSS_CODE"], axis=1)
corr_matrix = corr_data.corr()
#Using heatmap to visualize the correlation matrix
fig, ax = plt.subplots(figsize=(12, 12))
#mask = np.zeros_like(corr_data.corr())
mask = np.zeros_like(corr_matrix)
mask[np.triu_indices_from(mask)] = 1
sns.heatmap(corr_matrix, mask= mask, ax= ax, annot= True)
#Log transformation
london_dataframe_raw_v1 = london_dataframe_raw.drop(["GSS_CODE"],axis=1)
london_dataframe_raw_v1["CAR_USE"] = london_dataframe_raw_v1["CAR_USE"]*100
london_dataframe_raw_v1["PM25"] = london_dataframe_raw_v1["PM25"]*100
london_dataframe_raw_v1["ECOF"] = london_dataframe_raw_v1["ECOF"]*100
london_dataframe_raw_log = np.log(london_dataframe_raw_v1)
london_dataframe_raw_log["GSS_CODE"] = london_dataframe_raw["GSS_CODE"]
london_dataframe_raw = london_dataframe_raw_log

london_dataframe_raw.describe()
london_dataframe_raw = -london_dataframe_raw
london_dataframe_raw["GREEN%"] = - london_dataframe_raw["GREEN%"]
london_dataframe_raw["HRECYR"] = - london_dataframe_raw["HRECYR"]
#PCAs
ss = StandardScaler()
pca = PCA()
# PCA for air quality indicator
air_raw = london_dataframe_raw.filter(["NOX", "NO2", "PM10", "PM25"])
#sns.pairplot(air)
air_scaled = ss.fit_transform(air_raw)
pca_air = pca.fit_transform(air_scaled)
```

```
# checking pca components explianed variance and its ratio
print('Explained variation per principal component:
{}'.format(pca.explained_variance_ratio_))
print(pca.components_)
print(pca.explained_variance_)
# clearly first component has more value for explained variation ratio
# creating dataframe including each component
pca_air_Df = pd.DataFrame(data = pca_air
            , columns = ['pc1','pc2', 'pc3','pc4'])
# MinMax Scale for air
scale = MinMaxScaler(feature_range=(0, 1))
pca_air_scaled = scale.fit_transform(pca_air_Df)
pca_air_scaled_Df = pd.DataFrame(data = pca_air_scaled
            , columns = ['pc1','pc2', "pc3", "pc4"])
pca_air_scaled_Df

# PCA for environmental footprint
footprint = london_dataframe_raw.filter(["CARBONF", "ECOF"])
#sns.pairplot(footprint)
footprint = ss.fit_transform(footprint)
pca_footprint = pca.fit_transform(footprint)
print('Explained variation per principal component:
{}'.format(pca.explained_variance_ratio_))
print(pca.explained_variance_)

# MinMax Scale for footprint
scale = MinMaxScaler(feature_range=(0, 1))
pca_footprint_scaled = scale.fit_transform(pca_footprint)
pca_footprint_scaled_Df = pd.DataFrame(data = pca_footprint_scaled
            , columns = ['pc1','pc2'])
pca_footprint_scaled_Df
remaining_env_systems_id = ["GREEN%", "TEMP"]
remaining_env_systems = london_dataframe_raw.filter(remaining_env_systems_id)
```

```
remaining_env_systems = ss.fit_transform(remaining_env_systems)
remaining_env_systems_pca = pca.fit_transform(remaining_env_systems)
print('Explained variation per principal component:
{}'.format(pca.explained_variance_ratio_))
print(pca.explained_variance_)


remaining_env_systems_pca_list = []
for index_env in range(0,len(remaining_env_systems_pca)):
  remaining_env_systems_pca_list.append(remaining_env_systems_pca[index_env])
remaining_env_systems_pca = pd.DataFrame(list(remaining_env_systems_pca_list))


#MinMax Scale for environmental system domain
scale = MinMaxScaler(feature_range=(0, 1))
# Scale
remaining_env_systems_pca_scaled = scale.fit_transform(remaining_env_systems_pca_list)
remaining_env_systems_pca_scaled_df =
pd.DataFrame(list(remaining_env_systems_pca_scaled))
remaining_env_systems_pca_scaled_df.columns = ["pc1", "pc2"]
remaining_env_systems_pca_scaled_df


#Remaining Env Stress Indicators (Standardise and MinMax Scale)
remainin_env_stress_ind = ["HRECYR", "WASTE", "CAR_USE", "POPD", "FERTR"]
remainin_env_stress = london_dataframe_raw.filter(remainin_env_stress_ind)
remainin_env_stress = ss.fit_transform(remainin_env_stress)
#MinMax Scale
scale = MinMaxScaler(feature_range=(0, 1))
remainin_env_stress_scaled = scale.fit_transform(remainin_env_stress)
remainin_env_stress_df = pd.DataFrame(list(remainin_env_stress_scaled))
remainin_env_stress_df.columns = remainin_env_stress_ind
remainin_env_stress_df


env_stress_domain = (pca_footprint_scaled_Df["pc1"]
                + remainin_env_stress_df["HRECYR"]
```

```
                + remainin_env_stress_df["WASTE"]
                + remainin_env_stress_df["CAR_USE"]
                + remainin_env_stress_df["POPD"]
                + remainin_env_stress_df["FERTR"])/6


env_stress_domain = pd.DataFrame(list(env_stress_domain))
env_stress_domain.columns = ["env_stress_domain"]

env_systems_index = (pca_air_scaled_Df["pc1"] +
remaining_env_systems_pca_scaled_df["pc1+pc2"])

scale = MinMaxScaler(feature_range=(0, 10))
env_systems_index = scale.fit_transform(env_systems_index)
env_stress_domain = scale.fit_transform(env_stress_domain)
env_sust_index = (env_systems_index + env_stress_domain)/2
env_sust_index = pd.DataFrame(list(env_sust_index))
scale = MinMaxScaler(feature_range=(0, 10))
env_sust_index = scale.fit_transform(env_sust_index)
env_sust_index = pd.DataFrame(list(env_sust_index))
env_sust_index.columns = ["env_sust_index"]
env_sust_index
#Uncertainty Analysis
POPD = remainin_env_stress_df["POPD"]
FERTR = remainin_env_stress_df["FERTR"]
CAR_USE = remainin_env_stress_df["CAR_USE"]
HRECYR = remainin_env_stress_df["HRECYR"]
WASTE = remainin_env_stress_df["WASTE"]
footprint = pca_footprint_scaled_Df["pc1"]
air = pca_air_scaled_Df["pc1"]
air_scaled_df = pd.DataFrame(data = air_scaled
          , columns = ['NOX','NO2', 'PM10','PM25'])
```

```
ecological = remaining_env_systems_pca_scaled_df["pc1+pc2"]
CARBONF = remainin_env_stress_df["CARBONF"]
ECOF = remainin_env_stress_df["ECOF"]
NOX = air_scaled_df["NOX"]
NO2 = air_scaled_df["NO2"]
PM10 = air_scaled_df["PM10"]
PM25 = air_scaled_df["PM25"]
uncertainity_results = pd.DataFrame()

factor1 = ["air_pollution_pca", "air_pollution_equal", "air_pollution_PCA_without_PM25"]
factor2 = ["footprint_pca", "footprint_equal", "only_CARBONF"]
factor3 = {"all_env_stress" :[0.166, 0.166, 0.166, 0.166, 0.166, 0.166],
        "POPD_exluded":[0 ,0.2, 0.2, 0.2, 0.2, 0.2],
        "FERTR_exluded": [0.2 ,0, 0.2, 0.2, 0.2, 0.2],
        "CAR_USE_excluded": [0.2 ,0.2, 0, 0.2, 0.2, 0.2],
        "footprint_exluded": [0.2 ,0.2, 0.2, 0, 0.2, 0.2],
        "HRECYR_exluded": [0.2 ,0.2, 0.2, 0.2, 0, 0.2],
        "WASTE_exluded": [0.2 ,0.2, 0.2, 0.2, 0.2, 0]}
factor4 = {"all_env_systems" : [0,1], "air_exluded": [0,1], "ecological_exluded": [0.5,0.5]}



steps_taken=0

for a in factor1:
  if a== "air_pollution_equal":
    air = (NOX + NO2 + PM10 + PM25)/4
    air = pd.DataFrame(data=air)
    # MinMax Scale for air
    scale = MinMaxScaler(feature_range=(0, 1))
    air = scale.fit_transform(air)
    air = map(lambda x: x[0], air)
```

```
    air = pd.Series(air)
elif a =="air_pollution_PCA_without_PM10":
    air = pd.DataFrame({"NOX": NOX, "NO2": NO2, "PM10": PM10})
    air = pca.fit_transform(air)
    air = map(lambda x: x[0], air)
    air = pd.Series(air)
else:
    air = air
for d in factor2:
     if d== "only_CARBONF":
        ecological = CARBONF
        air = (NOX + NO2 + PM10 + PM25)/4
        air = pd.DataFrame(data=air)
        # MinMax Scale for air
        scale = MinMaxScaler(feature_range=(0, 1))
        air = scale.fit_transform(air)
        air = map(lambda x: x[0], air)
        air = pd.Series(air)
    elif d =="footprint_equal":
        ecological = (CARBONF+ECOF)/2
        # MinMax Scale for footprint
        scale = MinMaxScaler(feature_range=(0, 1))
        ecological = scale.fit_transform(ecological)
        ecological = map(lambda x: x[0], ecological)
        ecological = pd.Series(ecological)
    else:
        ecological = ecological
    for b in factor2:
        steps_taken += 1
        weights = factor2[b]
        env_stess_dom = (POPD*weights[0] + FERTR*weights[1] +CAR_USE*weights[2]
        +footprint*weights[3] +HRECYR*weights[4] +WASTE*weights[5])/6
```

```
    for c in factor3:
        weight = factor3[c]
        env_system_dom = (air*weight[0] + ecological*weight[1])
        env_sust_index = (env_system_dom + env_stess_dom)/2
        env_sust_index = pd.DataFrame(data=env_sust_index)
        scale = MinMaxScaler(feature_range=(0, 10))
        env_sust_index = scale.fit_transform(env_sust_index)
        env_sust_index = map(lambda x: x[0], env_sust_index)
        env_sust_index = pd.Series(env_sust_index)
        env_sust_index
        uncertainity_results[steps_taken] = env_sust_index
        steps_taken +=1
uncertainity_results = london_all.merge(uncertainity_results, left_index=True,
right_index=True)
uncertainity_results = uncertainity_results.drop(["Codes"],  axis=1)
uncertainity_results_borough = uncertainity_results.groupby(["Borough"]).mean()
```