UNIVERSIDADE CATÓLICA PORTUGUESA

Application of Artificial Intelligence in Efficiency Measurement

Maria Francisca da Rocha e Costa

Católica Porto Business School

2024

UNIVERSIDADE CATÓLICA PORTUGUESA

Application of Artificial Intelligence in Efficiency Measurement

Final Work in Academic Context presented to the Universidade Católica Portuguesa to obtain the Master's degree in Management with a specialization in Business Analytics

by

Maria Francisca da Rocha e Costa

under the guidance of

Aydin Teymourifar, PHD

Católica Porto Business School

April, 2024

Acknowledgments

I sincerely thank all those who have touched my life in countless ways during this significant phase, enriching it with their presence and making it a profoundly beautiful journey.

First, I am deeply grateful to my advisor, Professor Aydin Teymourifar, for his support and guidance throughout this project. His availability, knowledge and recommendations were fundamental in shaping this work. I extend my gratitude to the jury committee members for their dedication in reviewing my dissertation and for their invaluable insights and suggestions to improve it.

To my parents, my deepest thanks for giving me the opportunity to do this Master's degree and for their unconditional support throughout my whole academic career.

To my family, Francisco and my friends, thanks for all the encouragement and understanding throughout this process.

Abstract

This study evaluates the efficiency of OECD countries in achieving 14th Sustainable Development Goal - "Life underwater". Using a comprehensive model that integrates data analysis, Data Envelopment Analysis (DEA) models, and Artificial Neural Networks (ANN), the research focuses on marine resource management, specifically analyzing the impact of support to the fisheries sector on environmental indicators.

The results reveal significant variations in efficiency between OECD countries, with Sweden systematically occupying the lowest position. Correlation analysis identifies support for sectoral services as crucial, suggesting that reducing support to the fisheries sector could increase the efficiency of marine conservation. The integrated DEA-ANN approach provides a customizable framework for assessing marine sustainability, offering valuable information for policymakers and stakeholders.

Future research should extend the analysis to more countries, refine the model based on regional characteristics, and explore temporal dynamics for a comprehensive understanding of 14th SDG implementation.

Keywords: Efficiency, Artificial Neural Networks, Data envelopment analysis, OECD Countries, Marine sustainability.

Resumo

Este estudo avalia a eficiência dos países da OCDE na concretização do Objetivo de Desenvolvimento Sustentável 14 - "Vida debaixo de água". Utilizando um modelo abrangente que integra a análise de dados, modelos de Análise de Envoltória de Dados e Redes Neurais Artificiais, a investigação centra-se na gestão dos recursos marinhos, analisando especificamente o impacto do apoio ao sector das pescas nos indicadores ambientais.

Os resultados revelam variações significativas de eficiência entre os países da OCDE, com a Suécia a ocupar sistematicamente a posição mais baixa. A análise de correlação identifica o apoio aos serviços sectoriais como crucial, sugerindo que a redução do apoio ao sector das pescas poderia aumentar a eficiência da conservação marinha. A abordagem integrada DEA-ANN fornece um quadro personalizável para a avaliação da sustentabilidade marinha, oferecendo informações valiosas para os decisores políticos e as partes interessadas.

A investigação futura deve alargar a análise a mais países, aperfeiçoar o modelo com base nas características regionais e explorar a dinâmica temporal para uma compreensão abrangente da implementação do ODS 14.

Palavras-chave: Eficiência, Redes Neuronais artificiais, Análise envoltória de dados, Países da OECD, Sustentabilidade marítima.

Abbreviations

AI – Artificial Intelligence

ANN – Artificial Neural Networks

BCC – Banker, Charnes, and Cooper

CCR – Charnes, Cooper, & Rhodes

CRS – Constant Returns to Scale

DEA – Data Envelopment Analysis

DMUs – Decision-making units

DRS – Decreasing Returns to Scale

FSE – Fisheries Support Estimate

IRS – Increasing Returns to Scale

OECD – Organization for Economic Cooperation and Development

PF – Production Frontier

PPC – Production-Possibility Curve

PPF – Production Possibility Frontier

14th SDG – 14th Sustainable Development Goal

SDGs – Sustainable Development Goals

VRS – Variable Returns to Scale

Index

Figure Index

Table Index

1. Introduction

Within this section, the fundamental framework of the dissertation topic is outlined, clarifying its significance. The research objectives are introduced along with the approach used to develop them and the general structure of this work is outlined.

1.1. General Framework

The pursuit of Sustainable Development Goals (SDGs) in the field of ecology and environment has become increasingly relevant for countries from The Organization for Economic Cooperation and Development (OECD) as they strive to address pressing environmental challenges and transition towards sustainable economies. Ecological problems are increasingly an everyday issue, and it is important for countries to be concerned about their environmental indicators for safeguarding the planet's well-being, ensuring sustainable development, and preserving a habitable environment for present and future generations. Countries use various measures to assess their performance in different domains of the SDGs and several authors have studied the application of the most common efficiency measures and some combinations. To contribute to the literature, we will study the application of Artificial Intelligence (AI) in these existing approaches to overcome some of their limitations.

1.2. Objectives and Research Methodology

Concerning this issue, the final goal is to find suitable data on OECD countries to measure their efficiency in managing and utilizing their marine and coastal resources relative to their inputs. Specifically, our model would evaluate resource utilization efficiency, environmental sustainability, and balancing development and conservation.

This analysis would provide insights into how different countries perform in terms of sustainable marine and coastal resource management, considering both economic and environmental aspects. It can highlight best practices and the potential for more sustainable resource use and conservation strategies.

We want to identify the explanatory factors of efficiency and find any significant difference between the results of the different approaches used. We will investigate the benefits of hybridization of these approaches, as well as their disadvantages. This knowledge allows for a more efficient allocation of available resources, prioritizing countries with the greatest need and accelerating their convergence with these objectives.

To meet our objectives and explore the gap in the literature, throughout this research, we will answer two research questions:

(1) How does the AI improve existing efficiency measures and what advantages does it offer to overcome the limitations of traditional approaches?

(2) What are the main factors influencing efficiency in OECD countries in the context of marine and coastal resource management?

1.3. Macrostructure

This dissertation is divided into six chapters. The first introduces the topic, objectives and working methodology. The second presents the literature review, to understand the research already carried out and the gaps found. The third defines the research problem and its questions. The fourth describes the model developed. The fifth refers to the experimental results. It begins with the characterization of the SDG evaluated, the characterization of the data used, and the results of the experiment and their discussion. The sixth contains the main conclusions of this study and possible areas for future work. The last section consists of the bibliographical references used throughout this article.

2. Literature review

This section aims to clarify concepts used throughout the paper. As such, we will focus our attention on the knowledge of the Data Envelopment Analysis (DEA), focusing on some of its variants, and Artificial Neural Networks (ANN) approaches, to have deep learning about each one, as well as their specificities and limitations. Hence, we are sure that the development of the study will be done more clearly and concretely.

2.1. Efficiency vs. Productivity

Efficiency is the ability to achieve desired outcomes (outputs) while minimizing the use of resources (inputs) (Brockway et al., 2021). It evaluates the effectiveness of an asset (such as a company, organization, or economy) in transforming inputs into outputs. Efficiency involves a dual classification, each with its explanatory domain: Technical Efficiency, which assesses whether inputs are used without room for waste, thus ensuring that the production process operates at its maximum potential (Narayanan et al., 2022), and Allocative Efficiency, which explores the distribution of resources, seeking to determine whether resources are allocated in a way that corresponds to the optimal balance between inputs and outputs (Liu et al., 2023).

On the other hand, productivity is the ability to generate positive outcomes per unit of input, so it focuses on measuring the effectiveness of resource use in achieving desired results (Huang et al., 2019). The essence of productivity analysis resides in discerning whether a given entity is reaching peak production efficiency relative to its inputs and whether there are ways to improve performance without increasing inputs or decreasing outputs.

2.2. Production Frontier

Production Frontier (PF) also known as Production Possibility Frontier (PPF) or Production-Possibility Curve (PPC), is a curve in the graphical representation that illustrates the possible quantities that can be produced of two goods or services (outputs) if both depend upon the same finite resources and technology for their manufacture (Ayouba, 2023; Khezrimotlagh et al., 2020). The convexity assumption specifies that if two points belong to a PF then every point that lies on a line segment or plane joining those two points also belongs to that PF (Khezrimotlagh et al., 2020). The PF demonstrates that the production of one commodity may increase only if the production of the other commodity decreases (trade-off). Therefore, the production frontier is an essential concept for economic agents, that reflects the trade-offs between outputs, thus proving to be fundamental when making decisions in terms of production and allocation of resources.

The idea of the PF is central to two main methods that are available for performance (efficiency and productivity) assessment: parametric and non-parametric approaches (Ayouba, 2023).

2.3. Non-parametric vs. Parametric approaches

Non-parametric approaches make little or no assumptions about the underlying data distribution or the functional form of variable linkages (Wu et al., 2006). Moreover, parametric approaches involve explicitly assuming the functional form of the population distribution or the relationship between variables in advance, so they offer direct control of the model properties (Makieła & Mazur, 2022; Olesen & Petersen, 2016). The advantages and limitations of the two types of models are explored in greater detail in Appendix A.1.

2.3.1. Statistical noise

Statistical noise can be defined as unexplained variability in a sample of data (Fried et al., 2002) and is a phenomenon that arises in data due to random processes, errors, or uncertainties. It has a significant impact on statistical analysis, leading to distorted conclusions and reducing the reliability of findings. In the application of our study, statistical noise felt by DEA method will be overcome by ANN.

2.4. Data Envelopment Analysis

DEA is a non-parametric approach that uses a linear programming technique (Fried et al., 2002) to evaluate the relative efficiency of multiple decision-making units (DMUs). A DMU is a unit that produces certain outputs that result from the production of consumed inputs, which can be businesses, organizations, or any other entity participating in the manufacturing process (Wu et al., 2006). DEA provides a holistic view of efficiency by considering multiple inputs and outputs simultaneously. Furthermore, DEA not only reveals the extent of inefficiencies in the units under analysis but also quantifies the potential improvements that can be achieved for these inefficient units (Wu et al., 2006).

DEA is considered the best way to compare and analyze data since it allows efficiency to change over time and does not require any previous assumptions about the specification of the best practice frontier (Wu et al., 2006), it only requires input and output quantities. However, the DEA frontier, obtained from the dataset's top-performing DMUs, meaning the utmost level of efficiency attainable based on the observed data (Lin & Yu, 2023; Olesen & Petersen, 2016), is very sensitive to the presence of outliers and statistical noise (Bauer, 1990) ignoring the observation error (Makieła & Mazur, 2022). DEA is deterministic so has no explicit error term, and attributes all deviations from the frontier to inefficiencies, which suggests that the frontier derived from the DEA model can be deformed if the data are contaminated by statistical noise (Wu et al., 2006). These impacts would be captured by a disturbance term in a stochastic model (Fried et al., 2002).

2.4.1. Charnes, Cooper, & Rhodes, 1978

Charnes, Cooper, & Rhodes (CCR) is a DEA model using non-linear programming that provides a new vision aimed at resolving the limitations inherent in traditional DEA models, as it allows multiple inputs and outputs in the evaluation of efficiency. Remarkably, this is achieved without the need for predetermined weights (Charnes et al., 1978).

The model provides a single numerical indicator of efficiency for each participating entity by designing systematic techniques to assign weights objectively, based on observed data covering the various outputs and inputs. In addition, the model draws parallels with conventional linear programming models, facilitating computational processes. This approach allowed for a clearer understanding of the relative positions of all units, both efficient and inefficient, about the best practice frontier, as well as the possibility of gaps (Charnes et al., 1978).

2.4.2. Banker, Charnes, & Cooper, 1984

Banker, Charnes, & Cooper (BCC) model is an advance on the conventional DEA, as it encompasses the assessment of efficiency through the consideration of multiple inputs and outputs for decision-making units, emphasizing the involvement of inefficient units, and providing a more comprehensive perspective (Banker et al., 1984).

The BCC model incorporates a new approach to dealing with inefficiencies. It focuses on broadening the set of production possibilities to encompass inefficient units, while touching the efficient frontier. Therefore, it effectively quantifies the inefficiency of the decision units about the surrounding hyperplane (Banker et al., 1984).

This model aims a more comprehensive and adaptable approach to different scenarios to meet the varied needs of different DMUs. In this way, it allows the model to be better aligned with the complexities of real-world situations.

2.4.3. Input / Output oriented DEA

In general, DEA uses an input-oriented or an output-oriented approach. The input-oriented model defines a strategy for minimizing resource consumption while maintaining an equivalent level of production. The production-oriented model outlines a methodology for improving production efficiency to reach a reference level while keeping input levels constant (Khan & Karam, 2019).

2.4.4. Constant Returns to Scale and Variable Returns to Scale

DEA studies can be based on two different assumptions regarding scale efficiency (Khan & Karam, 2019). Constant Returns to Scale (CRS) and Variable Returns to Scale (VRS) both relate to how changes in input quantities affect output in a production process, but they represent different scenarios and implications. CRS refers to a situation in which a proportional change in all inputs results in an exactly proportional change in output. In the context of sensitivity analysis, CRS implies that the production process has a constant efficiency so that a variation in production factors doesn't lead to changes in production efficiency (Hicks, 1989).

On the other hand, VRS refers to a situation in which a proportional change in inputs does not result in an exactly proportional change in output (Jones, 1968). There are two subcategories: Increasing Returns to Scale (IRS) and Decreasing Returns to Scale (DRS). In the case of IRS, the increase in all factors of production leads to a more than proportional increase in production. This implies that, as a company increases its production it becomes more efficient and its average costs decrease, achieving economies of scale. In the case of DRS, the increase in all factors of production leads to a less-than-proportional increase in output (Fielbaum et al., 2023).

While CRS is often used as a simplifying assumption, VRS is used to study the actual behavior of efficiency in production processes.

2.5. Artificial Neural Networks

As good alternatives to assist in estimating efficiency frontiers for decision-makers, ANN was introduced, which plays an increasingly important role in integrated environmental modelling and has been used primarily for prediction and forecasting (Maier et al., 2023). ANN, a non-parametric approach, is a machine learning method inspired by the structure and functioning of the human brain, grouped in input, hidden, and output layers, which has gained considerable importance because it models complex physical systems, especially noisy and/or non-linear processes (Ren et al., 2022; Wu et al., 2006). The obtained input signal, a set of analyzed data (also called input vector), is processed by the network and converted into the output signal, which can be treated as the classification of the case into a certain category (Milewski et al., 2016). Therefore, ANN can be utilized to model complex relationships between inputs and outputs and predict efficiency scores.

ANN consists of artificial neurons linked by weighted connections, whose weights change during learning, with the input data being fed into the first layer, transmitting signals to the following layers. The neurons process the information and learn at the same time. The weights are adjusted in response to the results provided by the researcher, to match the network's output. The aim is to create a network that aligns the input responses with the data (Milewski et al., 2016). The findings reveal that ANN trained on the 'efficient' subset of training data exhibits superior predictive capabilities compared to one trained on the 'inefficient' subset of training data (Wu et al., 2006). When using ANN, all input and output data must be acquired and normalized (Sabiston et al., 2020).

ANN has already been shown to have applications in various fields, demonstrating their versatility and effectiveness as image recognition, patterns identification, facial recognition systems, and improving the diagnosis of medical images. ANN also plays a crucial role in autonomous vehicles, processing sensor data to make decisions in real-time, increasing safety and efficiency. In the financial sector, these models help with fraud detection, risk assessment, and algorithmic trading. In addition, ANN are applied in recommendation systems for the personalized distribution of content on platforms such as Netflix and Amazon.

ANN has some advantages over DEA, since ANN can capture non-linear relationships between inputs and outputs, which is often the case in real-world scenarios, whereas DEA assumes linear relationships, which can limit their ability to accurately model complex systems (Wu et al., 2006). Neither DEA nor ANN make assumptions about the functional form that links its inputs to outputs (Wu et al., 2006). ANN can also be adapted to various problem domains and data types, making them versatile for different applications of efficiency analysis (Wu et al., 2006). This flexibility allows them to handle both continuous and categorical data. Finally, ANN learn from data and can discover hidden patterns that may not be evident through traditional methods (Maier et al., 2023). This capability is especially useful in scenarios where the relationship between inputs and outputs is not well understood. ANN often have lower data requirements than other types of models and can deal very well with incomplete or missing data (Maier et al., 2023).

Wu et al. (2006) conclude that incorporating the ANN approach with DEA produces a more robust frontier and identifies more efficient units as more patterns of good performance are explored. Nevertheless, this combination provides worse guidance on how agents can improve their performance to different efficiency ratings. In their studies, Song et al. (2022), concluded that it overfit noisy labels owing to their high capacity in totally memorizing all noisy training samples.

Moreover, ANN often require large amounts of data for training to avoid overfitting and produce reliable results and the inputs must be selected carefully to ensure good model performance (Maier et al., 2023).

2.6. Regression

In the realm of regression models, it is essential to recognize the distinction between linear and non-linear regression approaches, as they represent distinct methodologies for modeling relationships between variables.

Linear regression is a modeling technique that seeks to establish causal relationships between a target variable and one or multiple predictor variables. When the dependent variable is correlated with other independent variables, we are dealing with a cause-and-effect relationship (Papalexopoulos & Hesterberg, 1990).

Non-linear regression delineate complex curved relationships between two variables, incorporating a combination of linear and non-linear parameters, offering greater flexibility and complexity, which makes them more challenging to develop (Afifi et al., 2019). Non-linear regression covers a diverse set of functions, including exponential and logarithmic forms, accommodating unlimited potential relationships.

2.7. Correlation

A data correlation analysis is a statistical method used to assess the strength and direction of the relationship between two variables. Essentially, it measures the closeness between two sets of data. The main result of this analysis is the correlation coefficient, which is a number between -1 and 1. A coefficient close to 1 indicates a strong positive relationship (as one variable increases, the other increases), while a coefficient close to -1 indicates a strong negative relationship (as one variable increases, the other decreases). A coefficient close to 0 indicates a null or very weak relationship. It is important to note that correlation does not imply causality; it only indicates that there is a relationship between the variables (Asuero, A. G. et al., 2006).

The results of a correlation analysis in the context of DEA can help to understand the linear relationships between different inputs and outputs. A high correlation between an input and an output does not necessarily mean that changes in the input directly cause changes in the output. There may be other underlying factors or variables that influence this relationship.

3. Problem Definition and Research Questions

3.1. Problem Definition

Managing a country today involves managing natural resources and maximizing their efficiency so that we don't waste their use and preserve their assets. Countries often face challenges in managing their marine resources due to competing economic interests, inadequate regulatory frameworks, and limited enforcement capacities. Balancing the exploitation of fisheries for economic purposes with sustainable practices to avoid overfishing is a persistent struggle. In addition, the transboundary nature of marine ecosystems complicates resource management, requiring international cooperation and effective governance mechanisms to ensure the efficient and equitable use of these valuable resources without depleting them.

Therefore, this research aims to understand how countries can improve their efficiency in the use of marine and coastal resources.

3.2. Research Questions

Considering the problem previously defined, this research aims to understand the advantages that AI can bring to the measurement of unit efficiency. To extract measurable results, this study aims to assess which factors influence the efficiency of the countries analyzed in the context of marine and coastal resource management.

The research questions addressed in this dissertation are presented in Table 3.

Research Question	**Objectives**
How does AI improve existing efficiency measures and what advantages does it offer to overcome the limitations of traditional approaches?	Evaluate how AI improves existing measures of efficiency in the management of marine and coastal resources, overcoming the limitations of traditional approaches.
What are the main factors influencing efficiency in OECD countries in the context of marine and coastal resource management?	Identify and analyze the main factors influencing the efficiency of OECD countries in the context of marine and coastal resource management, understanding what can be implemented by countries to improve their performance in this area.

Table 1- Research questions and objectives

4. Research Method

This thesis aims to develop a model through the hybridization of DEA with ANN to assess the efficiency of the outputs against the inputs used. The said approach serves as a valuable tool to overcome statistical noise and enhance the accuracy of classifications and predictions generated.

Traditional statistical methods can have difficulties in capturing complex relationships and patterns hidden in the data due to noise and uncertainties. ANN can learn from the data, detect underlying patterns, and establish complex relationships between various factors. Through a process of training on historical data, ANN can identify the most significant variables and parameters that contribute to the efficiency of outputs, also considering the specific inputs. Moreover, ANN possesses the capability to adapt and improve its performance over time. As the model encounters new data, it can continually update its internal weights and biases to refine its predictions and classifications. This adaptability ensures that the model remains robust and accurate even in the face of changing circumstances.

Then, we will apply the model created to real data from OECD countries to evaluate their efficiency concerning the 14th SDG. With the model developed through the hybridization of the various approaches mentioned above, it will be possible to classify countries with a higher level of accuracy and to have a real view of the actual position of countries about their management of environmental problems.

For the development of the model resulting from the DEA method and consequent application of ANN, MATLAB will be used.

4.1. Model

We will mathematically explain the model developed by delving into DEA and ANN approaches.

The selection of DEA models is justified by the need for a comprehensive and nuanced assessment of country efficiency. Thus, the different advantages of each model, allow for a comprehensive exploration of various dimensions of efficiency, enabling a deeper understanding of how countries are progressing towards the SDG. This diverse set of models helps to capture the multidimensional nature of sustainability, providing valuable information to policymakers, highlighting best practices, and identifying areas for improvement in different countries.

4.2. 14th Sustainable Development Goal

To apply the model developed above, we proceeded to find appropriate data on SDGs in OECD countries. As the research aims to evaluate the development of these countries in terms of the environment, we selected objective number 14 to be explored in such a way as to quantify the efficiency of each country.

According to the United Nations, this goal is about Life below water, more specifically described as: "Conserve and sustainably use the oceans, seas and marine resources for sustainable development". To better manage the development of the goal, the UN has developed certain targets and indicators, presented below (United Nations, 2022).

The 14th SDG aims to promote the conservation and sustainable use of the world's oceans, seas, and marine resources. The goal includes a set of targets and indicators to address various aspects of marine sustainability. The main objectives include preventing and reducing marine pollution, sustainably managing, and protecting marine and coastal ecosystems, mitigating the impacts of ocean acidification, regulating fishing practices to restore fish stocks, conserving coastal and marine areas, and promoting economic benefits for Small Island Developing States and Least Developed Countries through the sustainable use of marine resources. In addition, the 14^{th} SDG underlines the importance of increasing scientific knowledge, research capacity, and technology transfer to improve ocean health and support the development of developing countries. The goal emphasizes the need for international cooperation and the implementation of legal frameworks, such as the United Nations Convention on the Law of the Sea, to achieve the conservation and sustainable use of the oceans and their resources. Overall, the 14^{th} SDG aims to promote a holistic and integrated approach to ensure the health, resilience, and sustainable use of the world's oceans for current and future generations.

4.3. DEA

The DEA model defines a set of n DMUs and meticulously specifies the inputs and outputs associated with each one. Inputs, denote the crucial resources or factors used in each process, while outputs represent the results generated by each DMU. This comprehensive representation captures the essence of the DEA framework since each DMU is characterized by a vector that includes its specific inputs, designated by "x", and its outputs, designated by "y". In essence, the DEA model is an efficiency analysis tool that encompasses several sets of inputs and outputs, effectively modeling DMUs as production systems. This analytical framework encompasses n DMUs under analysis, each presenting unique consumption patterns of m distinct inputs to produce different quantities of s distinct outputs.

DEA operates based on a linear programming model for each DMU, to maximize its efficiency, ensuring that it is at least as efficient as the other DMUs. The DEA model for a DMU can be represented as follows:

$$Max\ E_{i^0} = \frac{\sum_{k=1}^{s} \lambda_k y_{k0}}{\sum_{j=1}^{m} \theta_j x_{ji}} \qquad (1)$$

$$s.t. \quad \frac{\sum_{k=1}^{s} \lambda_k y_{ki}}{\sum_{i=1}^{m} \theta_j x_{ji}} \leq 1; i = 1, \dots, n$$

$$\lambda_k, \theta_i \ \geq 0\ ;\ \ k = 1, \dots, s\ ;\ \ j = 1, \dots, m.$$

In this mathematical model, λ_k, which is the weight of output *k*, and θ_j, which is the weight of input *j*, are the decision variables to be determined. Then, each evaluation involves the assignment of an efficiency score, E, to the DMU under scrutiny, alongside the determination of weight coefficients corresponding to each reference *i*-th DMU for the calculation of the relative efficiency. This process relies on the amount of output *k* obtained by the *i*-th DMU, represented by $y_{ik,}$ and the amount of input *j* consumed by the *i*-th DMU, represented by x_{ji}. The primary aim of the DEA model is to seek out the optimal values of E and coefficient weights, ensuring that they adhere to the predefined constraints.

In the context of each DEA assessment, the model computes a crucial metric known as the relative efficiency score, e, for the specific DMU under investigation. This score is derived as the ratio between the weighted total output and the weighted total input of the DMU in question. This efficiency ratio constitutes the core objective function of the DEA model and is applied independently to each DMU within the dataset. It serves as a quantitative indicator of how efficiently each DMU transforms its inputs into outputs, enabling comparative assessments and insightful rankings among the DMUs.

The model incorporates normalization restrictions to maintain the limited nature of the efficiency values. These constraints have the crucial aim of limiting the calculated efficiency values to a range between 0 and 1. In this context, a DMU under evaluation is given "efficient" status if its calculated efficiency score is equal to the upper limit of 1, meaning that it fully uses its inputs to produce outputs. On the other hand, any DMU with an efficiency score below 1 is classified as "inefficient", suggesting potential areas for performance improvement or resource optimization. This normalization process not only standardizes efficiency scores but also facilitates meaningful comparisons and actionable insights in the context of decision-making and resource allocation.

To achieve the normalization of the model, Equation (1) is linearized, resulting in the following model:

$$max\, E_{j0} = \sum_{k=1}^{s} \lambda_k y_{ik_0}$$

$$s.t. \quad \sum_{j=1}^{m} \theta_j x_{ji_0} = 1$$

$$\sum_{k=1}^{s} \lambda_k y_{ki} - \sum_{j=1}^{m} \theta_j x_{ji} \leq 0, \forall i$$

$$\lambda_k, \theta_j \geq \varepsilon, \forall k, j$$

In this model, the decision variables do not change, while a small scalar, ε, is introduced to create a restriction preventing the weights from being equal to zero.

4.3.1. BCC vs. CCR

Both the BCC and CCR models were conceived within the Management Science tradition. In some applications of these models, this involvement permits the incorporation of management judgments into the analysis. Furthermore, the inherent complexity of real-world decision-making often introduces a degree of randomness or deviation from normality in the characteristics under consideration, which can partly stem from the unique nature of the data involved in these applications (Olesen & Petersen, 2016).

The BCC was developed as an extension of the CCR and provides a more flexible approach to efficiency analysis. In the BCC, DMUs are evaluated based on their ability to produce multiple outputs with multiple inputs while considering VRS. One of the primary distinctions is the consideration of VRS in the BCC, allowing for a more realistic assessment of efficiency. While the CCR model assumes CRS, the BCC allows for the identification of increasing or decreasing returns to scale, offering a more accurate representation of DMU performance. These advancements make the BCC a more robust tool for analyzing efficiency, which is why it will be the method used to develop our model.

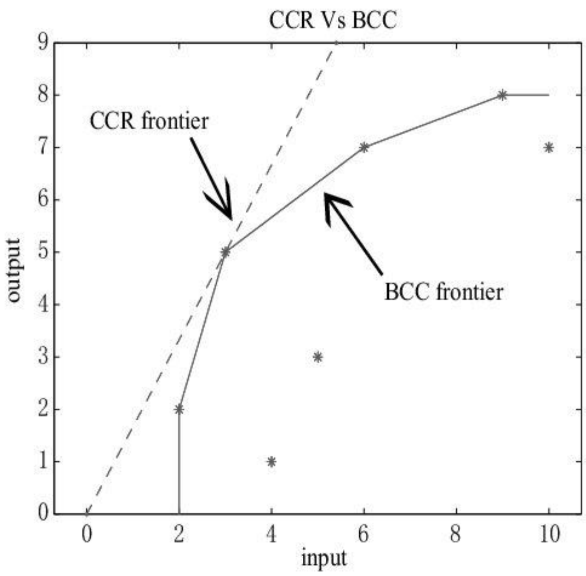

Figure 1 - Frontiers of CCR model and BCC model (Li, X. et al., 2008)

4.3.1.1. BCC – Input-oriented

The input-oriented BCC model is a widely used methodology for assessing the relative efficiency of decision-making units whose main objective is to optimize the allocation of input resources while keeping the level of outputs constant. This means that the model aims to identify the extent to which an inefficient unit can reduce its consumption of inputs while maintaining the same level of outputs, thus becoming more efficient. By comparing the input use of various units in a data set, the input-oriented BCC model provides valuable information on which units are operating most efficiently and which may have the potential to improve their performance by using their inputs more judiciously.

Accordingly, when we are in a CRS situation, both the λ and θ variables are equal to zero, while when we are in a VRS situation, both the λ and θ variables are free.

The BCC model when following the input orientation is written as follows:

$$max \sum_{k=1}^{s} \lambda_k j_{ik} + \lambda$$

$$s.t. \quad \sum_{j=1}^{m} \theta_j x_{ij} = 1$$

$$\sum_{k=1}^{s} \lambda_k j_{ik} + \lambda \leq \sum_{j=1}^{m} \theta_j x_{ij} \quad {\scriptstyle i=1,2,\ldots,n}$$

$$\lambda_k \geq 0, k = 1,2, \ldots, s$$

$$\theta_j \geq 0, j = 1,2, \ldots, m$$

The model includes the variable λ, which works as a margin for the model.

4.3.1.2. BCC – Output-oriented

The production-oriented BCC model offers a different perspective on assessing the efficiency of decision-making units. This model focuses on maximizing the production of outputs while keeping the level of inputs fixed. Thus, the main objective is to determine how efficiently the DMU can increase its outputs while keeping its inputs fixed at the current level. The production-oriented BCC model is a valuable tool for assessing the relative performance of various units in a data set and comparing their production efficiency. It helps organizations understand how they can allocate their resources to achieve higher levels of production to improve overall performance.

In this model, when we are in a CRS situation, both the λ and θ variables are equal to zero, while when we are in a VRS situation, both the λ and θ variables are free.

The BCC model when following the output orientation is written as follows:

$$min \sum_{j=1}^{m} \theta_j x_{ij} + \theta$$

$$s.t. \quad \sum_{k=1}^{s} \lambda_k j_{ik} = 1$$

$$\sum_{k=1}^{s} \lambda_k j_{ik} \leq \sum_{j=1}^{m} \theta_j x_{ij} + \theta \qquad i = 1, 2, \dots, n$$

$$\lambda_k \geq 0, k = 1,2, \dots, s$$

$$\theta_j \geq 0, j = 1,2, \dots, m$$

In this model is introduced the variable θ, which works as a margin for the model.

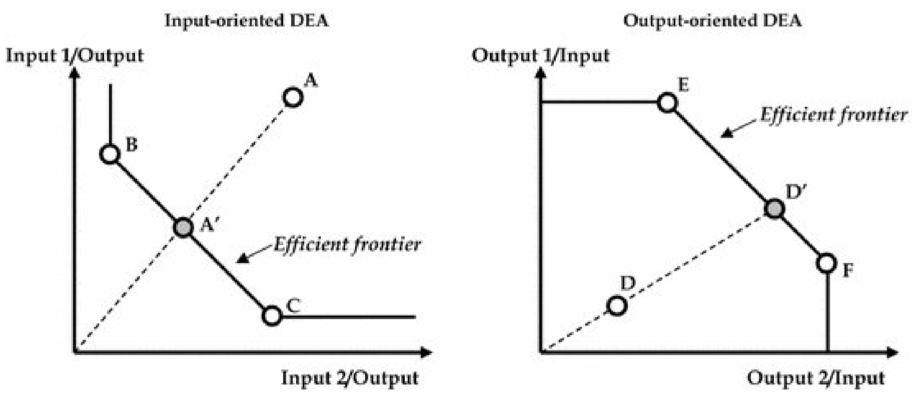

Figure 2 - Efficient frontiers for input-oriented and output-oriented DEA (Yang, F. C., 2017)

4.3.1.3. CCR Model

Thus, in a situation where there are n DMUs, with index i where i = 1, 2, …, n, and each has m inputs and s outputs, the following model is applicable. The efficiency score of each DMU_i is written as follows, representing how efficiently the DMU_i uses its inputs to produce outputs.

$$\rho_i = \frac{\sum_{k=1}^{s} \lambda_k y_{ik}}{\sum_{j=1}^{m} \theta_j x_{ij}} \leq 1$$

The objective is to maximize, so the efficiency of the j-th DMU is defined as the ratio of the weighted sum of outputs to the weighted sum of inputs, as shown in the following equation.

$$\max \sum_{k=1}^{s} \lambda_k y_{ik}$$

From the inequality that follows, the efficiency of each DMU is, at most, equal to 1.

$$\text{s.t.} \quad \sum_{j=1}^{m} \theta_j x_{ij} = 1$$

$$\sum_{k=1}^{s} \lambda_k y_{ik} \leq \sum_{j=1}^{m} \theta_j x_{ij} \ ; \ i = 1, 2, \dots, s$$

In addition, the weights of the inputs and outputs are positive, as shown in the inequalities below.

$$\lambda_k \geq 0 \ ; \ k = 1,2, \dots, s$$

$$\theta_j \geq 0 \ ; \ j = 1,2, \dots, m$$

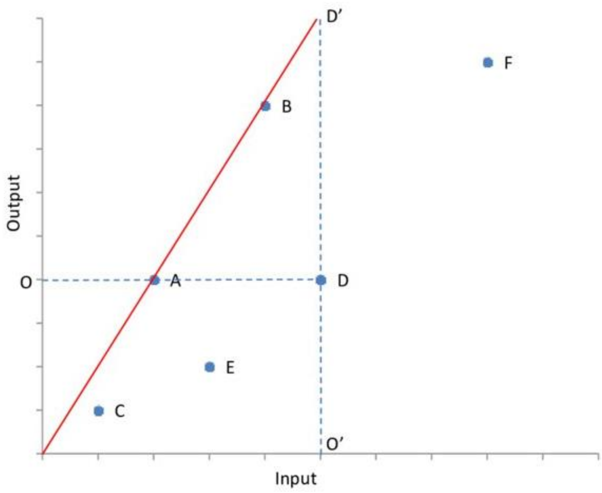

Figure 3 - CCR efficient frontier for one input and one output (Costa, C. A. B. et al., 2016)

4.3.2. DEA Additive

The authors Charnes et al. (1985) presented the additive DEA model as an extension and improvement of traditional DEA methodologies. Specifically, the additive DEA model allows for a more flexible and realistic representation of production processes by accommodating the additive nature of inputs and outputs. The model proves particularly useful in cases where the traditional DEA model is insufficient, as it allows for a more precise identification and classification of efficient decision-making units.

Traditional DEA models assume a constant return to scale, while the additive DEA model overcomes this limitation, providing a tool for assessing the efficiency of decision-making units that can operate on variable scales or have non-proportional relationships between inputs and outputs. Essentially, the additive DEA model allows for a more accurate representation of the complexities inherent in real-world production functions, offering a better understanding of the efficiency of decision-making units in various sectors.

This model is solved for each DMU_h and in this way, the frontiers are formed. It should also be pointed out that, unlike the other models, λ_i is the weight of an input or output for a DMU_j. Thus, the model is defined in the following equation and its restrictions.

$$Min \sum_{j=1}^{m} S_j^+ + \sum_{k=1}^{s} S_k^-$$

$$s.t. \sum_{i=1}^{n} x_{ij}\,\lambda_i + S_j^+ = x_{jh} \quad j = 1, \dots, m$$

$$\sum_{i=1}^{n} y_{ij}\,\lambda_i + S_k^- = y_{kh} \quad k = 1, \dots, s$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

$$S_j^+ \geq 0, S_j^- \geq 0, \lambda_i \geq 0$$

4.3.3. DEA Super-Efficiency

The DEA super-efficiency model emerged as a response to the limitations and shortcomings of traditional DEA models. Originally developed to assess the relative efficiency of decision-making units, DEA faced challenges when dealing with situations where there were several optimal solutions and it failed to distinguish between units operating on the efficiency frontier. Considering these issues, the super-efficiency model was designed to increase the accuracy of efficiency measurements. By addressing the drawbacks of traditional DEA models, such as their inability to classify fully efficient units and their sensitivity to outliers, the super-efficiency model seeks to provide a more robust framework for evaluating and benchmarking the efficiency of decision-making units in various sectors (Zhu, J. 2001).

Thus, the units evaluated are removed from the set of DMU and the evaluation is carried out on the updated frontier. In this way, the efficiency indices of the inefficient units remain unchanged, and the efficiency values of the efficient units are greater than one. Thus, the model is defined in the following equation and its restrictions.

$$Min \quad E_{i0} = \theta$$

$$\sum_{i=1}^{n} \lambda_i \, x_{ji} \leq \theta x_{j0} \quad , j = 1, \dots, m; \quad i \neq 0$$

$$\sum_{i=1}^{n} \lambda_i y_{ki} \geq y_{k0}, \qquad k = 1, \dots, s$$

$$\lambda_i \geq 0; \quad i = 1, \dots, n$$

$$\theta \; free$$

4.3.4. CRS vs. VRS

The BCC model offers two fundamental approaches to assessing efficiency: CRS and VRS. These approaches differ in the way they treat the scale of operations. In CRS, it is assumed that units operate at an optimal scale and that any deviations from this scale indicate inefficiency. VRS, on the other hand, allows for scale inefficiencies, recognizing that units may not be operating at their optimal scale. The choice between CRS and VRS depends on the context and the specific objectives of the analysis. CRS is suitable for comparing units and identifying global inefficiencies, which makes it valuable in sectors where economies of scale are vital. SRV, on the other hand, is useful for investigating the sources of inefficiency, as it distinguishes between pure technical inefficiency and scale inefficiency, which makes it applicable in sectors with varying scales of production and different constraints.

When applied to the mathematical models presented above, these differences between CRS and VRS can be seen in the free variable added to the objective function. When it comes to the CRS model, the λ and θ variables, respectively in the input and output-oriented model, are equal to zero, leaving no room for inefficiency margins. On the other hand, when it comes to the VRS model, these variables (λ and θ) are free, leaving room for inefficiencies in the DMUs.

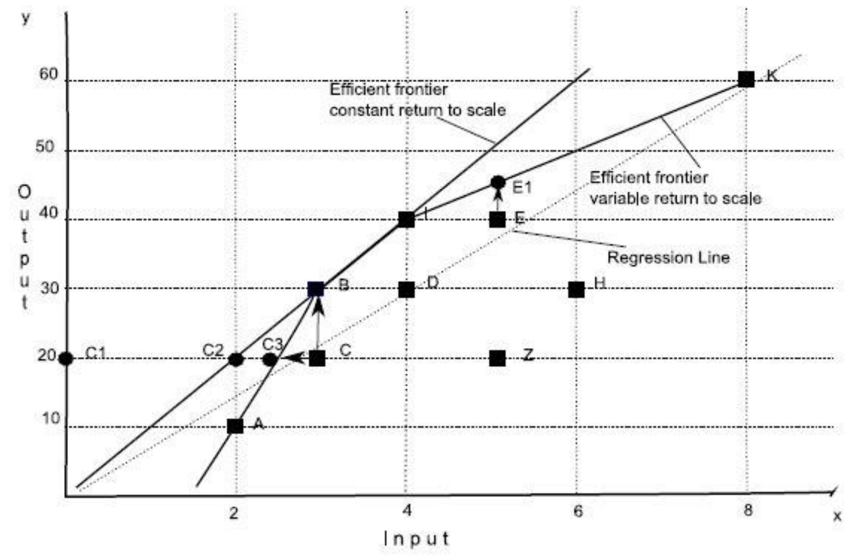

Figure 4 - DMUs, DEA and Efficient Frontier Diagram (Farantos, G. I., 2015)

4.4. Sensitive Analysis

In DEA, sensitivity analysis, known as "shadow pricing" or "dual analysis", evaluates the impact of changes in inputs and outputs on the DMUs' efficiency scores and rankings.

In DEA, DMUs receive efficiency scores based on input-output performance. Shadow prices, double values or multipliers, indicate how efficiency scores respond to adjustments in inputs or outputs, providing information for management decisions, helping to optimize resources and improve performance.

Sensitivity analysis in DEA, whether VRS or CRS, examines how changes in inputs and outputs affect efficiency, with different considerations of scale effects. Efficiency scores in VRS models can be sensitive to changes in input and output levels. Shadow prices provide information on the marginal impact of these changes on efficiency, helping to identify the optimal scale of operations for each DMU. In this way, decision-makers understand how adjusting the scale of production influences efficiency. In CRS, efficiency scores are not influenced by changes in the scale of production. Sensitivity analysis focuses on individual input and output changes, with shadow prices indicating the marginal impact on efficiency, without considering global scale adjustments.

It should be noted that a fundamental distinction between DEA and SFA is the assumption about the production frontier. DEA assumes a fixed production frontier, i.e. the efficiency frontier is considered constant and the inefficiency of DMUs is measured on the basis of their deviation from this fixed frontier, not including stochastic elements or statistical noise in the data. On the other hand, SFA incorporates stochastic elements that affect the production process, through a stochastic frontier that represents maximum production, and inefficiency is attributed to the deviation from the observed results. The choice between DEA and SFA depends on the nature of the data, the assumptions of the production process and the objectives of the analysis.

Another approach to sensitivity analysis involves a variable or adaptive frontier, estimated using ANN or regression models based on varying levels of input-output. This approach is a form of sensitivity analysis in which the efficiency of DMUs is evaluated under different scenarios or variations in input-output relationships. In this thesis, we take advantage of this approach.

4.4.1. Artificial Neural Networks

To improve the process of forming the DEA model, we include the use of AI. The ANN model is particularly suitable for recognizing complex patterns and non-linear relationships between variables, which makes it very interesting when combined with traditional methodologies.

Using the original 2020 data set, we employed DEA models to determine the efficiency scores of the DMUs for that year. For a meaningful sensitivity analysis for policymakers, we introduced variations in input levels, assuming a corresponding impact on outputs. Using NN, we forecast outputs for various input scenarios, focusing specifically on changes to a particular input (e.g. input

2). Subsequently, for each combination of inputs and predicted outputs, we rerun the DEA models to recalibrate the efficiency scores. The DEA models are run for each generated set of inputs and outputs separately. This iterative approach identifies (near) optimal input levels to maximize the efficiency of a specific DMU or a group of DMUs, providing valuable information for decision-makers looking to increase efficiency in the context examined.

Thus, by using the MATLAB tool to use the ANN model, we can comprehensively explore the behavior of countries' results when varying the factors of production according to their interests from the point of view of resource management at the national level. By systematically manipulating the main input parameters, we can simulate various scenarios and assess their impact on the DEA model's results. This dynamic analysis allows for a deeper understanding of the model's sensitivity to different factors, providing valuable information to decision-makers. Through this innovative approach, we aim not only to increase the accuracy of our efficiency assessment but also to offer a practical tool for the country's management to anticipate and define strategies in response to changing conditions in the complex field of marine and coastal resource management.

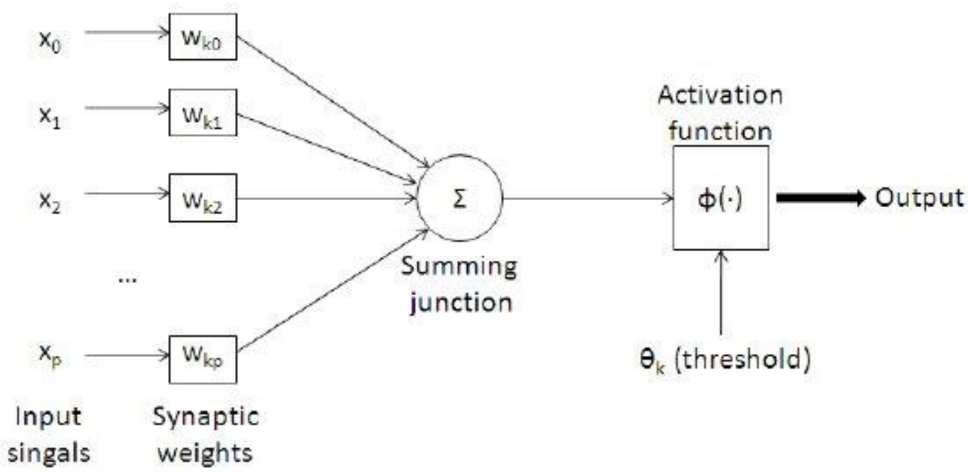

Figure 5 - Mathematical model of an ANN. (Martínez-Álvarez, F. et al., 2015)

5. Experimental Results

5.1. Data Characterization

To assess this objective, we gathered data from OECD database, that is divided into inputs and outputs to fit the model developed to assess efficiency, as well as the DMUs that are represented by each OECD country.

To begin the data processing step, we started by removing the OECD countries that do not have a coastal zone since they didn't have certain data that we want to analyze and because they would not have the same impact as the other countries on this specific objective. We also remove from the model countries whose data we need is incomplete because it would have jeopardized our analysis. Thus, the countries we included to test the model are the following: Australia, Denmark, Estonia, Lithuania, Netherlands, Poland, Spain, and Sweden.

5.1.1. Inputs

In the field of modelling and analysis, the input data section serves as the basis on which the entire structure of a thesis is built. They play a key role in shaping the model's predictive capabilities, as they allow the model to discern patterns, relationships, and trends in the data.

Thus, for the inputs, we collected independent variables that serve as the basis for our model. These include the population living in an area within 100 km from the coast of each country in 2020. The importance of this variable lies in gathering information on the influence that living in this area may have on future water conditions since the coastal population can affect and be impacted by activities related to fishing, aquaculture, and coastal protection. The work of Maul and Duedall (2019) combines the work of Burke et al. (2001), who reported the fraction of a country's population living within 100 km of the coast, with United Nations population figures. To standardize the data, for the population we used the data provided by OECD regarding the historical data that records the population of each country analyzed over the various years. We then combined this population information with the percentage of the population living on the 100km of coastline provided by the article.

Next, we have included data estimating support (in dollars) for services in the sector in 2020 (OECD, 2023). Fisheries support is defined as the annual financial transfers from governments to fisheries. The OECD developed the Fisheries Support Estimate (FSE) database to measure fisheries support policies, allowing users to compare how fisheries support differs across countries and evolves. Governments provide support to the fisheries sector through a wide range of policies, including stock management, safe working environments for fishermen, fleet capacity reduction, and, in some cases, income support for fishermen. For our model, we have specifically included the fisheries support component, which is directed to the management of available resources by the sector, within the category of support for services to the sector.

Finally, we have also included data concerning the fishing fleet, reflecting the total number of vessels in each country under analysis by 2020 (OECD 2023), i.e. the size and capacity of a country's fishing fleet. Incorporating this data allows for a holistic examination of their capacity, practices, and impact on marine ecosystems in different countries. Understanding the fleet's size and capabilities helps in quantifying a country's potential impact on marine ecosystems and the overall health of the oceans. All the inputs are summarized in Tabel 2.

Environmental Metrics	Input indicator
Coastal Population	Population living in an area within 100 km from the coast of each country (million).
Support for Services in the Sector	Financial support provided by governments to the fishing sector (million USD)
Fishing Fleet	Total number of vessels.

Table 2 - Explanation of the inputs

5.1.2. Outputs

On the other hand, within the architecture of the models, there is the output data section, a crucial dimension that encompasses the results, predictions, or knowledge derived from the intricate interaction of the input variables. Outputs are the manifestation of a model's analytical capacity, incorporating the transformation of raw data into actionable knowledge. Regarding the outputs, we selected certain dependent variables to assess the underlying impact.

Firstly, we used the variable of the percentage of marine protected areas in relation to the total area of exclusive economic zone from the various countries in the year 2020 (OECD, 2023). This output represents the extent or number of marine protected areas designated in a country's coastal waters. These areas are created to conserve and protect marine ecosystems and biodiversity.

We also used as an indicator the percentage of protected coastal area about the total area of each country's exclusive economic zone in 2020 (OECD, 2023). Like marine protected areas, this dataset provides valuable information on a country's commitment to marine conservation and sustainable resource management, making them essential components of the assessment.

Any margin of error in the latter indicators may be due to different sources of information on protected area coverage, arising from various issues, including the definitions of terrestrial, coastal, and marine areas; the national baselines used; the definition of the country; the area calculation technique used; the way protected areas recorded as points are treated; the time lag between national or regional data and WDPA updates; uncertainty about whether a particular type of protected area designation corresponds to the definition of a protected area; and which of the IUCN categories a protected area belongs to (OECD, 2021).

We also used as an output variable the total gross abstraction per capita of fresh water for each country by year until 2020, measured in cubic meters per capita. It is extracted for various purposes, which can include agricultural, industrial, and domestic use, as well as for aquaculture activities This indicator has significant implications for water resource quantity and quality issues. When using this data, it should be borne in mind that the definitions and estimation methods used may vary between countries.

Next, we include data on aquaculture production until 2020, measured in USD (OECD, 2023). Aquaculture production is defined as the production of farmed fish and crustaceans from marine and inland waters and marine tanks (OECD, 2023). Including this data is crucial as it provides a comprehensive perspective on their contributions to marine resource sustainability and conservation.

Finally, we also included data on wastewater treatment in 2020 (OECD, 2023). This data represents the percentage of the population connected to a wastewater treatment plant via a public sewage system. This indicator doesn’t consider independent private facilities, used when public systems are uneconomic. This output reflects the level of wastewater treatment and management practices in place to mitigate the environmental impact of activities in the fishing and coastal sectors. Wastewater treatment aims to conserve and sustainably use the oceans, seas, and marine resources. This data provides critical insights into the direct and indirect impacts of human activities on marine ecosystems, as well as the effectiveness of a nation's efforts to mitigate those impacts. All the outputs are summarized in Tabel 3.

Environmental Metrics	Output indicator
Marine Protected area	Percentage of marine protected area about the total area of exclusive economic zone (%).
Protected Coastal area	Percentage of protected coastal area about the total area of each country's exclusive economic zone (%).
Total freshwater abstraction	Gross abstraction per capita of fresh water for each country (m^3, millions).
Aquaculture production	Production of farmed fish and crustaceans (million USD)
Wastewater treatment	Percentage of the population connected to a wastewater treatment plant via a public sewage system (%).

Table 3 - Explanation of the Outputs

5.2. Descriptive Analysis

In the preliminary phase of this research project, a descriptive analysis of the data set was carried out, divided into inputs and outputs. This analysis involved setting up data visualization techniques to provide an in-depth understanding of the underlying data structure. By using a variety of graphical representations, we were able to reveal hidden patterns, trends, and outliers in the dataset. This visualization process not only facilitated a clearer understanding of the data, but also served as a critical basis for subsequent modelling efforts. Ultimately, this descriptive analysis not only increases the robustness of the resulting model but also underlines the importance of data exploration as an indispensable step in the research process.

Firstly, we created a dashboard with the indicators considered as inputs in our model for the year 2020. This dashboard gives us a clearer and more direct view of the data and allows us to draw some conclusions.

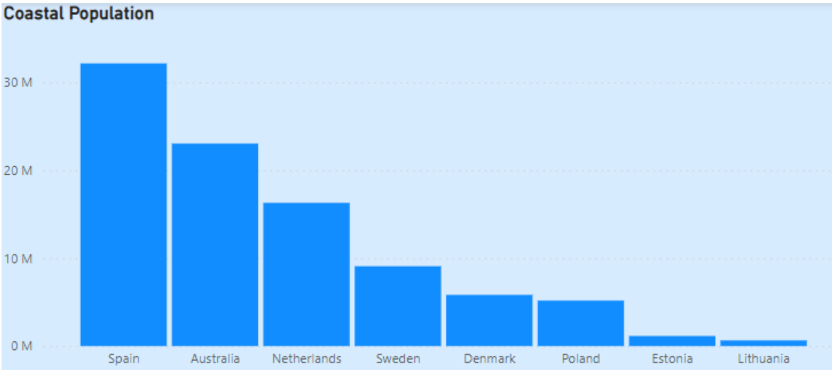

Figure 6 - Coastal population (2020)

Regarding the population of coastal areas, we can see that countries like Spain, Australia, and Netherlands stand out from the rest. The other ones have a relatively small coastal population (Figure 6).

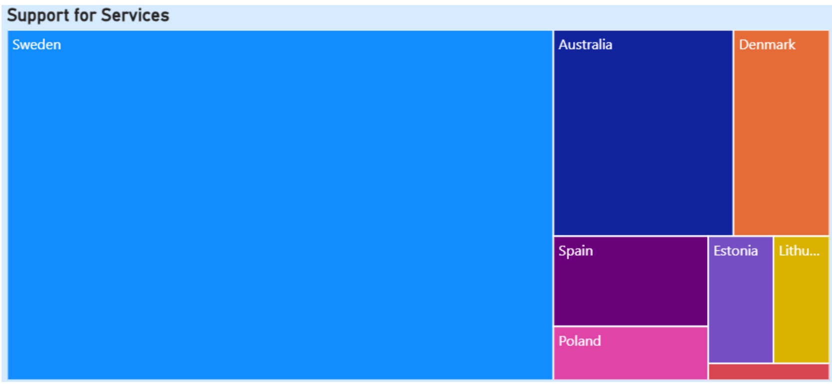

Figure 7 - Support for services (2020)

When we analyze the support given to the sector in 2020, we see that there is a clear discrepancy between Sweden and the other countries analyzed (Figure 7).

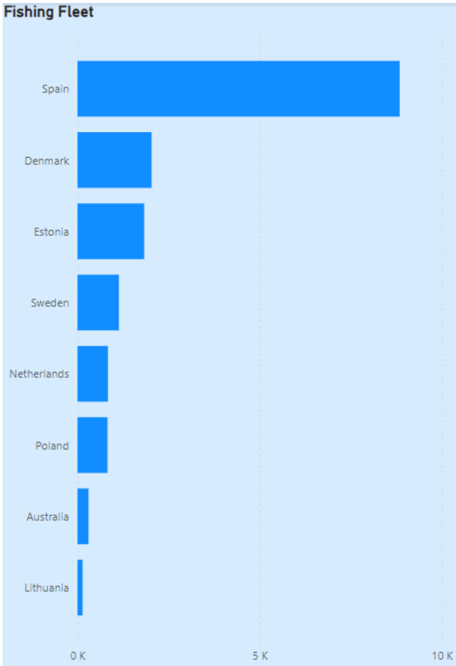

Figure 8 - Fishing fleet (2020)

Regarding the fishing fleet, we see that Spain registers a value close to 9K while the other countries all show values for this indicator below 3K units (Figure 8).

Next, we created the dashboard for the indicators considered to be the model's outputs in 2020.

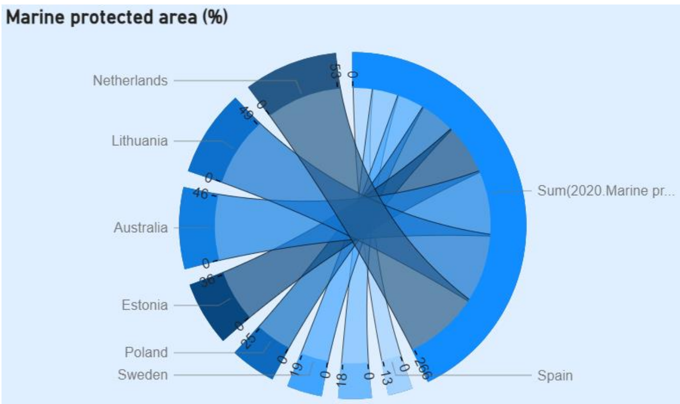

Figure 9 - Marine protected area (%) (2020)

Firstly, by analyzing the marine protected area, we can see that there are no countries that stand out considerably from the others, but the 3 countries with the highest percentage are the Netherlands, Lithuania, and Australia (Figure 9).

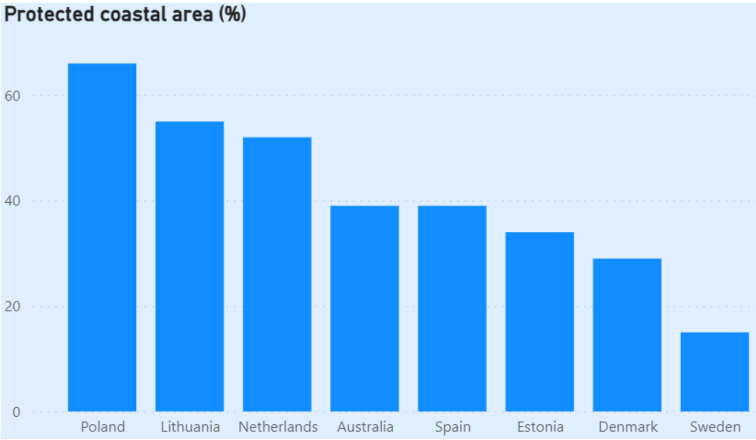

Figure 10 - Protected coastal area (%) (2020)

Regarding the protected coastal area, the same thing happens, there are no countries that stand out considerably, but there are 3 countries with percentages above 40%: Poland, Lithuania, and Netherlands (Figure 10).

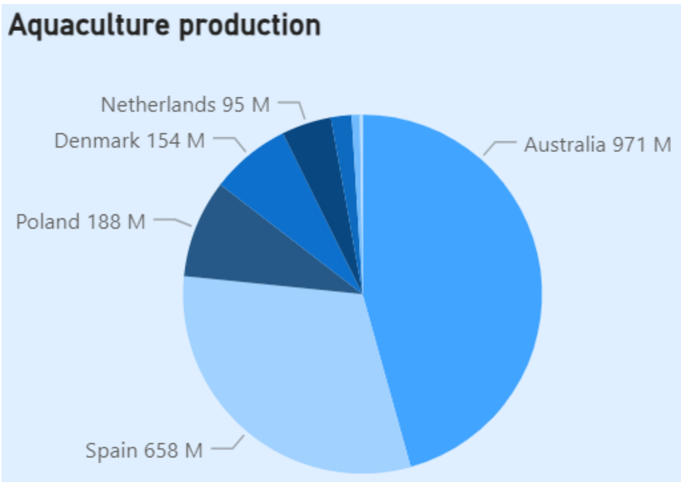

Figure 11 - Aquaculture production (2020)

In terms of aquaculture production, Australia and Spain stand out from the rest, with the indicator value for Australia being more than 5 times higher than the indicator value for Poland (the country in third position) (Figure 11).

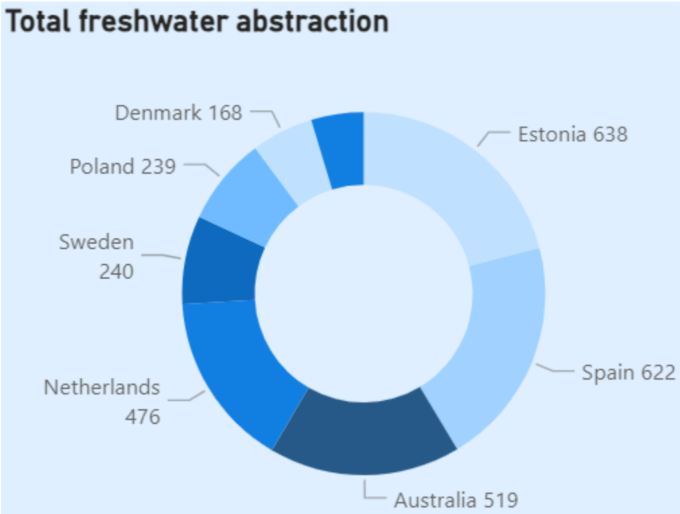

Figure 12 - Total freshwater abstraction (2020)

When analyzing freshwater abstraction, we conclude that Estonia, Spain, and Australia stand out considerably from the other countries (Figure 12).

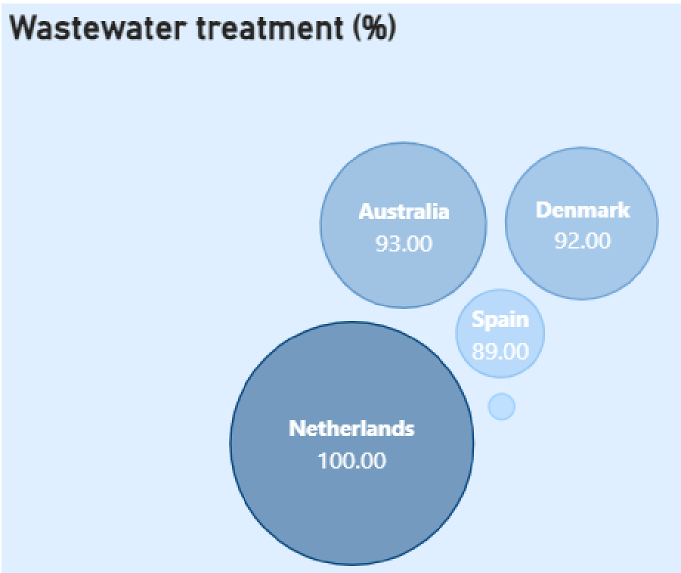

Figure 13 - Wastewater treatment (2020)

Finally, regarding the percentage of water treatment, the countries that stand out are those shown in the graph, while the rest are so low that they are not visible (Figure 13).

5.3. Results

This section aims to present the results obtained by applying the data collected to the models used.

When we applied the data on the countries to the various efficiency measurement models, we obtained the results as in Table 4.

	CCR	IO-BCC	OO-BCC	Additive	Sup-Eff
Australia	1	1	1	1	14.06
Denmark	0.67	1	1	1	0.67
Estonia	1	1	1	1	3.63
Lithuania	1	1	1	1	12.28
Netherlands	1	1	1	1	5.19
Poland	1	1	1	1	1.47
Spain	1	1	1	1	1.53
Sweden	0.2	0.77	0.96	0.90	0.20

Table 4 - The obtained results according to different models.

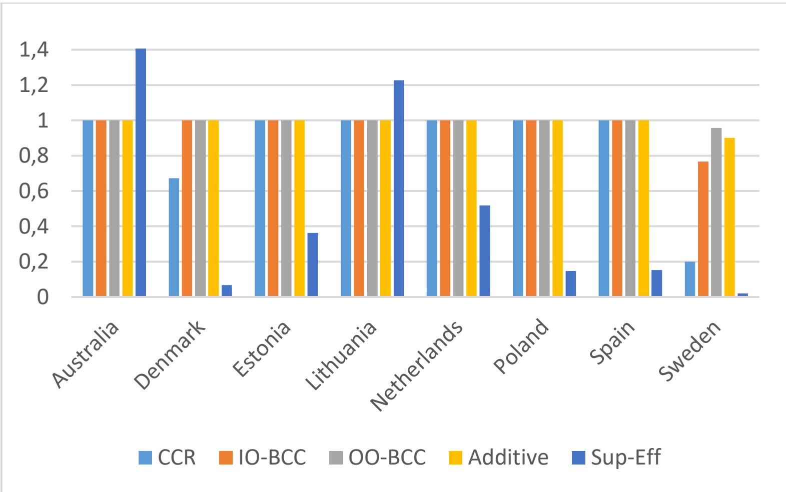

Figure 14 - Summary of Results from Table 4

Looking at Table 4 and Figure 14, we can see that Sweden has the lowest efficiency, and is also the only one with values below 1 for all models. As for Denmark, it is only inefficient when using the CCR model. On the other hand, when we look at the super efficiency model, we see that Australia and Lithuania stand out with values well above the rest of the countries.

Since Sweden has the lowest efficiency, we decided to continue our study with this country to understand how it can improve its efficiency about this SDG and when compared to the other countries. To understand the reason for Sweden's inefficiency, we analyzed the inputs and outputs of this country between the years 2011-2020, which are summarized in Table 5.

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
2020	9.08	51.45	1136	19.62	15.34	239.92	39.66	88
2019	9.01	61.15	1136	18.76	15.10	239.54	42.47	88
2018	8.92	64.38	1199	18.67	15	239.84	58.16	88
2017	8.82	38.63	1230	18.56	14.70	240.47	60.74	87
2016	8.70	32.62	1277	18.51	14.70	241.54	56.75	87
2015	8.59	36.48	1318	11.61	13.40	242.37	46.07	87
2014	8.50	45.38	1359	11.59	13.40	253.94	62.28	87
2013	8.42	46.48	1368	11.35	13	265.55	64.08	87
2012	8.35	32.08	1389	10.99	12.90	276.97	58.66	87
2011	8.29	44.81	1369	10.94	12.80	288.26	59.11	86

Table 5 - Inputs and outputs for Sweden between the years 2011-2020

To understand the behavior of Sweden's inputs and outputs, we carried out a correlation analysis of the data shown in Table 5. The results are detailed in Table 6.

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	0.57	1						
Input 3	-0.98	-0.62	1					
Output 1	0.92	0.43	-0.92	1				
Output 2	0.97	0.49	-0.95	0.98	1			
Output 3	-0.86	-0.30	0.77	-0.77	-0.84	1		
Output 4	-0.65	-0.29	0.69	-0.44	-0.54	0.49	1	
Output 5	0.86	0.66	-0.82	0.70	0.79	-0.73	-0.57	1

Table 6 - Correlation analysis between inputs and outputs of Sweden for the years 2011-2020

For policy-making purposes, we realize that attention should be focused on Input 2, which, according to Table 6, has a negative relationship with Outputs 3 and 4. From this relationship we can see that, theoretically, a decrease in input 2 would increase outputs 3 and 4 of the model. However, as can be seen in Tables 10 to 16 in Appendix B, this trend is not seen in the other countries. In addition, we can see that input 2 is the most manipulable in terms of decision-making by country management. Therefore, this will be the input of our model to focus our study on to understand how a given country can improve its efficiency.

We therefore realize that the analysis will aim to understand how Sweden's outputs behave when the value of Sector Services Support changes.

For our analysis, we used the ANN to estimate the countries' results according to different inputs. For example, Table 7 shows Sweden's results when the value of input 2 is equal to 1.1. The correlation of fisheries support provided by the Swedish government with freshwater abstraction and aquaculture production is negative. So, to understand the behavior of this country, we focused our input manipulation on reducing the support given to the sector, since the outputs were expected to increase.

Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
9.08	1.1	1136	25.55	16.02	232.99	59.24	87.51

Table 7 - The outputs for Sweden when the value of Input 2 is equal to 1.1

As can be seen in Table 8, by integrating the ANN analysis with the DEA models, we can see that Sweden's efficiency reaches a value of 1 in the CCR, IO-BCC, OO-BCC, and additive models when the value of Input 2 is equal to or less than 1.1. However, the super-efficiency model reveals that, even under these conditions, Sweden still maintains the lowest efficiency compared to other countries.

	CCR	IO-BCC	OO-BCC	Additive	Sup-Eff
Australia	1	1	1	1	0.56
Denmark	0.67	1	1	1	0.40
Estonia	1	1	1	1	0.91
Lithuania	1	1	1	1	12.28
Netherlands	1	1	1	1	4.53
Poland	1	1	1	1	0.89
Spain	1	1	1	1	0.30
Sweden	1	1	1	1	0.14

Table 8 - The efficiencies of various countries in response to Sweden reducing the value of Input 2 to 1.1.

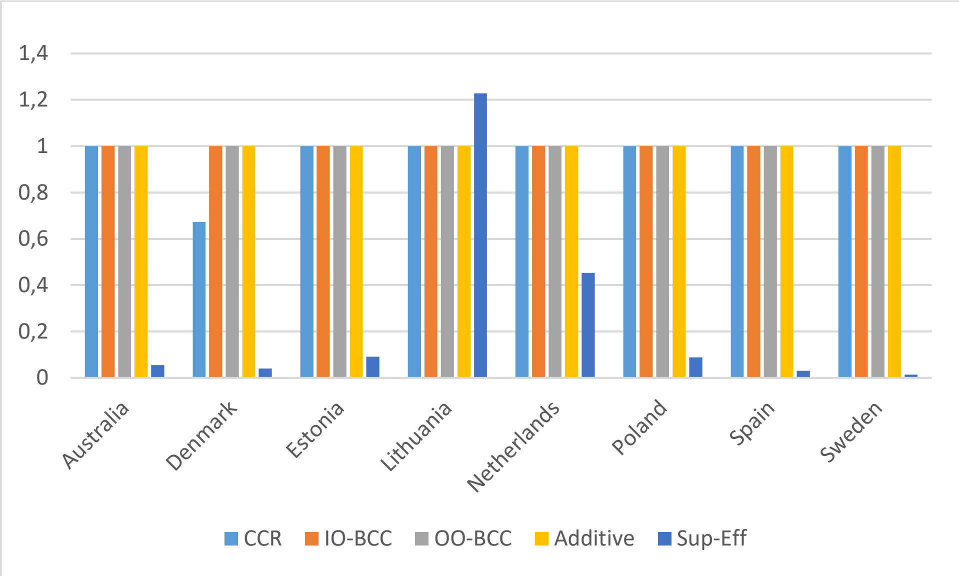

Figure 15 - Summary of results from Table 8

From the analysis we were carrying out, we realized that excluding certain outputs could help us better understand the impact of these outputs on the results of the DEA model. We therefore excluded outputs 3 and 4 from the model, and the results are shown in Table 9 and Figure 16.

	CCR	IO-BCC	OO-BCC	Additive	Sup-Eff
Australia	0.56	1	1	1	14.06
Denmark	0.40	1	1	1	0.67
Estonia	0.91	1	1	1	3.63
Lithuania	1	1	1	1	12.28
Netherlands	1	1	1	1	4.73
Poland	0.89	1	1	1	1.47
Spain	0.30	0.30	0.89	0.24	1.53
Sweden	0.14	0.77	0.96	0.84	1.02

Table 9 - The efficiencies of countries when outputs 3 and 4 are excluded from the model.

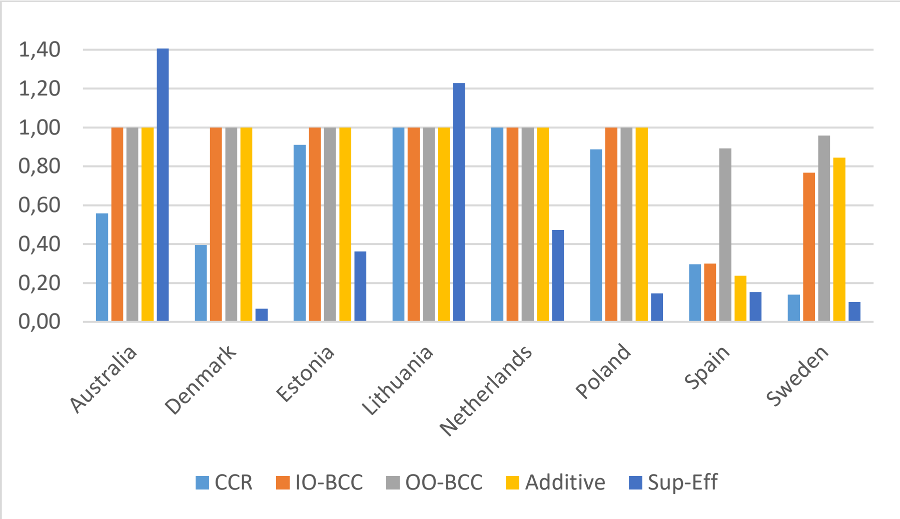

Figure 16 - Summary of results from Table 9

Table 9 shows that when we remove outputs 3 and 4 from the model, the behavior of several countries changes. In the CCR model, Sweden remains the least efficient country, but this is not the case with the other models. As we can see, the super-efficiency model results in Denmark being less efficient than Sweden.

5.4. Discussion of results

The inefficiency shown by Sweden in all the models highlights a noteworthy area for improvement. Therefore, and due to the relationship between input 2 and outputs, we focused on this input for further investigation. Our approach showed that a reduction in the support given to the sector (Input 2) could potentially lead to an increase in outputs, improving efficiency.

The efficiencies of various countries in response to Sweden's reduction in production factor 2, reveal improvements in the efficiency of most countries, but Sweden still lags in the super-efficiency model. This highlights the complexity of efficiency assessments and underlines the importance of considering several models for a comprehensive assessment.

From the analysis, we can see that greater support to the sector leads to lower total freshwater abstraction and aquaculture production. This may have several reasons. Firstly, the allocation of resources, i.e. countries can allocate a significant part of financial support to activities such as marine conservation or wastewater treatment. This allocation can limit the resources available for aquaculture or freshwater activities, which are the ones analyzed by the model. Moreover, environmental regulation, i.e. increased support for environmental services and protection, can be accompanied by stricter regulations on freshwater abstraction or aquaculture practices to ensure sustainability and reduce environmental impact. Thus, the impact of increased support will not necessarily be seen in an increase in the factors we analyzed. It may also be due to technological efficiency, where some countries may have more efficient technologies and practices for freshwater capture and aquaculture, allowing them to achieve higher levels of production with fewer resources, even when support is limited. It can also be due to geographical and environmental factors, i.e. local geography, climatic conditions, and the availability of natural resources can influence the viability and productivity of aquaculture and freshwater capture, regardless of support levels. Finally, we highlight the country's political priorities which may prioritize environmental conservation over increasing aquaculture production or freshwater abstraction, leading to a compromise between these activities.

Furthermore, our decision to exclude certain outputs reveals an interesting dimension of the analysis. The removal of outputs 3 and 4 resulted in changes in efficiency ratings, particularly in models such as CCR and super-efficiency. This suggests that the inclusion or exclusion of specific outputs can have a significant impact on a country's perceived efficiency, emphasizing the need for careful consideration and contextual relevance when designing policies and strategies to achieve the 14^{th} SDG.

6. Conclusion and Future Works

In conclusion, this research aimed to assess the efficiency of OECD countries in achieving the 14th SDG - "Life underwater" - using a model that integrated DEA and ANN. It focused on evaluating marine resource management, particularly the influence of support to the fisheries sector on various environmental indicators. The results are key to understanding the complexities of achieving the 14th SDG and provide valuable information to policymakers, researchers, and stakeholders.

The analysis of this SDG revealed remarkable variations in the efficiency of OECD countries in promoting the conservation and sustainable use of oceans and marine resources. Different DEA models were instrumental in quantifying each country's efficiency. Sweden emerged as the country with the lowest efficiency in all models, underlining the need to improve marine resource management strategies.

The integrated approach of combining DEA with ANN allowed for an understanding of the dynamics between inputs and outputs, demonstrating the versatility of the model. The model developed can be extended to other countries and can also be adjusted in the choice of inputs and outputs according to the objectives of the analysis, providing a customizable model for assessing marine sustainability.

The implications of this study for the countries' management are substantial. The identification of inefficiencies highlights the need for tailored policy interventions. Countries with lower efficiency may benefit from a reassessment of their support mechanisms for the fisheries sector to align them with sustainable practices. The study underlines the importance of considering several DEA models for a comprehensive evaluation and advocates a differentiated understanding of the relationships between inputs and outputs.

In addition, we realize that the adjustments must be considered in conjunction with the country's policies.

As for future studies, it is recommended to extended the analysis to more countries and the model is refined according to regional or country-specific characteristics. Furthermore, exploring the temporal dynamics of marine resource management and incorporating real-time data could increase the accuracy of the forecasts. Further research into the economic, social, and political factors influencing marine conservation efforts would contribute to a more holistic understanding of the implementation of the 14^{th} SDG.

In essence, this research provides a basis for advancement on sustainable development and marine resource management and our model can serve as a robust framework for evaluating and improving the efficiency of marine sustainability initiatives at a global level.

Bibliography

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of econometrics, 6(1), 21-37.

Afifi, A., May, S., Donatello, R., & Clark, V. A. (2019). Practical multivariate analysis. CRC Press.

Alshehhi, J. M., & Zervopoulos, P. D. (2023). The effect of institutional factors on environmental efficiency: A cross-country analysis using a Bayesian data envelopment analysis approach. *Journal of Cleaner Production*, 395.

Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1), 41-59.

Aydin Teymourifar., et al. "A Super-Efficiency Approach to Rank Units of a Hospital". *Acta Scientific Nutritional Health* 3.5 (2019): 15-18.

Ayouba, K. (2023). Spatial dependence in production frontier models. *Journal of Productivity Analysis*, 60(1), 21–36. https://doi.org/10.1007/s11123-023-00670-7

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078-1092.

Bauer, P. W. (1990). Recent developments in the econometric estimation of frontiers. *Journal of Econometrics*, 46(1–2), 39–56.

Brockway, P. E., Sorrell, S., Semieniuk, G., Heun, M. K., & Court, V. (2021). Energy efficiency and economy-wide rebound effects: A review of the evidence and its implications. In *Renewable and Sustainable Energy Reviews* (Vol. 141). Elsevier Ltd. https://doi.org/10.1016/j.rser.2021.110781

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 429-444. doi: 10.1016/0377-2217(78)90138-8

Costa, C. A. B., de Mello, J. C. C. S., & Meza, L. A. (2016). A new approach to the bi-dimensional representation of the DEA efficient frontier with multiple inputs and outputs. *European Journal of Operational Research*, 255(1), 175-186.

Farantos, G. I. (2015). The Data Envelopment Analysis Method and the influence of a phenomenon in organizational Efficiency: A literature review and the Data Envelopment Contrast Analysis new application. *Journal of Data Envelopment Analysis and Decision Science*, 15, 101-117.

Fielbaum, A., Tirachini, A., & Alonso-Mora, J. (2023). Economies and diseconomies of scale in on-demand ridepooling systems. *Economics of Transportation*, 34. https://doi.org/10.1016/j.ecotra.2023.100313

Fried, H. O., Lovell, C. A. K., Schmidt, S. S., & Yaisawarng, S. (2002). Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. In Journal of Productivity Analysis (Vol. 17).

Hicks, J. (1989). The assumption of constant returns to scale. In *Cambridge Journal of Economics* (Vol. 13, Issue 1). https://about.jstor.org/terms

Huang, J., Cai, X., Huang, S., Tian, S., & Lei, H. (2019). Technological factors and total factor productivity in China: Evidence based on a panel threshold model. *China Economic Review*, 54, 271–285. https://doi.org/10.1016/j.chieco.2018.12.001

Jones, R. W. (1968). Variable Returns to Scale in General Equilibrium Theory Author(s): Ronald W. In *Jones Source: International Economic Review* (Vol. 9, Issue 3).

Kao, C. (2014). Network data envelopment analysis: A review. *European journal of operational research*, 239(1), 1-16.

Khan, I. U., & Karam, F. W. (2019). Intelligent business analytics using proposed input/output-oriented data envelopment analysis DEA and slack-based DEA models for US-airlines. *Journal of Intelligent and Fuzzy Systems*, 37(6), 8207–8217. https://doi.org/10.3233/JIFS-190641

Khezrimotlagh, D., Cook, W. D., & Zhu, J. (2020). A nonparametric framework to detect outliers in estimating production frontiers. *European Journal of Operational Research*, 286(1), 375–388.

Kumar, A., Yadav, S., Gupta, P., & Mehlawat, M. (2021). A Credibilistic Multiobjective Multiperiod Efficient Portfolio Selection Approach Using Data Envelopment Analysis. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2021.3072543

Li, X., & Cui, J. (2008). A comprehensive DEA approach for the resource allocation problem based on scale economies classification. *Journal of Systems Science and Complexity*, 21, 540-557.

Liu, Y., Cui, L., Xiong, Y., & Yao, X. (2023). Does the development of the Internet improve the allocative efficiency of production factors? Evidence from surveys of Chinese manufacturing firms. S*tructural Change and Economic Dynamics*, 66, 161–174. https://doi.org/10.1016/j.strueco.2023.04.017

Makieła, K., & Mazur, B. (2022). Model uncertainty and efficiency measurement in stochastic frontier analysis with generalized errors. *Journal of Productivity* Analysis, 58(1), 35–54. https://doi.org/10.1007/s11123-022-00639-y

Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., Sànchez-Marrè, M., Acutis, M., Wu, W., & Humphrey, G. B. (2023). Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental Modelling & Software*, 105776. https://doi.org/10.1016/j.envsoft.2023.105776

Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11), 13162-13193.

Maul, G.A., Duedall, I.W. (2019). Demography of Coastal Populations. In: Finkl, C.W., Makowski, C. (eds) Encyclopedia of Coastal Science. Encyclopedia of Earth Sciences Series. *Springer*, Cham. https://doi.org/10.1007/978-3-319-93806-6_115

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, 435-444.

Milewski, R., Jankowska, D., Cwalina, U., Milewska, A. J., Citko, D., Więsak, T., Morgan, A., & Wolczyński, S. (2016). Application of artificial neural networks and principal component analysis to predict results of infertility treatment using the IVF method. *Studies in Logic, Grammar and Rhetoric*, 47(1), 33–46. https://doi.org/10.1515/slgr-2016-0045

Narayanan, E., binti Ismail, W. R., & bin Mustafa, Z. (2022). A data-envelopment analysis-based systematic review of the literature on innovation performance. In *Heliyon* (Vol. 8, Issue 12). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2022.e11925

OECD (2023), Aquaculture production (indicator). doi: 10.1787/d00923d8-en (Accessed on 29 September 2023)

OECD (2023), "Fisheries: Fishing fleet", OECD *Agriculture Statistics (database)*, https://doi.org/10.1787/data-00218-en (Accessed on 29 September 2023).

OECD (2023), Fisheries support (indicator). doi: 10.1787/1ff7e544-en (Accessed on 29 September 2023)

OECD. (2021). *OECD Statistics*. Oecd.org.

OECD (2023), Population (indicator). doi: 10.1787/d434f82b-en (Accessed on 29 September 2023)

OECD (2023), Wastewater treatment (indicator). doi: 10.1787/ef27a39d-en (Accessed on 29 September 2023)

OECD (2023), "Water: Freshwater - abstractions", OECD Environment Statistics (database), https://doi.org/10.1787/data-00602-en (accessed on 29 September 2023).

Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis - A review. In European *Journal of Operational Research* (Vol. 251, Issue 1, pp. 2–21). Elsevier. https://doi.org/10.1016/j.ejor.2015.07.058

Organization for Economic Co-operation and Development. (2023). OECD.Stat database. https://www.oecd.org/statistics/

Papadopoulos, A. (2022). The noise error component in stochastic frontier analysis. Empirical Economics. https://doi.org/10.1007/s00181-022-02339-w

Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on power systems*, 5(4), 1535-1547.

Ren, Y. M., Alhajeri, M. S., Luo, J., Chen, S., Abdullah, F., Wu, Z., & Christofides, P. D. (2022). A tutorial review of neural network modeling approaches for model predictive control. In Computers and Chemical Engineering (Vol. 165). Elsevier Ltd. https://doi.org/10.1016/j.compchemeng.2022.107956

Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, *2*(4), 367-374.

Shankar, S. (2015). Efficiency analysis under uncertainty: A simulation study. *Australian Journal of Agricultural and Resource Economics*, 59(2), 171–188. https://doi.org/10.1111/1467-8489.12055

Song, H., Kim, M., Park, D., Shin, Y., & Lee, J. G. (2022). Learning From Noisy Labels With Deep Neural Networks: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2022.3152527

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research*, 130(3), 498-509.

Tsionas, E. G., & Papadakis, E. N. (2010). A Bayesian approach to statistical inference in stochastic DEA. *Omega*, 38(5), 309–314. https://doi.org/10.1016/j.omega.2009.02.003

United Nations. (2022). Goal 14 | *Department of Economic and Social Affairs*. United Nations; United Nations. https://sdgs.un.org/goals/goal14

Wadgave, U., & Ravindra Khairnar, M. (2019). Parametric test for non-normally distributed continuous data: For and against. *Electronic Physician*, 11(2), 7468–7470. https://doi.org/10.19082/7468

Wei, T. (2013). A review of sensitivity analysis methods in building energy analysis. *In Renewable and Sustainable Energy Reviews* (Vol. 20, pp. 411–419). https://doi.org/10.1016/j.rser.2012.12.014

Wu, D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. Expert Systems with Applications, 31(1), 108–115. https://doi.org/10.1016/j.eswa.2005.09.034

Yang, F. C. (2017). Integrating corporate social responsibility and profitability into best practice selection: the case of large Taiwanese firms. *Quality & Quantity*, 51(4), 1493-1512.

Yavuz, E., & Şahin, M. (2022). Investigation of Parametric, Non-Parametric and Semiparametric Methods in Regression Analysis. https://orcid.org/0000-0003-3622-4543

Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of operational research*, 129(2), 443-455.

Appendix

A)Literature Review

A.1) Non-parametric vs. Parametric approaches

Some advantages of non-parametric methods include remarkable flexibility and resilience, drawing on their ability to circumvent the constraints imposed by rigid assumptions (Wu et al., 2006). These approaches free the analyst from the need to adhere to pre-defined probability distributions, making them particularly suitable for scenarios where the underlying data distribution remains undefined. This attribute gives researchers a versatile toolkit to deal with real-world complexities that often defy simplistic modeling assumptions.

However, it is imperative to acknowledge that non-parametric methods come with their own set of limitations as the potential requirement for larger sample sizes to achieve comparable statistical power (Alshehhi & Zervopoulos, 2023), a factor resulting from the intrinsic variability inherent in these methods. In addition, the process of estimating relationships directly from the data can have a computational cost, potentially lengthening the analysis process and affecting the efficiency of decision-making (Wu et al., 2006).

On the other hand, in parametric models, researchers make specific assumptions about the distribution of the data or how one variable relates to another. These assumptions usually rely on certain probability distributions or mathematical formulas as the basis for the analysis. Unlike the approaches presented above, these typically require a smaller volume of data points to produce robust parameter estimates, which is one of the advantages of these methods (Yavuz & Şahin, 2022). This intrinsic efficiency results from the rigorous nature of the assumptions underlying parametric models, which, when valid, generate greater statistical power, allowing researchers to detect and decipher subtle trends in the data with greater precision (Yavuz & Şahin, 2022).

However, they also present limitations such as the potential for conclusions to become skewed and untrustworthy when the underlying assumptions are not met (Papadopoulos, 2022). In addition, the selection of an inappropriate functional form can lead to problems of misspecification. Finally, these approaches base their models on linear regression which incorporates an underlying assumption of a linear correlation between the dependent and independent variables (Yan et al., 2009; Yavuz & Şahin, 2022). In addition, the basis of many parametric tests rests on the assumption that the data adhere to a normal distribution (Wadgave & Ravindra Khairnar, 2019).

A.2) Stochastic DEA

Stochastic DEA can deal effectively with noise in the data and allows the construction of finite-sample Bayes probability intervals for efficiency scores (Tsionas & Papadakis, 2010). The efficient frontier is a fundamental concept in DEA that helps identify the best-performing DMUs and defines the highest level of efficiency that can be achieved within the observed dataset (Kumar et al., 2021). These DMUs are considered efficient because they achieve the highest level of output for a given set of inputs, representing the most efficient and productive entities in the dataset.

B) Correlation analysis

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	-0.58	1						
Input 3	-0.54	0.34	1					
Output 1	0.72	-0.79	-0.61	1				
Output 2	0.94	-0.70	-0.68	0.87	1			
Output 3	-0.49	0.80	0.09	-0.50	-0.54	1		
Output 4	0.11	-0.28	0.10	0.48	0.23	-0.21	1	
Output 5	-0.65	0.25	0.62	-0.31	-0.58	0.23	0.62	1

Table 10 - Correlation analysis between inputs and outputs of Australia for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	-0.23	1						
Input 3	-0.98	0.23	1					
Output 1	0.93	-0.09	-0.89	1				
Output 2	0.91	-0.08	-0.87	1.00	1			
Output 3	-0.07	-0.12	0.04	-0.19	-0.19	1		
Output 4	0.40	-0.46	-0.37	0.36	0.38	0.30	1	
Output 5	0.99	-0.25	-0.97	0.95	0.94	-0.11	0.40	1

Table 11 - Correlation analysis between inputs and outputs of Denmark for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	0.51	1						
Input 3	-0.03	0.53	1					
Output 1	0.62	0.77	0.66	1				
Output 2	0.54	0.79	0.78	0.89	1			
Output 3	-0.64	-0.61	-0.67	-0.82	-0.91	1		
Output 4	-0.18	0.40	0.90	0.57	0.63	-0.44	1	
Output 5	0.03	0.68	0.71	0.58	0.64	-0.45	0.58	1

Table 12 - Correlation analysis between inputs and outputs of Estonia for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	-0.56	1						
Input 3	0.81	-0.40	1					
Output 1	-0.89	0.71	-0.72	1				
Output 2	-0.73	0.58	-0.74	0.91	1			
Output 3	0.82	-0.66	0.66	-0.90	-0.89	1		
Output 4	-0.95	0.48	-0.83	0.85	0.76	-0.79	1	
Output 5	-0.98	0.60	-0.79	0.90	0.70	-0.79	0.90	1

Table 13 - Correlation analysis between inputs and outputs of Lithuania for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	-0.47	1						
Input 3	-0.32	0.17	1					
Output 1	0.88	-0.32	0.01	1				
Output 2	0.92	-0.23	-0.52	0.80	1			
Output 3	-0.82	0.17	0.34	-0.84	-0.91	1		
Output 4	-0.60	0.72	0.05	-0.66	-0.45	0.60	1	
Output 5	0.96	-0.42	-0.16	0.93	0.86	-0.85	-0.61	1

Table 14 - Correlation analysis between inputs and outputs of Netherlands for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	-0.02	1						
Input 3	-0.18	0.67	1					
Output 1	-0.89	0.20	0.23	1				
Output 2	-0.90	0.15	0.13	0.74	1			
Output 3	0.96	0.01	-0.04	-0.77	-0.91	1		
Output 4	-0.92	0.03	0.10	0.67	0.94	-0.97	1	
Output 5	-0.91	0.41	0.45	0.85	0.84	-0.86	0.84	1

Table 15 - Correlation analysis between inputs and outputs of Poland for the years 2011-2020

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4	Output 5
Input 1	1							
Input 2	0.15	1						
Input 3	-0.41	-0.77	1					
Output 1	0.59	0.78	-0.91	1				
Output 2	0.17	0.79	-0.90	0.83	1			
Output 3	-0.50	-0.74	0.98	-0.90	-0.81	1		
Output 4	0.48	0.72	-0.68	0.80	0.67	-0.72	1	
Output 5	0.34	-0.02	0.31	-0.06	-0.33	0.19	0.22	1

Table 16 - Correlation analysis between inputs and outputs of Spain for the years 2011-2020