

Országos Tudományos Diákköri Konferencia 2023



Extensions to the Sum of Ranking Differences

Készítette: Ipkovich Ádám

Témavezető: Dr. Abonyi János

Pannon Egyetem, Mérnöki Kar, Folyamatmérnöki Tanszék

A pályamunka lezárásának dátuma: 2021.11.07

Contents

1	Introduction	2
2	Methods	8
2.1	Generalization of the SRD method	8
2.1.1	Data transformation	8
2.1.2	Aggregation of the variables – the generation of the gold standard .	9
2.1.3	Randomized analysis of the differences between the variables	13
2.2	SRD-based ordering of the parallel coordinates	15
2.2.1	SRD-based ordering for the reduction of intersections	15
2.2.2	MATLAB and Octave source codes	16
3	Results	19
3.1	Application example - analysis of greenhouse gas indicators	19
3.1.1	Total emissions as the gold standard	21
3.1.2	Percentile transformation	23
3.1.3	Derringer desirability function for climate change	25
3.1.4	Parallel Coordinates Visualization of Greenhouse gas emissions . . .	29
4	Discussion and Conclusion	36
4.1	Discussion	36
4.2	Summary and Conclusion	38

Abstract

This study proposes extensions to the classical sum of the ranking differences (SRD) method, both in applicability and visualization. Various problems of engineering require evaluation or decision making. The selection of the best objects becomes increasingly complex with the increase of the variables. Methods based on multi-criteria decision making were developed to solve them, such is the SRD method that focuses on the similarity of the rankings of the variables and a ranked reference to determine the relationships between the variables. The popularity of SRD algorithm is owing to the fact that it is a straightforward and easily comprehensible method. The gold standard reference can be either a selected ideal or be generated through aggregation functions. The dataset and gold standard are ranked, and the Manhattan distance is calculated. The SRD values reflect the similarity of a variable to the reference. The SRD is sensitive to the transformation of the data and is limited in terms of aggregation functions. Therefore, I extend the SRD algorithm using utility functions to transform the data in a problem-specific way. The aggregation is handled by multi-attribute utility functions requiring transformation functions to abide by independence laws. Moreover, a novel visualization technique is introduced to the SRD algorithm. The 1D ordering of the SRD values is capable of depicting the similarity of a high number of variables to the gold standard; the mutual similarities are not necessarily revealed. Hence, the parallel coordinates visualization technique is employed with the SRD, where the variables are placed as y-axes at their SRD value, the rankings of the neighboring variables are connected, and intersections measure correlation. The deviation angles between the SRD values are determined and placed left or right to the golden reference. As a use case, I employed the extended SRD to greenhouse gas emissions and economic data, with 147 countries and 40 variables present. I performed both criteria analysis and ranked nations based on the Derringer desirability function. Highly developed countries with fossil fuel-based economies such as Canada and Finland, as well as nations, namely Greece, Libya and South Africa to whom crises arose in the last decades are listed at the top of the list.

Keywords: Sum of ranking differences, Greenhouse gas emissions ranking, Data Transformation and Aggregation, Multi-attribute utility, Multi-criteria decision making, Derringer Desirability function, Parallel Coordinates; High-dimensionality data; Visualization; MATLAB/Octave Toolbox

Chapter 1

Introduction

Multi-criteria decision making (MCDM) tools are essential for the thorough analysis of complex engineering problems.

MCDM tools are frequently employed in engineering. One exact problem is materials selection, where a database of materials range from hundreds to tens of thousands possibilities and combination, but the one with the perfect properties have to be selected, otherwise the cost, durability of other essential aspect of the material may not be optimal [1, 2]. Another example set forth the question of optimal manufacturing processes, which can be selected with appropriate database and method [3]. An essential use of these methods are sustainability engineering [4], where sustainable materials [5], energy [6], climate change [7], supply chain management [8, 9] and transportation [10] are included.

This paper aims to extend the toolset of MCDM by a ranking-based method that can also be used to explore the hidden connections of the variables.

The proposed technique is based on the sum of ranking differences (SRD) method, a popular tool for chemometric and statistical problems as well as a support method for comparing MCDM results [11–13]. An example of its application is that the differences and similarities have been demonstrated in several machine learning algorithms [14]. Another example is post-Pareto analysis with swarm optimization, where the best of the Pareto optimal solutions is selected by the SRD technique [15]. In another paper, SRD as a supporting method for post-Pareto analysis provided a consensus on the solutions of eight MCDM algorithms [16]. As of September the first, 2021, Scopus lists 113 research papers in the field of SRD with an encompassing set of topics that include solutions to chemometric, agricultural, and data-science problems, however, outside of Scopus, more than 300 research papers have been published. In a recent one, the SRD and Analysis of variance (ANOVA) techniques are combined to provide a multi-criteria decision making method that selects the Pareto-optimal solutions, which are presented in 12 case studies from the field of food chemistry [17].

The SRD technique owes its popularity to being straightforward and efficient; it easily yields a unique (1D) ordering of variables and reveals grouping patterns. The steps of the method are depicted in Figure 1.1. The SRD method compares each data point with their respective reference point, collectively referred to as the gold standard. The steps consist of creating the reference, which can be an ideal or generated gold standard according to aggregation functions such as minimum, maximum, sum, arithmetic mean, *etc.* Then ranking the variables and the gold standard follows, after which their (City Block, L_1 -norm) Manhattan distance is calculated. Visualization is also trivial as the SRD value can be scaled between zero and a hundred. If the SRD value is close to zero, the criterion is ranked similarly to the ideal ranking. The maximum of the theoretical distribution can be calculated using permutations [12], and a probability distribution curve can be drawn, separating those variables that are ranked similarly to the ideal ranking and those ranked randomly as well as reversely. Variance can be rendered to individual SRD values by cross-validation.

The SRD algorithm may include transformation [18,19]. Depending on what transformation functions are applied, the structure of the data changes accordingly. Although the types of these functions include *linear*, *nonlinear* or *discrete*, classification based on the role of the function is another way to approach transformation methods. Functions may change the range of data by *scaling* (mostly multiplication/division) as well as alter the position of the data *en masse* by *centering* (mainly addition/subtraction) and *transform* the distribution of the data from one type to another (*e.g.* to a normal distribution). Transformation functions influence the result of the SRD technique, which provides a more detailed analysis of the variables. Another critical aspect of the data and the gold standard is their distribution. The most similar distribution to the standard is deemed the most accurate. If the distributions are known, the differences between the variables can be shown using statistical tests [20].

The visualization of data and their relationships are crucial to data science. Due to its straightforward use and comprehensibility, parallel coordinates is a popular visualization technique [21,22]. Axes are drafted for each variable, where the data points of each variable are depicted and connected to the respective neighbouring axis. The method visualizes the correlation between the variables as the intersection of the lines connecting the axes may indicate how different the variables are. The rank correlation coefficient (Kendall's tau, τ) measures the correlation if intersections are present [23]. If no intersections can be observed between the lines, the ordering of the two variables is identical, and their rank correlation is equal to one. To provide comprehensible visualization, the number of intersections between the axes should be reduced. Therefore, the parallel coordinates technique requires similar variables to be ordered as the correlation determines the number

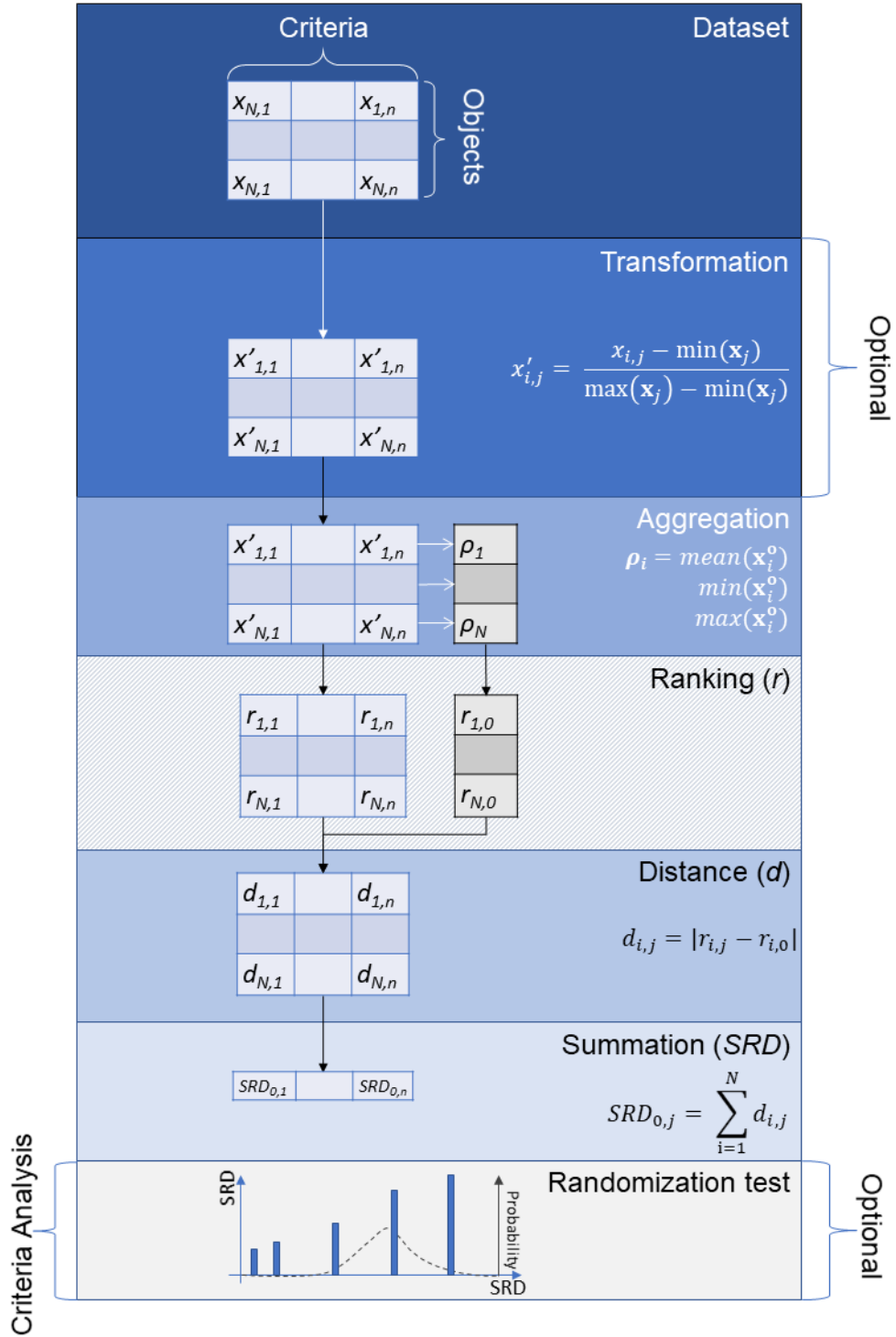


Figure 1.1: **Flowchart of SRD.** The steps can be summarized as transformation (optional depending on the generation of the gold standard), generation and ranking followed by calculating the distance between the data and the gold standard. Additionally, by comparing the criteria using randomization tests, a great deal of information can be extracted. Here, $x_{i,j}$ stands for the raw data and the transformed data is denoted by $x'_{i,j}$. The gold standard is represented by ρ_i , $d_{i,j}$ refers to the distance of the data from the gold standard and $SRD_{0,i}$ is considered the summed differences of a variable from the gold standard.

of intersections.

To provide a comprehensible visualization, the intersections of the lines can be optimized through the careful placement of the variables on the y-axes, which is achieved by measuring the similarity between the axes [24]. Multi-dimensional scaling curtails the number of intersections by reducing the variables into 2D coordinates and ordering based on their proximity to each other [25]. Employing hierarchical clustering and occlusion algorithms enable the number of lines to be reduced, improving the comprehensibility of its the visualization [26]. Some toolboxes include parallel coordinates for specific research areas such as the “Dynamics Visualisation based on Parallel Coordinates” (DYVIPAC) toolbox that was created with biological and mathematical applications in mind [27]. Moreover, parallel coordinates provide comprehensive visualization for ecological purposes, namely the effect of the proliferation of invasive species on water quality [28].

Hereinafter the use of the sum of ranking differences (SRD) method is proposed to provide an ordering based on how similar of the variables to a common reference are. The SRD values promote either the close proximity of a variable to ideal ranking or the random or reverse nature of the ranking to provide an unambiguous ordering, which may further enhance the parallel coordinates visualization technique. For the visualization of mutually similar variables, two directions namely on the left and right-hand sides on the x-axis are defined. As the deviation angle between the SRD values of the variables is another indicator of the consistency of the visualization; I developed an iterative algorithm that determines the angles between the last variables placed in two directions, and the selected variable is placed to the left- or right-hand side of the reference based on the lesser deviation angle. Additionally, the method is compared to simple, multidimensional scaling and unidirectional SRD-based orderings.

This paper presents SRD as an MCDM algorithm with broader transformation functions and aggregation implementation compared to the classical version. As the selection of an appropriate gold standard limits the usage of the SRD method, the alleviation of this limit of aggregation functions enables the reference to be tailored in a problem-specific way that benefits a broader range. I explore the influence of data transformation on rankings and the SRD algorithm as a whole. I interpret transformation as a utility function, where a specific function returns utility (preference) values concerning the variables. The construction of custom utility functions is possible; the key idea is to customize these functions depending on the properties of the variables. Moreover, I determine that the generation of the gold standard is the aggregation of utility values. The aggregation functions can be regarded as multi-attribute utility functions that require transformation functions to possess properties that enable their use. Therefore, I provide a thorough interpretation of both the transformation and the aggregation, which are employed to

generalize the SRD method.

Employing utility functions is a familiar feat in MCDM. Multi-objective linear problems are solvable by iteratively updating the approximation of the underlying utility function, and the weights of the variables may be determined based on pairwise comparisons [29]. Another use of utility functions is the evaluation of decision matrices, which results in a set of weights of the variables, with which the data can be processed by an MCDM method [30].

This research proposes a generalized method, the utility function based sum of ranking differences (uSRD). The utility functions delegate preference values to the variables, which can be analyzed similarly to the SRD.

In order to introduce the use of the extensions, I employ the uSRD with the classic 1D ordering to present the original visualization, and apply the parallel coordinates visualization algorithm on the environmental system, to explore the intricate web of emissions of various greenhouse gases and economic indicators worldwide. The analysis is based on emissions data from the database of Climate Watch [31–33]. Gross domestic product (GDP), population, percentage rural population to the total population, GDP growth and urban population growth rate are requested from the World Bank Database [34]. I aim to determine the relationships between the indicators and explain why highly ranked nations attain such dismal performance relative to other nations.

The contributions of this paper are as follows:

- I interpret transformation functions as utility functions that delegate preferences to the data, and the aggregation is presented as a multi-variate utility function with restrictions on its use.
- I propose that the selection of such transformation and aggregation functions be based on the goals and aspirations of the research, reflecting utilities and preferences.
- I introduce a utility function based generalization for the sum of ranking differences (uSRD) method, where the SRD values can only be influenced by the application of transformation functions.
- I utilize SRD as the axes of parallel coordinates, providing a thorough visualization and define two directions for dissimilar variables.
- I introduce a new visualization technique for the sum of ranking differences instead of the classical unidirectional ordering.
- I extend the parallel coordinates to two direction on the x-axis to provide information on mutually dissimilar variables.

I provide a toolbox for MATLAB and Octave to enhance the visualization of SRD-based parallel coordinates.

Chapter 2

Methods

2.1 Generalization of the SRD method

The goal of the present work is the extension of SRD. In this section, I influence the algorithm by applying transformation and aggregation functions. I regard the transformation functions as unitary utility functions and aggregation functions as multi-attribute utility functions.

2.1.1 Data transformation

$\mathbf{X}$ denotes the decision matrix, of dimensions N and n denoting the objects and variables, respectively. I can define a criterion, or column vector, as $\mathbf{x}_j = [x_{1,j}, x_{2,j}, \dots, x_{N,j}]$, $j = 1, 2 \dots n$, and an object, or row vector, as $\mathbf{x}^o_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n}]$, $i = 1, 2 \dots N$. Therefore, $x'_{i,j}$ is introduced as the transformed data and $u^j(x_{i,j})$ represents the transformation function. Transformation functions can center, scale or transform. Centering shifts the distribution and scaling changes the range of distribution, while the transformation alters the data distribution. The gold standard is denoted by ρ and 0 as an index in $SRD_{0,j}$, which stands for the SRD of j -th variable with the gold standard as the reference.

In Table 2.1, several transformation functions are defined, and each of them is discussed following the order of the table. Transformation functions are essential for data analysis as various MCDA methods require the data to be centered, scaled or transformed.

Min-max scaling [18] returns the data as a linearly scaled set between zero and one. Min-max scaling with different norms is similar but scaled according to the Manhattan (L_1 -norm), Euclidean (L_2 -norm) or generally L_p -norm distances. As the data can be compared with a chosen reference by scaling as well, a comparison can also be made with the leader [19]. Level scaling, standardization, and Pareto scaling [18] center and scale the data, which is essential for providing data within the required range and position. The table incorporates normalization functions, which reduce a criterion to a standard or a

norm, hence the name. Normalization to unit length [35] is a scaling method that divides the data by a unit distance of the variables, stabilizing the data as each scaled criterion falls within a uniform range and given a particular unit length. Rank transformation and rank-based normalization [36] is required if the ranking is to be analyzed, albeit information is lost as the data takes on a rank value and is given a uniform distance from other points. Customized rank normalization methods frequently appear in gene expression technology [37]. The Box-Cox transformation [38] is a power transformation that attempts to force the distribution into a standard normal one, however only non-negative numbers are allowed. For this reason, the Yeo-Johnson transformation [39] was designed to accept negative values.

The logarithmic transformation [40] transforms the data in a non-linear way if the data contains only positive numbers. A similar concept is a logistic function [41] for neural networks, where the input is transformed so that the neural network learns non-linearly as if real-world problems were barely linear.

Other functions, such as cumulative distributions can be perceived as transformation methods as well, along with quantile transformations. These functions are essential to describe a distribution, as the former returns the area of the probability density function (the distribution) and the latter returns the position of the data in quantiles.

The direction of the transformation functions can be changed; if the range is between zero and one, the inverse function may be provided by subtracting the function from one, multiplying it by minus one, or taking their reciprocal.

I consider transformation functions as a way of projecting the data into utility values. A utility function delegates numerical values of preference (with an added meaning) to the input. The outcome is, in many cases, a value between zero and one. Functions such as Harrington [42] or Derringer [43] desirability functions are examples of such utility functions.

2.1.2 Aggregation of the variables – the generation of the gold standard

The selection of an aggregation function depends on the intended goal of the application, which is immensely beneficial if a purposeful meaning is delegated. In Table 2.2, I compiled a set of functions for generating the gold standard, along with their properties.

The SRD technique requires a proper gold standard to provide comprehensive results as all data points are compared to their respective reference values. Without an ideal reference, the gold standard can be generated from the variables in accordance with functions such as the arithmetic mean or summation. These aggregation functions are

Table 2.1: **Transformation functions including centering, scaling.** The table contains the name, formula, the range and other properties of the functions. $\bar{x}$ denotes the arithmetic mean ($\frac{1}{N} \sum_{k=1}^N x_{k,j}$) of the criterion, while $\hat{\sigma}$ represents the estimation of its standard deviation ($\sqrt{\frac{1}{N-1} \sum_{l=1}^N (x_{l,j} - \bar{x})^2}$) of the criterion.

Name	Formula	Range/Property
Min-max scale [18]	$x'_{i,j} = \frac{x_{i,j} - \min(\mathbf{x}_j)}{\max(\mathbf{x}_j) - \min(\mathbf{x}_j)}$	$x'_{i,j} \in [0, 1]$
Min-max scale in p norm	$x'_{i,j} = \frac{x_{i,j} - \min(\mathbf{x}_j)}{\sqrt[p]{\sum_{k=1}^N (x_{k,j} - \min(\mathbf{x}_j))^p}}$	$x'_{i,j} \in [0, 1]$
Compare with leader [19]	$x'_{i,j} = \frac{x_{i,j}}{\max(\mathbf{x}_j)}$	$x'_{i,j} \in (-\infty, 1] \neq 0$ $[1, \infty); x_{i,j} \neq 0$
Level scale [18]	$x'_{i,j} = \frac{x_{i,j} - \bar{x}}{\bar{x}}$	$x'_{i,j} \in (-\infty, \infty)$
Standardization [18]	$x'_{i,j} = \frac{x_{i,j} - \bar{x}}{\hat{\sigma}}$	$x'_{i,j} \in (-\infty, \infty)$
Pareto scale [18]	$x'_{i,j} = \frac{x_{i,j} - \bar{x}}{\sqrt{\hat{\sigma}}}$	$x'_{i,j} \in (-\infty, \infty)$
Normalization to unit length [35]	$x'_{i,j} = \frac{x_{i,j}}{\sqrt{\sum_{k=1}^N (x_{k,j})^2}}$	$x'_{i,j} \in [0, 1]$
Rank transformation	$x'_{i,j} = \text{rank}(x_{i,j})$	$x'_{i,j} \in \mathbb{Z}$
Rank-based normalization [36]	$x'_{i,j} = \frac{\text{rank}(x_{i,j}) - 1}{\max(\text{rank}(x_{i,j})) - 1}$	$x'_{i,j} \in [0, 1]$
Log transformation [40]	$x'_{i,j} = \log_{10}(x_{i,j}) - \frac{1}{N} \sum_{k=1}^N \log_{10}(x_{k,j})$	$x'_{i,j} \in [0, 1]$
Box-Cox transformation [38]	$x'_{i,j} = \frac{x_{i,j}^\lambda - 1}{\lambda}; \lambda \neq 0$	$x_{i,j} \in [0, \infty)$ $x'_{i,j} \in (-\infty, \infty)$ $\lambda = [-5, 5]$
Yeo-Johnson transformation [39]	$x'_{i,j} = \frac{(x_{i,j} + 1)^\lambda - 1}{\lambda}; \lambda \neq 0$ $x'_{i,j} = \ln(x_{i,j} + 1); \lambda \neq 0$ $x'_{i,j} = -\frac{(-x_{i,j} + 1)^{2-\lambda} - 1}{2-\lambda}; \lambda \neq 2$ $x'_{i,j} = -\ln(-y + 1); \lambda \neq 2$	$x_{i,j} \geq 0$ $x_{i,j} < 0$ $x'_{i,j} \in (-\infty, \infty)$
Logistic function (sigm) [41]	$x'_{i,j} = \frac{1}{1 + e^{-x_{i,j}}}$	$x'_{i,j} \in [0, 1]$
Utility function	$x'_{i,j} = u^j(x_{i,j})$	$x'_{i,j} \in [0, 1]$

chosen depending on what is considered appropriate for the analysis.

The arithmetic means is a characteristic function used to the determine gold standard, which is a special case of the summation if the weights of the variables is not symmetric. The average performance is ranked and used to calculate the SRD, which returns a value for each criterion. If I use the median as a gold standard, it is indifferent to outliers. The exceptional or detrimental level of performance is not taken into account due to the median splitting the distribution into two equal parts, where the best and worst are in the opposite direction. Aggregating according to the minimum or maximum is also possible and is perceived as a conjunction (AND) or disjunction (OR) logical operators, respectively.

Each aggregation method requires properties to be fulfilled by the utility functions. These functions can be defined as logical or arithmetic functions, where logical functions are based on fuzzy logic and arithmetic functions are based on the utility-independent

Table 2.2: **Aggregation functions to determine the gold standard.** The properties must be satisfied by the utility functions for the aggregation to be valid. Here, $\checkmark$ denotes that a utility function is required to have such a property, $\times$ represents that such a property is not required and ρ_i stands for the respective aggregated value.

Name	Formula	Range	De Morgan	Symmetry	Dominance
Conjunction ($\wedge$)	$\rho_i = \min(x_i^{\prime o})$	$\times$	$\times$	$\checkmark$	$\checkmark$
Disjunction ($\vee$)	$\rho_i = \max(x_i^{\prime o})$	$\times$	$\times$	$\checkmark$	$\checkmark$
Additive	$\rho_i = \sum_{j=1}^n w_j \cdot x'_{i,j}$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$
Multiplicative	$1 + w\rho_i = \prod_{j=1}^n (1 + \omega w_j x'_{i,j})$	$\checkmark$	$\times$	$\checkmark$	$\times$
Preference	$\rho_i = w_j \max(x_i^{\prime o})$	$\times$	$\times$	$\checkmark$	$\checkmark$

attributes. The same range, symmetry, De Morgan's laws and dominance (having a local extremes, and a variance) are listed in Table 2.2.

Depending on the situation, not all aggregation functions are applicable. The gold standard is created from the utility values of similar utility functions that have the same range so that summation and the arithmetic mean may be coherent. However, when the utility functions differ, the utility values may not be interchangeable. The use of the same scale for each criterion is required for the application of such an aggregation. As aggregation may handle various variables, the aggregation function can be perceived as a multi-attribute utility function constructed from the transformation functions applied to the variables. In this section, transformation functions are referred to as (unitary) utility functions.

When the multi-attribute utility function is assessed, the utility functions may be monotonic. If not, the ranking becomes ambiguous, therefore, the ranking may change. Determining whether the functions are monotonic can be expressed as:

$$\text{rank}(u_j(\mathbf{x}_j)) = \text{rank}(\mathbf{x}_j), \forall j \quad (2.1)$$

The utility functions have to satisfy either additive, utility or preference independence [44–46] to be constructed into multi-attribute utility functions. In the case of the strongest, the additive independence, the functions have to satisfy the De Morgan's laws [47] so that the aggregation functions return coherent reference values. An additional consequence of De Morgan's laws is that two distinct utility functions can be replaced by one. Note that the scaling or centering transformation functions are elementary functions, which only require the same range to be used in additive aggregation, whereas transformation-type functions may not be additive. For example, an x/y type function is not additive as y is not a constant.

$$u(x_{i,j}, x_{i,m}) = u^j(x_{i,j}) + u^m(x_{i,m}) \quad (2.2)$$

where $\rho_i = u()$ denotes a multi-attribute utility function and $u^j()$ as well as $u^m()$ represent utility functions for different variables.

If the utility functions satisfy Equation 2.2, then the SRD can be extended so that the gold standard is generated as an additive aggregation with utility functions that satisfy additive independence:

$$u(\mathbf{x}_i^o) = \sum_{j=1}^n w_j \cdot u^j(x_{i,j}) \quad (2.3)$$

where w_j denotes the weight of a criterion and $\rho = u(x_{i,j})$ represents the multi-attribute utility function, whereas $x'_{i,j} = u^j(x_{i,j}), j = 1, 2 \dots n$ stands for the (unitary) utility functions.

When the utility functions are not additive independent, the second, referred to as utility independence, is to be met. Here, one of the following equations must be satisfied as utility independence is asymmetric. The symmetric property implies that the order of grouping of variables transformed by the same utility function is inessential to the aggregation. The change in the order of the variables does not alter the aggregate. Non-symmetric utility independence can be expressed as the linear transformations of a utility function with regard to the other:

$$u(x_{i,j}, x_{i,m}) = c_1(x_{i,m}) + s_1(x_{i,m})u(x_{i,j}, 0) \quad (2.4)$$

$$u(x_{i,j}, x_{i,m}) = c_2(x_{i,j}) + s_2(x_{i,j})u(0, x_{i,m}) \quad (2.5)$$

where c_1 and c_2 denote centering functions and s_1 as well as s_2 are scaling functions for one-way utility independence.

If Equations 2.4 and 2.5 are both satisfied, then the utility functions are mutually independent (symmetric):

$$u(x_{i,j}, x_{i,m}) = u^j(x_{i,j}) + u^m(x_{i,m}) + \omega u^j(x_{i,j})u^m(x_{i,m}) \quad (2.6)$$

where ω denotes a scaling constant determined by the equation of $1 + \omega = \prod_{j=1}^n (1 + \omega w_j)$. If $\omega = 0$, the utility functions are additive, therefore, additivity implies multiplicativity.

If the multiplicative condition can be satisfied, then the SRD can be extended with a multiplicative aggregation method, which results in the linear transformation of the utility function:

$$1 + \omega u(\mathbf{x}_i^o) = \prod_{j=1}^n (1 + \omega w_j u^j(x_{i,j})) \quad (2.7)$$

where $u(\mathbf{x}_i^o) = \rho_i$. If $\sum_{j=1}^n w_j = 1$, then the utility functions are additive as $\omega = 0$.

If neither independence is satisfied, the weakest one, namely the preference indepen-

dence can be met. A criterion is preferentially independent of the other variables, if the order of preference is not altered by the change in the other variables.

Preference focuses on the outcome of the utility functions determining, which variables contain higher utility values. One criterion is preferred to another ($\succ$) as long as:

$$u_j(x_{i,j}) + u_m(x_{i,j}) > u_j(x_{i,m}) + u_m(x_{i,j}) \quad (2.8)$$

$$u_j(x_{i,j}) + u_m(x_{i,m}) > u_j(x_{i,m}) + u_m(x_{i,m}) \quad (2.9)$$

If both conditions are satisfied, the criterion is preferentially independent of the other.

This aggregation is similar to logical disjunction, but weights are taken into account. In preferential independence-based aggregation, only the most preferable value of all variables is selected to be the gold standard.

$$u(\mathbf{x}^o_i) = w_j u_j(x_{i,j}) \succ w_m u_m(x_{i,m}), m = 1, 2, 3 \dots n, j = 1, 2, 3 \dots n, j \neq m \quad (2.10)$$

where u_j represents the variables with the strongest preference independence and u_m denotes the other variables.

Independence-based aggregation should be applied by taking the data into consideration. If the data cannot be analyzed by the former two independences, preferential independence can provide the necessary aggregation. If the variables are interchangeable, additive aggregation is a suitable approach. If the variables are closely related, then it is better to employ multiplicative aggregation. If additive aggregation is the strongest form of independence, the use of multiplicative and preference aggregation is possible, however, if multiplicative is the strongest, preference aggregation can still be employed.

2.1.3 Randomized analysis of the differences between the variables

To understand the similarities and groups of variables, the visualization test is introduced to determine the relationship between the variables. Moreover, statements related to the behavior of the methods and variables are provided. In order to compare the performance of the variables, their distance from a gold standard must be measured. First, the SRD values for each transformed variables is calculated:

$$SRD_{j,m} = \sum_{i=1}^N |rank(x'_{i,j}) - rank(x'_{i,m})| \quad (2.11)$$

The criterion that possesses the smallest SRD value is the most similar to the aggregate. In the classical SRD, the randomization test is implemented by approximating a correct

asymmetric distribution with a normal distribution. A histogram can be drawn from the SRD values of an immense set of random numbers that can be compared to the original SRD values. The SRD values are normalized by dividing with the theoretical maximum, which is calculated as described in Ref. [12]:

$$SRD_{max} = \begin{cases} 2(\frac{N}{2})^2 & \text{if } N \text{ is even} \\ 2\frac{N-1}{2}(\frac{N-1}{2} + 1) & \text{if } N \text{ is odd} \end{cases} \quad (2.12)$$

The visualization of the SRD values in form of the randomization test is introduced in Figure 2.1.

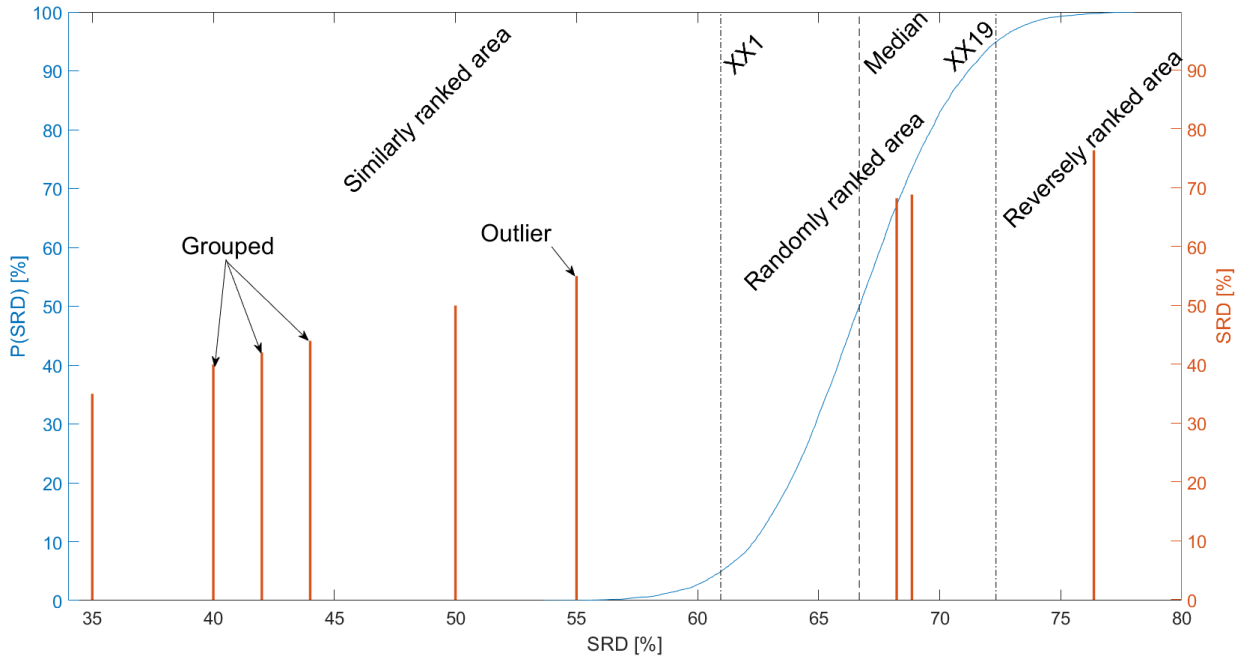


Figure 2.1: **Explanation of the randomization test.** SRD values are depicted on both axes, while $P(SRD)$ is drawn to present the probability curve of the randomly distributed numbers. SRD can range from 0 to 100 on the SRD axes and is assigned a percentage. The large empty range between zero and the first SRD value shows the possibility to an eventual improvement finding better indicators of utility function. The values to the left-hand side to the XX1 bar are the similarly ranked ones to the ideal solution. The randomly ranked variables are located between XX1 and XX19 bars, while to the right-hand side of XX19 bar, the reversely ranked variables can be found. The grouping of variables implies that they are similarly ranked, while sole variables can be perceived as outliers that are different to neighbouring groups. The area right to the XX19 bar indicates the probability of the ranking not being reversed is 5 %

In the case of the similarly ranked variables, their SRD values may form clusters. Outliers can also be observed as a single criterion, far from nearby groups. Here, the cumulative probability curve is created by Monte Carlo simulations. The positions of the

variables help to identify which are similarly, randomly and reversely ranked. If the SRD values are placed well after the 95th percentile of this curve, it is deemed to be reversely ranked. The randomly ranked variables fall between the two bars, namely XX1 and XX19, if a 5 % first kind error is allowed on both sides. At the reversely and randomly ranked areas, it can be concluded that either the data is incomparable to the gold standard, not transformed correctly or incompatible with the other variables.

2.2 SRD-based ordering of the parallel coordinates

The proposed parallel coordinates algorithm orders the variables according to their respective SRD values, before the angle of deviation is calculated between each variable and the last axes from each direction so that the intersection are minimized.

In this visualization technique, the positions of the axes are determined by the SRD value of each variable. Therefore, the distance between the axes is inconsistent, providing information on how similar the variables from the viewpoint of the golden standard can be considered to be. The rankings are depicted and the ranks of the objects are connected to their respective neighbouring ranks for each axis (variable).

2.2.1 SRD-based ordering for the reduction of intersections

Given that the number of intersections can rate how good the variables are, the relationship between two variables with similar SRD values can be appropriately described and the generated visualization classified by calculating the sum of the intersections. For parallel coordinates with N elements (ranks), the number of intersections can be calculated using the Kendall's τ , or rank correlation [48]:

$$c_{j,m} = \frac{(1 - \tau)(N(N - 1))}{4} \quad (2.13)$$

The visualization ordering can be performed in several ways. As I define the negative values of the horizontal axis as mutually dissimilar to the positive axis, where the differences between the neighbouring variables can be visualized. Kendall's τ [23] can determine how significantly they differ from each other, as it is a non-parametric measure of the association between two rankings.

If the number of axes is not two, then the summation of intersections can perfectly depict how proficient a method of ordering is, which can be divided by the theoretical maximum number of intersections:

$$\iota = \frac{\sum_{l=1}^{n-2} c_l}{(n - 2)N(N - 1)} \quad (2.14)$$

where s denotes a percentage of intersections with regard to the maximum number of intersections, c_l represents the intersections in a section and $n - 2$ is the number of sections between the axes.

Moreover, if the SRD is calculated based on Eq. 2.11, mutual dissimilarities can be showed by determining the angles between the variables. If the two respective gold standard-based SRD and the SRD of the two variables (one is defined as a reference) is constructed as a triangle, the angle between two SRD values:

$$\cos(\gamma_{j,m}) = \frac{SRD_{0,j}^2 + SRD_{0,m}^2 - SRD_{j,m}^2}{2SRD_{0,j}SRD_{0,m}}; j \neq m \quad (2.15)$$

where $\gamma_{j,m}$ denotes the angle between two gold standard-based SRD values, namely $SRD_{0,j}$ and $SRD_{0,m}$, while $SRD_{j,m}$ stands for the summed difference between the rankings of the j -th and m -th variables. Figure 2.2 represents the dissimilarity of the variables between two axes in terms of parallel coordinates visualization by measuring the angles. The method is described in Algorithm 1. As the mutual similarity between items is not necessarily revealed by the SRD method, the calculation of the deviation angle between the SRD values measures the alignment of the values and thus the differences in similarities if the triangle equality is satisfied.

In the proposed algorithm, first, the SRD values of the variables are determined, before the angles between the SRD values are calculated by Eq. 2.15 and the variables iteratively compared. On the x-axis, the normalized SRD values between zero and a hundred are plotted. The gold standard and the variable with the least SRD value are defined as starting reference axes to the left-, and right-hand sides, respectively. The angle of deviation between the variable and the reference determines on which side the variable should be placed on. If the angle of deviation with regard to the variable and the reference axis on the left-hand side is less than the other, the variable is placed to the left-hand side and selected as the following reference and *vice versa*.

A third possibility to determine differences between the variables by calculating the dissimilarity value using classical multidimensional scaling (MDS) [25] and the rank correlation. A similarity matrix is required that can be defined by the rank correlation coefficients of the variables, which is converted into dissimilarity by the equation $1 - \tau_{j,m}$ before being converted into a dissimilarity matrix. Then the data is converted into a one-dimensional distance matrix with MDS.

2.2.2 MATLAB and Octave source codes

The angle-based visualization of the SRD technique is provided in a script (.m extension). The functions that build up the visualization technique are also included in the toolbox.

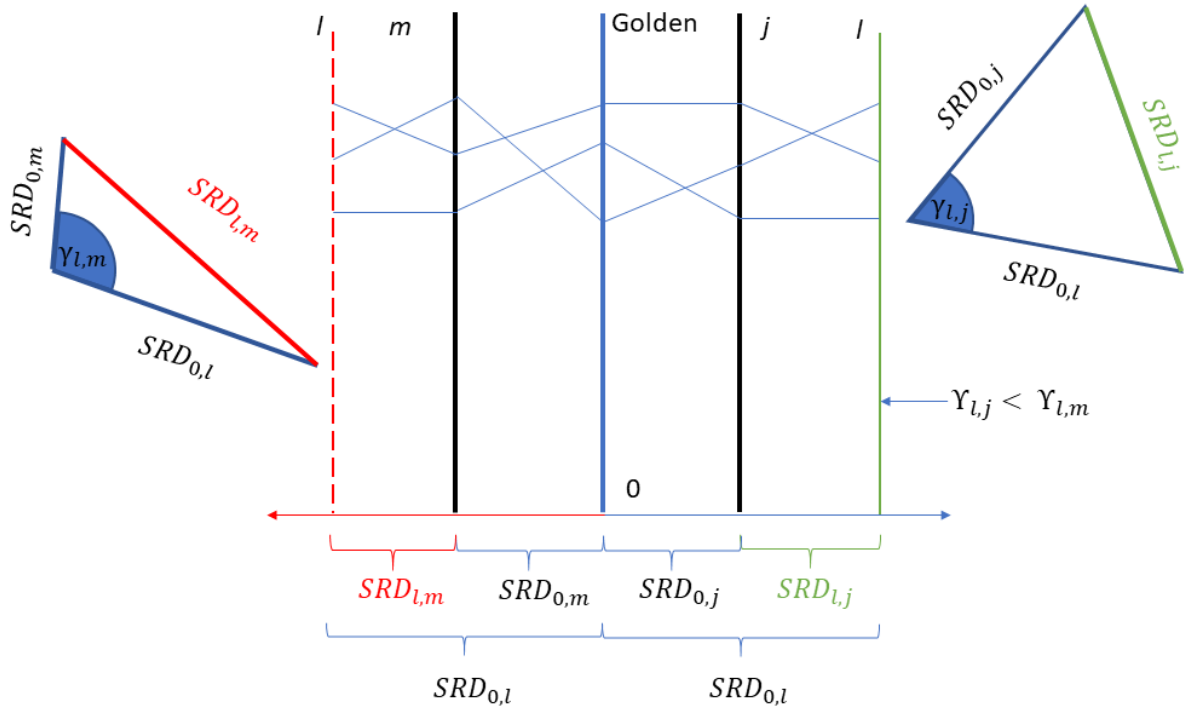


Figure 2.2: Placement of the axes based on the angle of deviation calculated with the law of cosines. The angle of deviation between the SRD values (distance from the reference axis) of one axis and its neighbouring axis can be determined by constructing a triangle with the SRD value (one angle as a reference) between the two axes as the third side. The smaller the angle of deviation is, the more mutually similar the variables are.

The scripts are only dependent on ‘tiedrank’ and ‘corr’ functions. The following functions are included in the toolbox:

- `parcoord(args)` for the visualization of parallel coordinates (in general);
- `SRD(args)` for the calculation of the gold standard-based SRD and SRD matrix;
- `srd_angle(args)` for the calculation of the angle of deviation between the j -th and m -th variables;
- `parcoord_angles(args)` for the placement of angle-based SRD in parallel coordinates;
- `pc_srd_demo.m` for a demo that presents the usage of the functions, creating the figures presented in this research.

All the functions require the gold standard and matrix of SRD value vectors to be included separately. The angle-based SRD visualization is also provided a compact function:

Algorithm 1 SRD-based visualization of parallel coordinates

```

 $N \leftarrow \dim(\mathbf{X}, 1);$ 
 $n \leftarrow \dim(\mathbf{X}, 2);$ 
 $\rho \leftarrow \text{gen\_reference}(\mathbf{X});$ 
 $SRD \leftarrow [0, SRD(\mathbf{X}, \mathbf{ref})];$ 
 $lft \leftarrow 1;$ 
 $rgh \leftarrow 2;$ 

for ( $j = 3$  to  $n + 1$ ) do  $aleft \leftarrow \text{srd\_angle}(SRD_{0,j}, SRD_{0,lft}, SRD_{lft,j});$ 
 $aright \leftarrow \text{srd\_angle}(SRD_{0,j}, SRD_{0,rgh}, SRD_{rgh,j});$ 

    if ( $aleft < aright$ ) then  $lft = j; \text{place\_left}(SRD_{0,j});$ 
    else  $rgh = j; \text{place\_right}(SRD_{0,j});$ 
    end if
end for

```

```

[tau, intersections] = parcoord_srd_angle(data, golden, axesnames,
goldenname);

```

where ‘data’ is either raw or transformed, ‘golden’ gold standard is either ideal or generated, the ‘axesnames’ denotes the labels of the variables, and ‘goldenname’ stands for the label of the gold standard.

The ‘io’ and ‘nan’ packages are required for the demo and the functions to be functional in Octave.

Chapter 3

Results

3.1 Application example - analysis of greenhouse gas indicators

The generalized method is employed on emissions indicators to analyze climate change and provide helpful remarks on the urgent problem of global warming.

The analysis is based on the Climate Analysis Indicators Tool (CAIT) [31] database of Climate Watch, which contains the total greenhouse gases (all GHGs), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and fluorinated greenhouse gases (F -gases). Sectors emitting harmful gases are elaborated on, including categories such as energy, industry and transportation. Moreover, the emission data includes the FAOSTAT land-use, change and forestry indicator provided by the Food and Agriculture Organization of the United Nations [32], as well as combustion data (other fuel combustion) that originates from the Organisation for Economic Co-Operation and Development (OECD) [33]. The economic indicators are requested from the World Bank Database [34], namely, the gross domestic product (GDP), population, percentage rural population to the total population, GDP growth and urban population growth rate. The GDP indicator is divided by the population, therefore in the analysis, it is referred to as GDP per capita. Countries with less than 500,000 inhabitants are excluded. False entries with inadequate data are removed. The final database consists of 147 nations and 40 variables from 2018. Table 3.1 contains all emission categories as well as their IDs (k) and codes. All categories are provided in metric tons of CO_2 equivalent divided by the population. An indicator is presented in the following formula: category code + ‘e’ + emission type, where ‘e’ denotes the category and emission separator.

In the parallel coordinates visualization, only the all GHGs and the CO_2 emissions are included to reduce the cluttering of the lines for exemplary purposes: First, the categories characterizing the climate change are reduced: constants (all zeroes) as well as variables

with low standard deviation are eliminated, and the variables is arranged in alphabetic order. The total number of nations and categories are 147 and 20 (excluding population), respectively.

Table 3.1: **Category lists.** This table contains the ID denoted by k as well as the name and code of the categories.

ID / k	Category name	Category Code
1	Agriculture	AGR
2	Building	BLD
3	Bunker Fuels	BNK
4	Electricity/Heat	ELH
5	Energy	ENG
6	Fugitive Emissions	FUE
7	Industrial Processes	IND
8	Land-Use, Change and Forestry	LUCF
9	Manufacturing/Construction	MAN
10	Other Fuel Combustion	OFC
11	Total excluding LUCF	TOT
12	Transportation	TRP
13	Waste	WAS

Various researches incorporates parts of the database for the purpose of a thorough analysis of emissions. The land-use, change and forestry indicator is employed to estimate the CO_2 emissions that are released as a consequence of removing forests, which resulted in the net forest intake of CO_2 to be reduced from 4.0 to 2.9 Gt per year from 1991–2000 to 2011–2015 [49]. Moreover, the methods of assessing emissions datasets are compared and analyzed to evaluate the emissions of China [50], where it was proposed that China should increase the frequency and accuracy of its emissions data. The Climate Watch system is managed by the nonprofit organization World Resources Institute that has partnered with more than 50 companies to provide solutions to sustainable development focusing on seven different areas: food, forests, water, energy, cities, climate and oceans.

I demonstrate the emissions per capita presented in Figure 3.1 to determine which countries are required to change their lifestyle the most. The Figure includes more than 147 nations as their total emissions per capita indicators have been published by Climate Watch [31]. Even though micronations seem significant in terms of emissions per capita, they are not of statistical significance due to the magnitude of their distinctive emissions and populations, therefore, were omitted in both the ranking and the criteria analysis. The average emissions is calculated as 6.26 tons per capita.

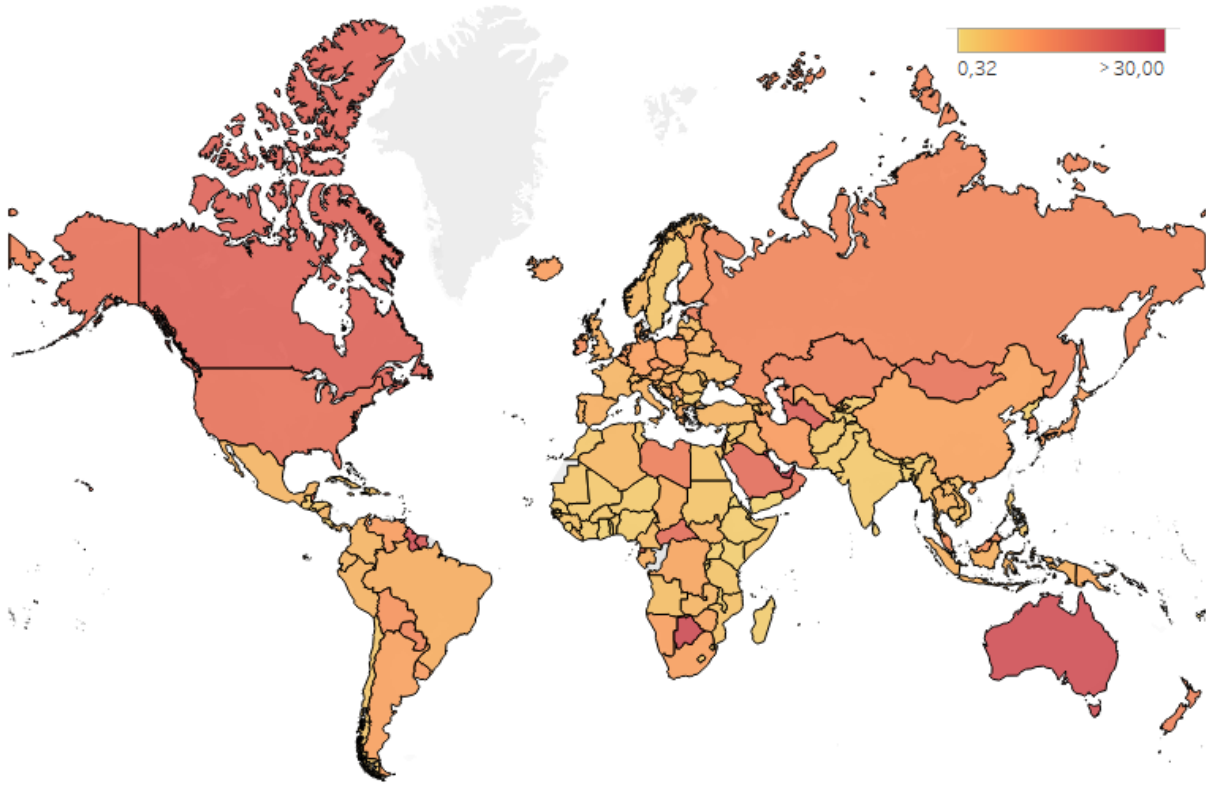


Figure 3.1: **Map of the total emissions per capita.** Although the range of emissions is between 0.32 and 70 tons per capita, each country with emissions of more than 30 is assigned the same color. The average emissions is 6.26 tons per capita.

3.1.1 Total emissions as the gold standard

I employed the SRD algorithm to analyze the relationship between the variables. Figure 3.2 presents the SRD against the total emissions per capita as the gold standard.

As is illustrated in Figure 3.2, the majority of the indicators are ranked similarly to the total emissions (TOT), including the indicator GDP per capita. The ‘all GHG’ and CO_2 energy (ENG) category as well as the CO_2 total emissions (TOT) are ranked the most similarly to the total ‘all GHG’ emissions. According to the database, about 78 % of the total emissions is energy-related. The GDP per capita indicator is located in the vicinity of the fields of electricity (ELH), transportation (TRP) as well as manufacturing/construction (MAN). Industry (IND) and building (BLS) energy can be found on the other side of the GDP per capita, indicating that higher and more advanced production lines lead to more value creation and proposes that the positive-sum game is played. However, sustainability must not be neglected. The closeness of industry (INDe CO_2 , INDe N_2O) and waste (WASeAllGHG) indicators propose a connection.

Waste is a by-product of human life such as from packaging, household, biological and industrial waste which are integral to everyday life. Wastewater treatment and safe processing of excrement are primary sources of N_2O , whose magnitude depends solely

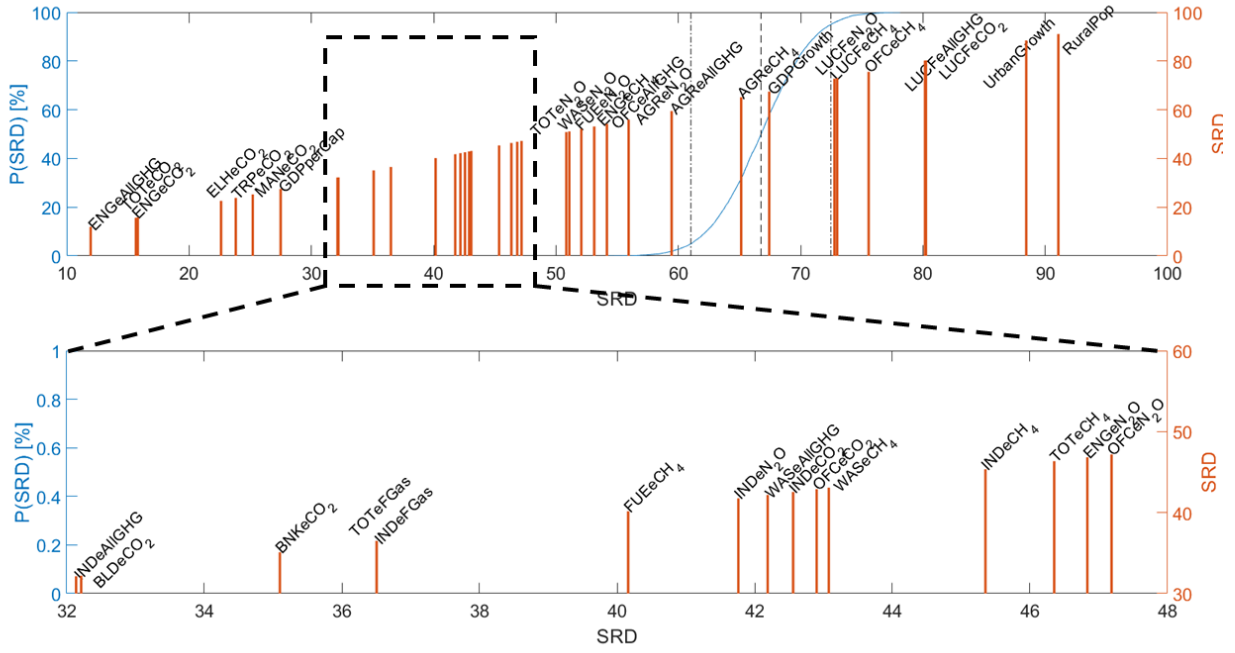


Figure 3.2: **Depiction of the SRD to represent the relationship between the emissions categories in tons per capita and the all greenhouse gases emissions as the gold standard.** Most categories are located in the similarly ranked area. The energy category of ‘all GHG’ and CO_2 is the most similar to the total emissions. Agricultural CH_4 , and GDP growth rate are ranked randomly, while the LUCF, urban growth and rural population indicators are ranked in reverse order.

on the population and water treatment technology. The fugitive emissions indicator is in the vicinity of the N_2O waste (WAS) indicator. Various wastewater treatment plants use fossil fuels for on-site sludge disposal [51], and the unwanted waste can be burnt or disposed of in landfills. The centralized wastewater management is incredibly efficient compared to local wastewater management [52], therefore the investment in researches that cost-effectively eliminates all forms of pollution from water and the by-product N_2O in water treatment plants is essential to tackle the greenhouse gases that result from the incorrect handling of some waste particles.

GDP growth rate and agricultural (AGR) methane are placed in the randomly ranked area. As the most polluting category of agricultural methane is cattle and cattle-related manure management [53], it is crucial to focus on the reduction of the by-product emissions of enteric fermentation, due to the digestion process in herbivorous animals. Meat is one of the staple foods in the West and it would be most reasonable to innovate the production of meat and milk. Such example is cultured meat that is produced by tissue engineering processes. The environmental impact of such meat was examined in 2012 [54] which concluded that *in vitro* meat production produces 78–96 % less greenhouse gas emissions along with reduced land use. Research suggests that the increase in meat consumption and

GDP growth are connected [55], as low-income nations export meat cheaply moreover, the farming techniques systems are obsolete in developing nations. In contrast, in developed nations the income levels from the production of lamb and poultry are not as preferable, therefore, the levels of production of pork and beef are sustained, resulting in heavier pollution.

In the reversely ranked area, the rural population (RuralPop), urban growth rate (UrbanGrowth), land-use, change and forestry (LUCF) indicators as well as CH_4 from other fuel combustion (OFC) are placed. The deforestation and change in land use pollute clean water and reduce biodiversity as well as healthy soils. Land reclamation of wetlands is also significant for emitting greenhouse gases as the removal results in the emissions of an immense CO_2 . Wetlands store CO_2 [56], and emit some amount of methane and nitrous oxide in exchange. Therefore the destruction of wetlands is a natural source of CO_2 , but the regrowth of forests and wetlands is becoming a hot topic. A possible explanation for the reverse nature of urban growth and rural population indicators may be the environmental Kuznets curve [57], according to which the GDP per capita and environmental deterioration are non-linearly correlated. Higher GDP per capita reduces emission due to the shift from the industrial to the service sector. Emissions are rarely significant in a developing country, where the population lives off agriculture and can barely afford electricity.

3.1.2 Percentile transformation

Next, I transformed the data into percentiles. This transformation presents which nations are placed in the upper, more problematic percentiles. The total ‘all GHG’ emissions category is also included in the analysis as well. I employed the SRD technique with the data transformed by an empirical distribution function to convert the values to the respective percentiles, while the arithmetic means and the maximum were employed as gold standards. Figures 3.3 and 3.4 are assessed to present the relationship between the variables.

The order of the variables is similar to Figure 3.2 to the point that no significant change can be mentioned. Even though the range of the SRD values is smaller than that recorded by the first approach, the algorithm does not differ significantly in distinctiveness. The distances between the variables of the groups variables are more substantial and therefore more distinguishable. In Figure 3.4 I analyzed the percentiles by also using the maximum as the gold standard.

The analysis of Figure 3.4 suffers from a small range and excessive similarity; the SRD values range from 48.32 to 68.57. The variables are lumped together, and many take on values in the vicinity of other categories. Employing the maximum as the aggregate is

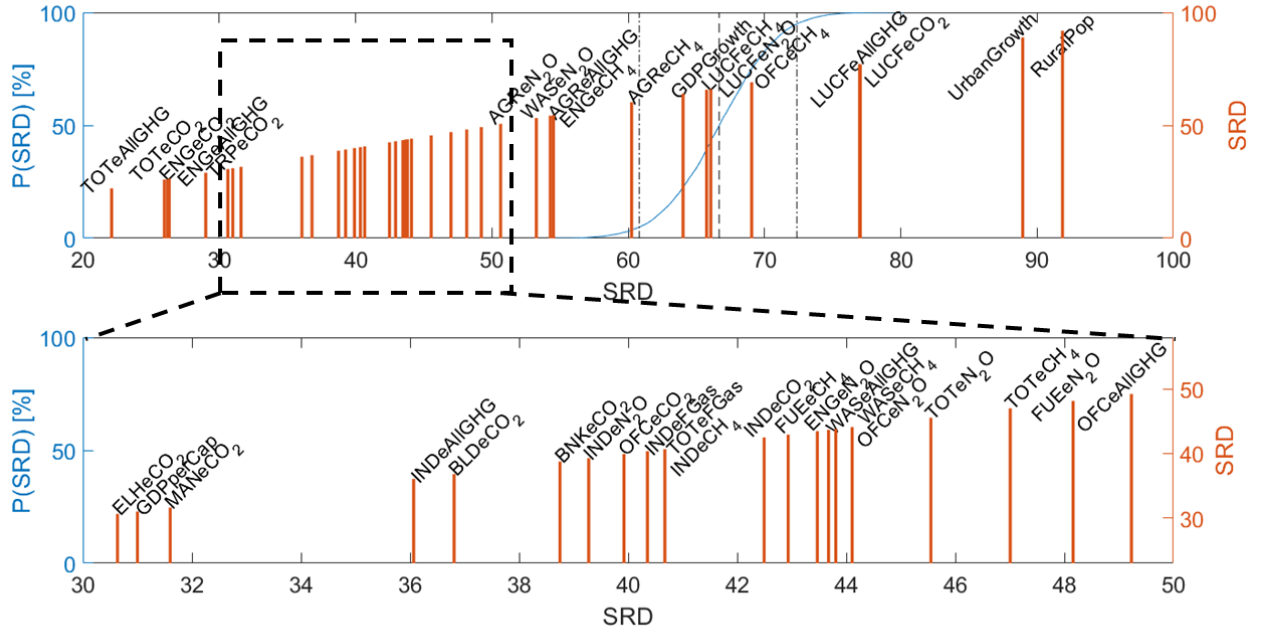


Figure 3.3: **Depiction of SIRD based on percentile transformation and the arithmetic mean as the aggregate.** Although the order of the categories is similar to that found in Figure 3.2, here the arithmetic mean of the categories is the aggregate. The total ‘all GHG’ indicator is also included.

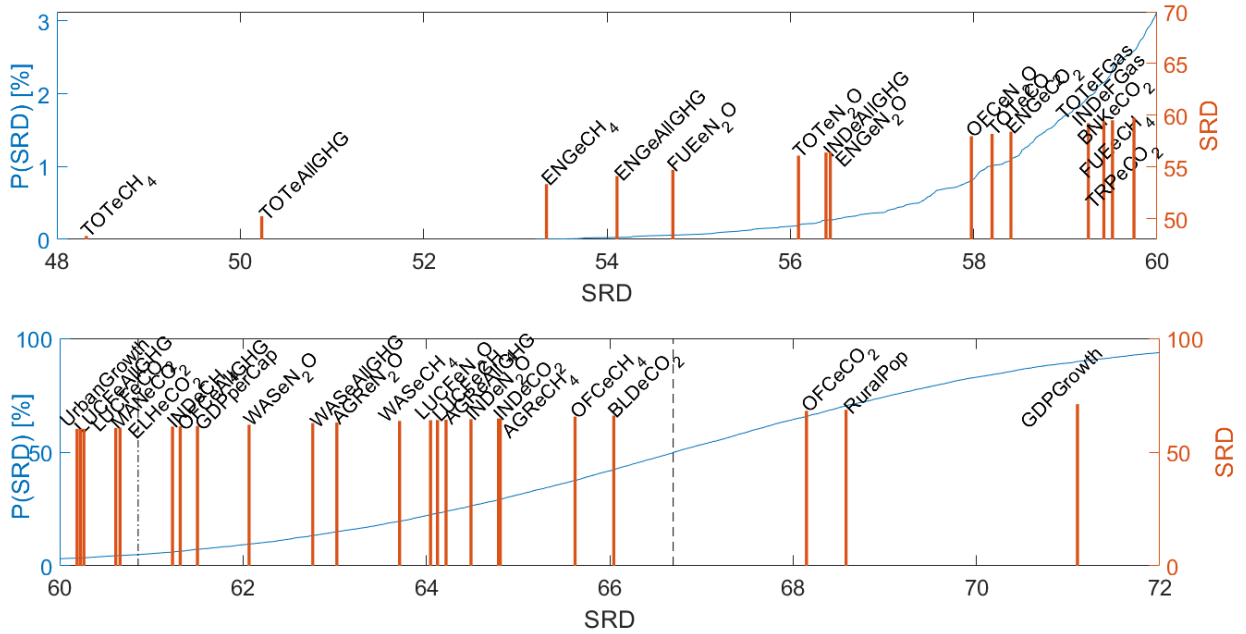


Figure 3.4: **Depiction of SIRD based on percentile transformation with maximum as the aggregate.** This aggregation functions is not as efficient as the other two, as the categories are not as distinctive. Many categories are in the close vicinity of each other; therefore, these are considered similar.

not preferable due to the capped range and fixed transformation of the data, unless the aim is to determine close similarity.

3.1.3 Derringer desirability function for climate change

As utilities imply preference values, I rank based on “Is it preferable?”. I regard it essential to analyse climate change using the generalized SRD algorithm, as various connections and insights may prove helpful to the world as far as mitigating the disasters and helping humanity develop a sustainable industry. Therefore, the defined transformation functions are only sensitive in a determined range. The functions are based on the Derringer desirability function [43], where the values of each criterion are based on a proportion. These proportional classes are determined based on the ABC analysis [58], which follows the 50/30/20 % ratio. Three classes of nations are determined, where the emissions per capita values are insignificant or acceptable (C - 50 %), another should focus more on the reduction (B - 30 %), and in the last, the emissions rates are imperative to reduce (A - 20 %). Therefore, the most preferential class falls within the first 50th percentile (C).

$$u_j(x_{i,j}) = \begin{cases} 1 & \text{if } x_{i,j} > lb \\ 0 & \text{if } x_{i,j} < ub \\ \left(\frac{x_{i,j}-lb}{ub-lb}\right)^s & \text{if } lb \leq x_{i,j} \leq ub \end{cases} \quad (3.1)$$

where lb denotes the emissions per capita of the 50th percentile, before which only one is assigned, while ub represents the limit value denoted by the 80 % mark percentile. The parameter s is determined as one so that the transitional values remain linear, where a value between zero and one is returned. Zero is assigned to all values contained in the 80th and 100th percentile. The cumulative empirical distribution and the Derringer desirability function of the total of the ‘all GHG’ emissions category is presented in Figure 3.5. In terms of the determined Derringer desirability function, the first border (50 %) is designated at five tons per capita, while the second border (80 %) is located at 9.8 tons per capita.

The economic indicators are similarly transformed. The direction of growth indicators are inverted; an additional ‘i’ is added to the name of the categories to demonstrate that these indicators obtained a reversed meaning. The limits are defined in a similar way to the emissions indicators. Data aggregation is performed using additive aggregation, where the weights of emission categories are determined as the ratio of the worldwide emissions of a category to the value of the worldwide total ‘all GHG’ emissions. The weight of the economic indicators is designated as one. The weights of emissions are calculated as in Eq. 3.2:

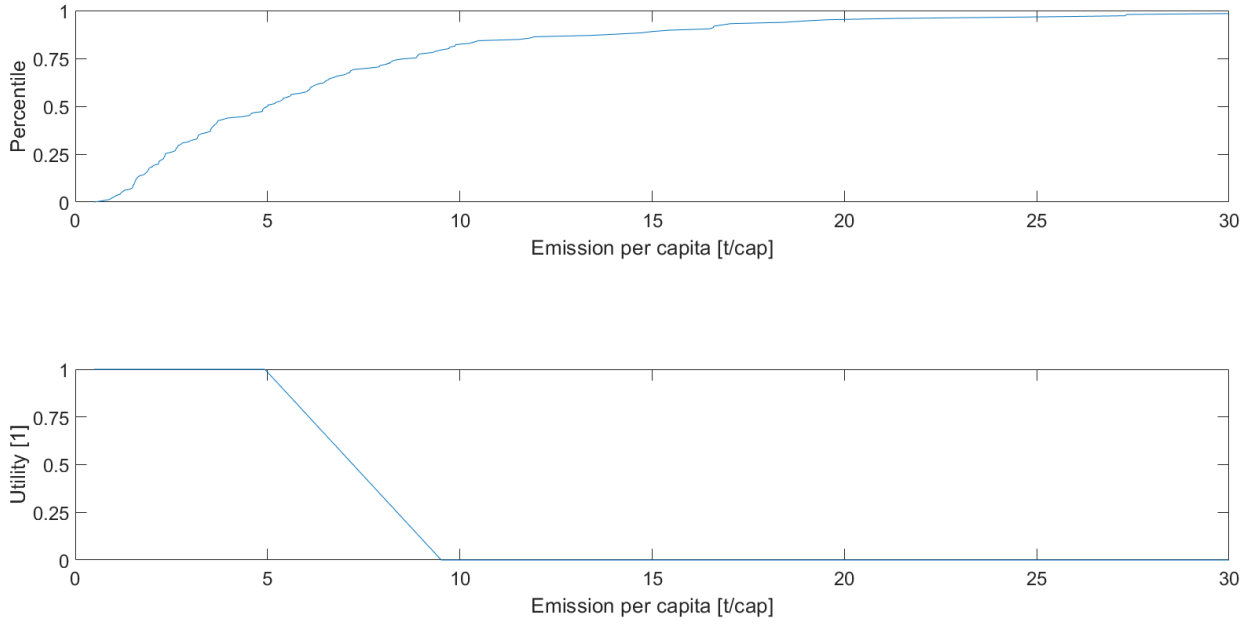


Figure 3.5: **Depiction of the cumulative empirical distribution and Derringer desirability function of the total ‘all GHG’ emissions indicator.** Half of the countries (73) whose emissions per capita are less than about 5 tons per capita are assigned a preferable value of one, while the transition area is determined between 5 and about 9.8 tons per capita. Zero is assigned to every nation that produces more than 9.8 tons per capita.

$$w_k = \frac{x_{world,k}}{x_{world,total}} \quad (3.2)$$

where k is the index of a category.

The SRD-based evaluation of the variables is presented in Figure 3.6, where the variables are transformed by the Derringer desirability function determined for each criterion. Additive aggregation is performed to generate the gold standard.

In the area where the variables are similarly ranked to the gold reference, the starting order strongly resembles that found in Figure 3.2. The total ‘all GHG’ emissions resembles most closely to the sum of desirability due to the weights assigned. Four indicators can be perceived: ‘all GHG’ and CO_2 version of total emissions and energy categories. From the single gas type categories, CO_2 is the most frequent, indicating that I should be the wariest of is CO_2 that is the most abundant gas emitted. Such indicators are electricity, transport and manufacturing, which are grouped. Fluorinated gases are placed in between the manufacturing and industry indicators. The total industry category is in the vicinity of emissions from bunker fuels indicators as shipping the manufactured goods can be

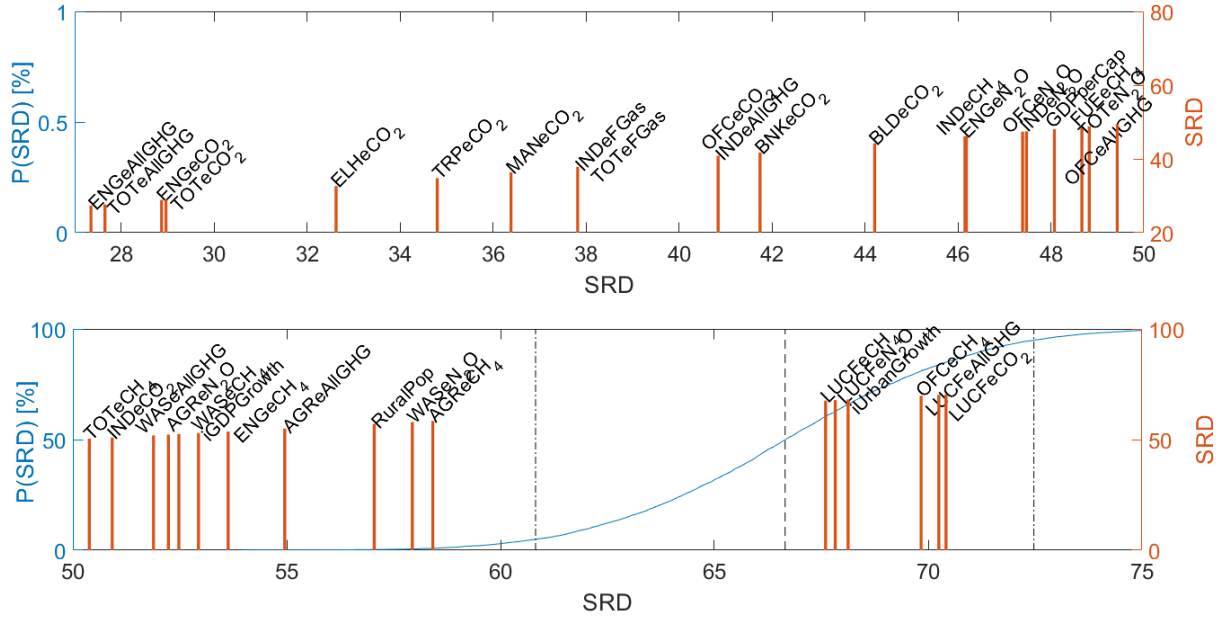


Figure 3.6: **Depiction of the Derringer desirability function-based SRD with additive aggregation as the aggregate.** The variables are distinguishable and operate over a wide range. The starting order of the variables is similar to the classical SRD.

considered as a connection.

The building category may be grouped with the latter indicators, as on-site construction and building maintenance (BLD) requires immense energy that can be provided with combustion from fuels such as bunker (BNK) fuels, petrol, coal or other fuel combustion (OFC). Industry (IND) requires building for operating production lines. Another explanation is that the larger the number of buildings are maintained in urban areas, the more significant a city becomes, leading the establishment of ports and maritime industries such as transport, or *vice versa*, which may be the cause of correlation in these indicators. A real-life example is Singapore, where due to the strategical location, the city flourishes. Even though emissions from burning bunker fuel is not as significant as those from the energy sector, the reduction of pollutants is of utmost importance to the local population as it decreases the chances of developing lung cancer and asthma [59].

The fugitive (FUE) emissions of CH_4 , waste (WAS) CH_4 and ‘all GHG’, CO_2 from industry as well as the GDP per capita indicators are grouped. Fugitive emissions are regarded as a consequence of the inefficient handling of methane [60], similarly, in waste management, the a lack of funding and obsolete technology to treat of organic waste may further damage the environment. Even though landfills are now widespread throughout the world, Sweden burns its solid waste [61] to generate electricity and decrease emissions from inefficient handling. In landfills, CH_4 emissions produced by local microorganisms are rarely captured and become significant greenhouse gas emissions in developing coun-

tries [62], suggesting the inefficient handling of the gas similarly to fugitive emissions. One solution to this problem is the mitigation of CH_4 with controlled microbial oxidation processes [63]. The inverse of GDP growth is located in the vicinity of such indicators due to the fact that in some countries, burning biomass might be cheaper than importing natural gas; therefore, when the growth decreases and the population starts to struggle earning a preferable income, they turn to alternatives.

The total CH_4 indicators and the agricultural N_2O indicators are grouped as some plants may rupture the soil releasing the confined methane. Additionally, due to manure management, methane and nitrous oxide is released during the processing phase [64]. The agricultural all greenhouse gas emissions stands alone between the GDP growth and rural population. A possible explanation is that agriculture in developing countries has been sustained by slash and burn agriculture and burning forests to fertilize the local soil. One specific action is the incineration of oil palms where CO_2 and CH_4 are emitted even after the fire [65]. On the contrary, in developed nations, improper management of fertilizers may increase N_2O gas emissions [66].

N_2O waste emissions, the rural population, and agricultural CH_4 may be connected. In developing countries, a high proportion of the population lives off subsistence agriculture, where the production margin does not resemble that of a specialized company but the amount to satisfy a family. If cities are incapable of sustaining their inhabitants, then the migration to the countryside may cause an increase in agricultural activities. In developing countries, the techniques for grain farming are often similarly inefficient as modern equipment as well as resources are not affordable. Moreover, developing countries may not be able to afford the efficient treatment of human waste and have barely been urbanized. In Asian countries, rice cultivation is a backbone of agriculture, which is responsible for 10 % of its methane emissions [67].

The indicators of slow or decreasing urban growth, LUCF, methane originating from other fuel sources are randomly ranked. Wetlands [56], and tundra [68] release CH_4 and N_2O as well that is somewhat independent of human activities and rather a function of overtime accumulation, thus are placed in the randomly ranked area. Countries with more negligible emissions footprint maybe those that are not developed or significant in terms of emissions per capita enough as urbanization has not destroyed natural habitats. Additionally, if the urban areas do not grow, the wetlands and arable lands may not be removed or built upon.

Although the GDP per capita seems to have attained a rank close to the randomly ranked area, it is essential to note that the GDP per capita of a nation is closely tied to their direct emissions even if this correlation is not linear, *e.g.* the environmental Kuznets curve [57].

3.1.4 Parallel Coordinates Visualization of Greenhouse gas emissions

Regarding the visualization of parallel coordinates, the red axis is considered to be the gold standard or origin axis. The quantiles (25% - yellow, 50% - magenta, 75% - blue, 100% - green) and the first decile (10% - red) of the ranks are colored to distinguish between their positions. The rank numbers for each variable are plotted on the y-axes, and the x-axis depend on the input data. In the case of Figure 3.7, the indicators are provided in alphabetical order for the purpose of visualization. The neighbouring indicators are in an arbitrary ordering without regard to their relationship to each other; therefore, the minimization of the total crossings is not probable. This method yields 76,509 intersections for a ranking of 147 members in terms of CO_2 and ‘all GHGs’ emissions.

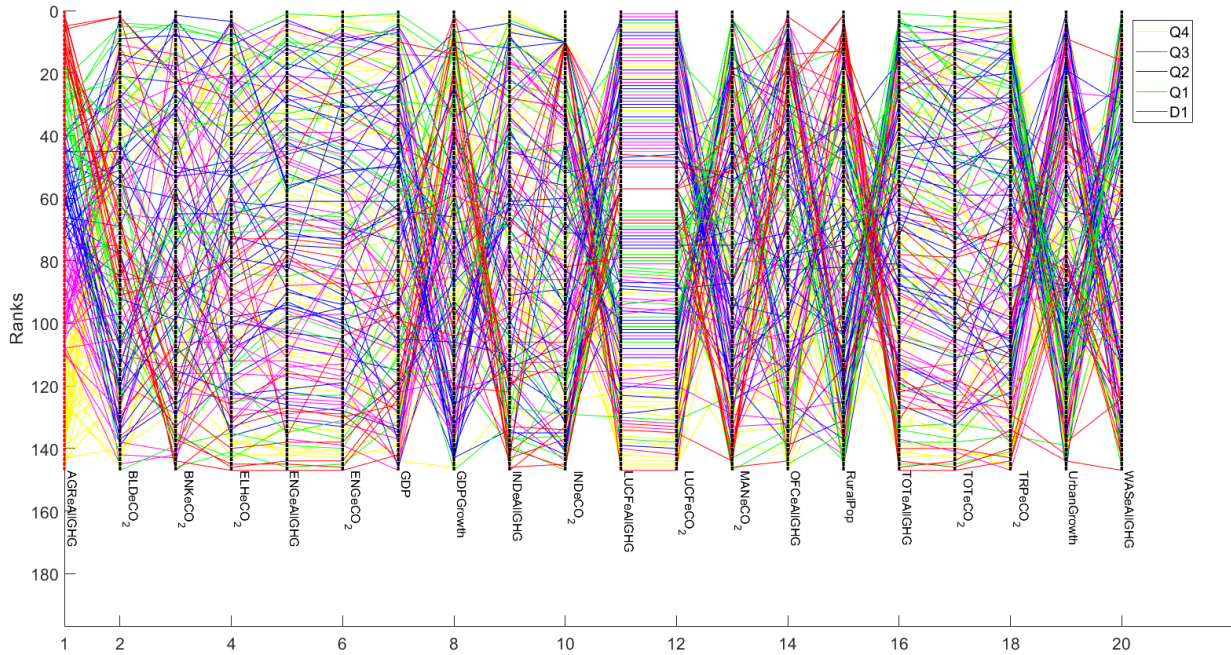


Figure 3.7: Ranked climate indices in alphabetical order in a parallel coordinate system. The legend presents the color of the first decile (D1) and the quantiles (Q1-Q4) for the ranks. The total number of intersections is 76,509.

The MDS-based analysis yielded 53,591 intersections, and both negative and positive dissimilarities. In Figure 3.8, the total of all greenhouse gas emissions categories axis is denoted with red so that the gold standard is quickly identifiable. On the left-hand side, various categories are placed in the vicinity of each other, creating a cluster whose quantiles remain in a similar position, reducing the number of intersections. As the figure proceeds towards the positive limits of the x-axis, the intersections between the categories

urban growth, and rural population are ranked in reverse order to the gold standard.

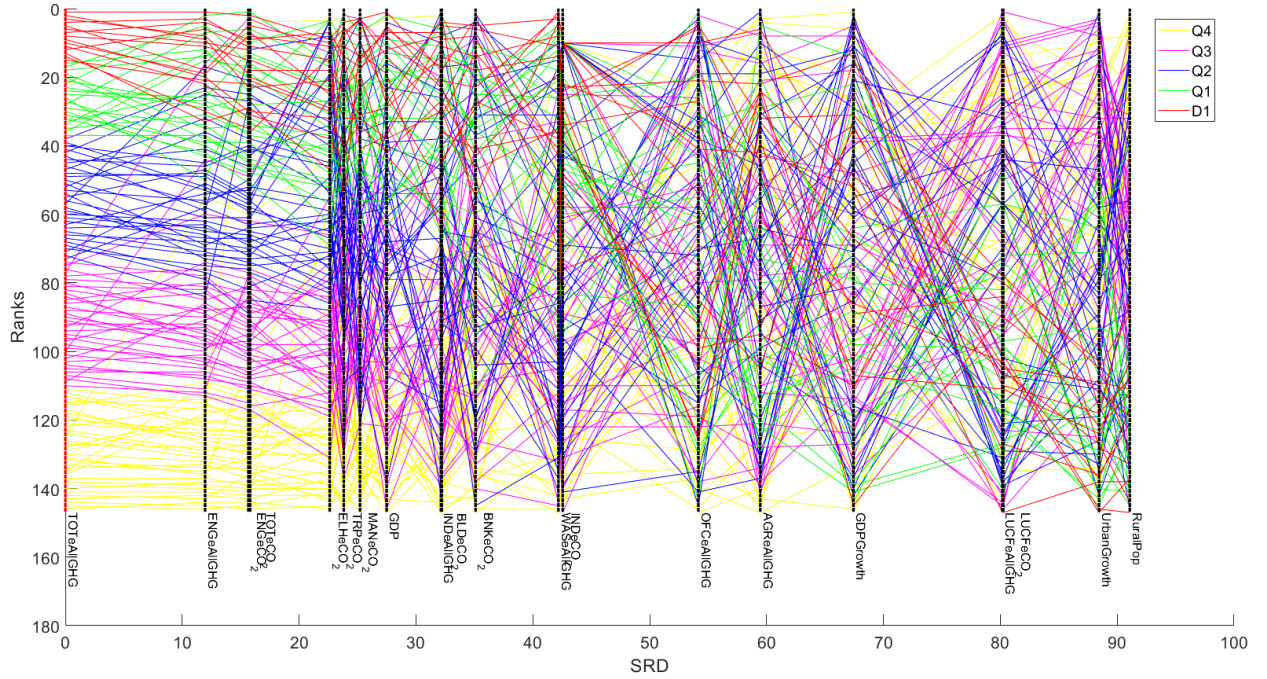


Figure 3.9: Ranked climate indices based on the SRD in a parallel coordinate system. Color codes are explained in Figure 3.7. The total number of intersections is 52,200.

In Figure 3.10, the angle-based ordering of the SRD values is depicted. On the left-hand side, the position of the quantiles is reversed, indicating that the left-hand side of the figure contains reversely ranked categories, whereas, on the right-hand side, the quantiles remain within the same range, with a neglectable degree of fluctuation. It is only reasonable that the randomly and reversely ranked categories are placed in the reverse direction as if one side is filled with similarly correlating variables, the angles between the remaining non-correlating variables may be less to each other than to the other side. The total crossings can be further reduced (to 52200) if the asymmetric distances are plotted to the right and the left from the gold standard, whose SRD value is 0. The total number of intersections is determined as 52,200, 24,309 less than in the alphabet-based approach.

Here, all greenhouse gas emitted by the energy sector, CO_2 emissions of transportation, manufacturing, GDP per capita, industry, bunker fuels and waste are mutually similar, which implicitly suggests the significant contribution of the heavy industry to the total emissions. Because of that, the economic models of nation should be shifted towards circular sustainable economies as circular economies may provide benefits as mitigation, reduced costs of manufacturing, recycling and efficient waste management [69–71]. On the left-hand side, the total CO_2 and CO_2 emitted by energy and electricity are similar variables. However, from the CO_2 emissions of buildings category, the correlation steeply

decreases. The left-hand side consists of indicators that may provide useful in determining either the service or agricultural sector, but not essential to the industry sector. For example, the maintenance cost in terms of environmental footprint of office buildings has been reduced [72]. Passive houses have appeared that are under strict regulation of heat insulation and electricity consumption [73] although the actual manufacturing emissions of the required materials for the buildings may be overlooked [74]. Industry is a requirement for the service sector, and agriculture is boosted with manufactured machines. The GDP growth is reversely ranked along with agriculture, yet it is clear that the two barely correlates. In many developing countries, the agriculture is based on subsistence farming, and a high proportion of the population may only feed themselves and their families and live below the poverty line and cannot contribute to the economy [75]. The figure contains 50,252 intersections, 1,948 less than the classical SRD-based ordering.

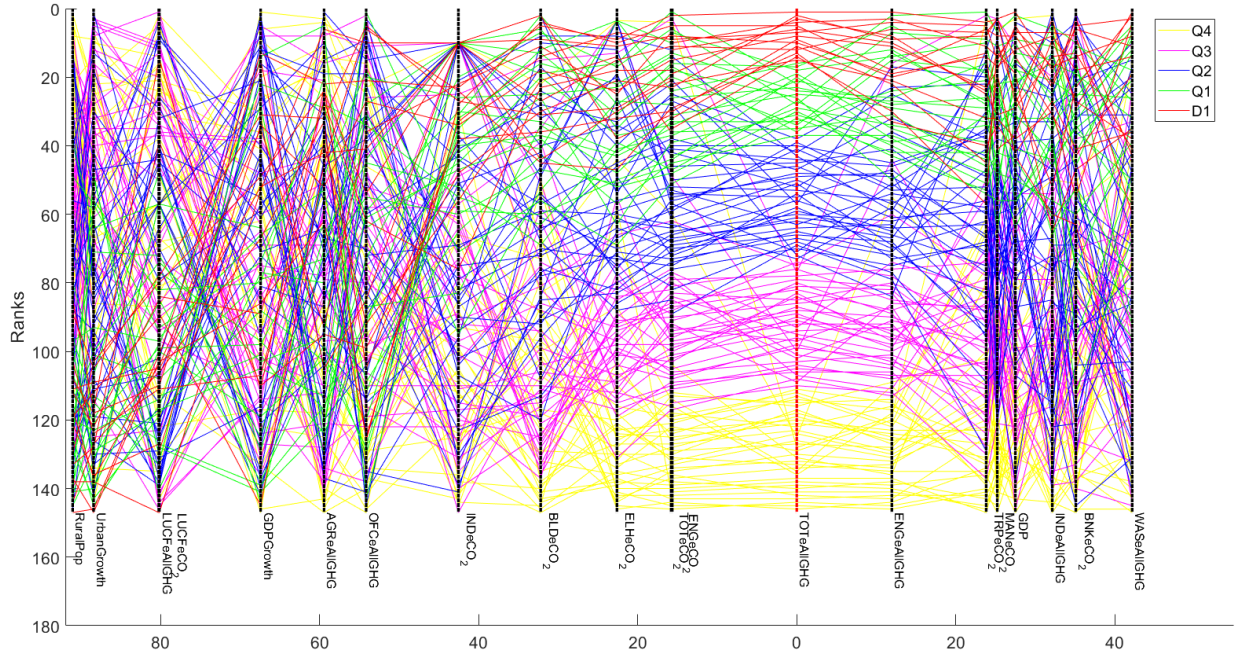


Figure 3.10: Ranked climate indices based on the angles between neighbouring SRD values in a parallel coordinate system. Color codes are explained in Figure 3.7. Number of total crossings are 50252.

Ranking of countries

As the funding towards and awareness of this topic increases, the question arises that which nations should change policies or develop more dynamically for the sake of humanity. The methods include different approaches. Therefore, a ranking for each approach was established in Table 3.2.

Table 3.2: **Ranking of nations based on their total emissions per capita in 2018, the arithmetic mean, maximum percentiles as well as desirability.** The emissions per capita ranking consists of nations with oil and mineral extraction as well as processing capabilities. The mean percentile ranking selects the nations that are responsible for a significant amount of emissions per capita in most categories. The maximum percentile-based ranking is invalid. The last ranking is based on the desirability of a country. The nations that are ranked high are the worst 20 % in terms of emissions and economic indicators performance in the ABC classes.

Country	Emissions per Capita [tCO_2e/cap]	Country	Mean Percentile
Qatar	35.88	Australia	0.83
Bahrain	31.19	Canada	0.82
United Arab Em.	27.33	United States	0.73
Kuwait	27.30	Argentina	0.68
Australia	24.63	Qatar	0.67
Turkmenistan	21.37	New Zealand	0.66
Canada	19.55	Poland	0.66
Saudi Arabia	18.93	United Arab Emi.	0.64
United States	18.43	Kazakhstan	0.64
Oman	17.04	Malaysia	0.64
Country	Max Percentile	Country	Desirability
United Arab Emi.	0.99	Finland	1.73
Uganda	0.99	Malaysia	2.15
Turkmenistan	0.99	Canada	2.27
Solomon Islands	0.99	South Africa	2.60
Singapore	0.99	Bahrain	2.66
Qatar	0.99	Greece	2.75
Oman	0.99	Libya	2.83
New Zealand	0.99	Estonia	2.89
Montenegro	0.99	Czech Republic	2.97
Mongolia	0.99	Australia	3.04

The order of the nations differs for each approach. According to the emissions per capita-based analysis in Figure 3.2, the countries that are ranked high possess an immense amount of natural resources such as oil, iron, bauxite and gold as well as have an industry to extract and process these minerals. Their emissions are enormous per capita if compared to the global average. The first five countries are generally small developing nations, whose economy is based heavily on the oil and petroleum industries, which explains its unusually high emission per capita. The petroleum and oil industries are significant contributors to the economy of a country, providing a more prosperous and better quality of life if performed profitably [76]. It is important to note that a nation is required to have a developed economy for the purpose of funding mitigation technologies to reduce emission per capita [57]. The problem appears when the acquired riches are not invested in a sustainable future are used to momentarily satisfy the citizens or a group of citizens who are only concerned with selfish gains. The other countries possess a significant amount of natural resources, including gold, diamond, iron, coal and bauxite, the mining and processing of which emit an incredible amount of greenhouse gases [77,78] and cause air pollution [79]. The use of petroleum and oil is one of the driving forces of greenhouse gas production, although without which countries can hardly be developed. One of the most crucial categories of present-day economics, namely electricity generation, makes up about 32 % of the global emissions per capita in all greenhouse gases. 17 % of the global per capita is derived from transportation and fuel combustion in vehicles, while agriculture contributes 12 %. Figure 3.2 depicts the relationship between the emissions categories based on the classical approach of SRD and the total ‘all GHG’ emissions indicator which is regarded as the ideal reference, and here, what the ranking is based on.

In Figure 3.3, where the mean percentile is employed as a gold standard, the highest mean percentile proposes that a nation is frequently placed in the most emissions. The top-ten countries are highly developed or have developed intensely over recent decades. Most nations possess a considerable amount of purchasing power and a service-based economy along with an adequate financial sector. Australia and Canada are particularly problematic as they are placed higher with at least 0.1 more in the average percentile. The energy sector of Canada is mainly comprised of oil, and petrol [80], while Australia has an extensive resource extraction-based industry. Profits from these natural resources were invested back into the development of the countries, resulting in highly developed economies, however not enough has been invested in emission-reducing technologies.

If the maximum is employed as an aggregation function, as in Figure 3.4, the ranking is incorrect as the percentiles are calculated based on the relative position of a value with a capped range of one and not on measurable amounts. Countries are selected depending on whether it was placed in the highest percentile in a category. As no weights are assigned,

the highest percentile from energy, 'all GHG' and F-Gases industry may have the same value. These categories are far from equal in terms of environmental degradation, which might be an explanation for why the maximum performed so poorly as a gold standard as well.

In the last ranking, the weighted sum of the desirability values of a nation determines its rank, where the most negligible value of desirability is the 'worst.' The direction of GDP per capita and growth indicators is inverted so that the nations that emit the most per capita and have insignificant GDP per capita as well as slow economic and negative urban growth are listed as of low desirability (refer to Figure 3.5: the less, the better; thus I changed the direction of some indicators). Countries with low GDP per capita, slow or negative growth indicators and a high percentage of the rural population are delegated zero score. Moreover, the worst in the ranking are those, which emit the most in total or energy-related emissions. Two categories can be observed in the rankings: mostly those who produce minerals and petroleum and those where crises occurred in the recent decades. Finland, Australia and Canada are generally developed countries rich in minerals and petrol. Greece have gone bankrupt, South Africa bears unrest due to inequality and racial segregation. In Lybia, two civil wars have been waged. The instability of a nation can be measured by the economic indicators that may modify the greenhouse gas emissions.

Chapter 4

Discussion and Conclusion

4.1 Discussion

I generalized the SRD method for ranking-based multi-criteria decision-making (MCDM) tools. The gold standard can be determined as an aggregated utility value and the main message is to check whether the condition of additive independence is valid. If the aggregation function has already been decided, the analyst may only influence the SRD through the careful selection of the transformation function, similarly to the Derringer desirability-based transformation of the emission categories. If not, I propose new approaches for aggregation, namely multiplicative aggregation or the preference aggregation. Table 4.1 summarizes the properties of the generalized method.

As mentioned above, the ordering of the variables is a major factor that reduces the number of intersections. It is not recommended to visualize the parallel coordinates based on an arbitrary ordering, such as the alphabet, as the number of intersections in this exact example is 76,509. Each ordering method reduced the number of intersections significantly (each by about 25,000), proving that the ordering is a requirement to a good visualization.

Although the angle-based ordering algorithm reduced the number of intersections slightly more than the MDS-based method. However, the hierarchical ordering-based MDS has not been compared to SRD. The importance of the ordering was realized before the employment of the SRD. I compared to a similar rank correlation-based MDS technique, which performs slightly worse. The y-axes of the MDS should be employed with a unit length from each other, as the method provided enormous or minute distances between the axes. It is important to note that even though another analysis may not yield such characteristics. By dropping the fixed-position of the axes, the MDS-based method may reduce the resemblance of the figure and provides information on the relationship between the axes.

The inclusion of the SRD method in the visualization of the parallel coordinates pro-

Table 4.1: The main steps of the generalized method and its properties.

Any transformation can be considered as a utility function and it should be based on the interpretation of the variables
Transformation functions are interpreted as utility functions, so the transformation can be specified for each variable/criterion and can express preferences.
A transformation may change the distribution of the variables/criteria. Monotonous transformations do not influence the ranking.
Aggregation – Generation of the “gold standard”
The aggregation function should be selected based on the intended goal of the analysis.
The aggregated variables are called “gold standard” and can be considered as multi-attribute utility functions.
The transformation of the variables influences the SRD values through the aggregation step.
The transformation and interpretation of the transformed variables determine the applicability of the aggregation functions – the assumptions are employed in utility aggregation: Additive, Utility and Preference-independence.
Criteria analysis – Distances to gold standard
The SRD technique represents the dissimilarities between the criteria/variables and the aggregated utility (gold standard).
The difference in the ranking does not necessarily reflect the utility differences as ranking causes information loss.
The SRD values reflect the consistency of the decision problem that can be influenced by the tuning of the transformation/utility functions and the selection of the aggregation function.

vided a satisfactory reduction in the number of intersections, which has been reduced and compared to multi-dimensional scaling. “Dynamics Visualisation based on Parallel Coordinates” (DYVIPAC) focuses on a set of problems and not on general use. However, the SRD angle method-based visualization alone is not capable of complex biological nor mathematical problems and dynamics analysis. It is rather a technique that focuses on the correlation and relationship between the variables.

The parallel coordinates technique is widely used, *e.g.* in ecology where a complete framework was created to ease the visualization and clarification of water quality as well as invasive species proliferation. The visualization is interactive and the framework provides features such as line coloring, data filtering, highlighting as well as tables. In the SRD-based parallel coordinates visualization toolbox, I only provide a general SRD-based visualization with angle-based ordering that can be further extended in MATLAB or Octave depending on the requirements of the analyst.

4.2 Summary and Conclusion

The case study for the generalized method was the analysis of greenhouse gas emissions indicators. The emissions indicators provide comprehensible information on the emissions under ordinary consequences as the data originates from before the COVID-19 pandemic. The ranking based on desirability proposes that nations change their ways drastically or begin to finance sustainable or renewable energies, as the highest emissions category is based on the energy sector. In the energy sector, most electricity is supplied by fossil fuels, moreover the availability of cheap local coal, oil, petroleum and natural gas may encourage the population as well as companies to use dirt-cheap fossil fuels instead of funding large-scale innovative renewable energy. Moreover, it should be emphasized that struggling nations should be provided as much help as possible so that the financial situation enable them to fund innovative technologies and their application. The connection between the variables is also established. A developing country with high proportion of rural population may not be capable handling the methane and nitrous oxide emissions caused by agriculture and human waste, or the equipment may be too obsolete to do so. Low GDP growth may correlate with ‘all GHGs’ waste and N_2O from agriculture as during a recession people may be required to find jobs in simpler and stable jobs such as farming. The analysis also incorporates the connection between the methane and nitrous oxide from land use, change and forestry and slow urban growth as wetlands and arable land are left untouched. It might be preferable to propose that nations reduce imports of energy and goods as well as attempt to become more sustainable, green and autonomous to reduce the usage of fossil fuels and emissions from shipping. Moreover, various technologies are available or in prototype phase to reduce the emissions per capita. What is important is the funding and the hastened application for commercially viable products, such as *in vitro* meat production.

The visualization of parallel coordinates in itself and as the representation of the sum of ranking differences (SRD) is extended, where the order and distribution of the variables are calculated by the SRD method. The SRD and the parallel coordinates technique were modified by placing the variables not only on the right-hand side of the gold standard but on the left-hand side, as well, thereby increasing the consistency of the visualized relationship system (i.e., showing that variables adjacent to each other are similar). I developed a method for classifying the arrangement according to SRD based on the properties of SRD as a distance norm for triangular inequality (and the cosine theorem). The variables are iteratively placed based on the angle of deviation between the selected side reference variables on left- and right-hand side. The developed method is suitable for parallel coordinate-based visualization with a small number of intersections, as it significantly reduces the quantity of the intersections in the figure.

In this work, I presented that the generalized method has a wide range of applications if the transformation and aggregation functions are selected with an intended goal. The method is capable of exploring the relationship between the variables for multi-criteria decision problems. Moreover, instead of the 1D ordering of the SRD, the parallel visualization technique can be used for more informative visualizations in terms of correlation.

Software and data availability

Generalization of the SRD

Both the MATLAB codes (`preproc_main.m`, `plot_srd.m`, `derringer.m`, `a_ecdf.m`) and the database (`climate_data.xlsx`) is accessible at https://github.com/abonyilab/Preproc_SRD, written by János Abonyi and Ádám Ipkovich, 2021.10.10.

Parallel coordinates visualization

Both the MATLAB/Octave codes described in subsection 2.2.2 and the database (`co2_data.xlsx`) is accessible at <https://github.com/abonyilab/>, written by János Abonyi and Ádám Ipkovich, 2021.10.10.

Acknowledgments

This research was funded by the National Laboratory for Climate Change (NKFIH-471-3/2021). Also, I would like to give my heartiest appreciation to my supervisors, János Abonyi and Károly Héberger who encouraged me to give my best to complete this work and spared incomprehensible amount of time to help my efforts come true.

Bibliography

- [1] Ali Jahan, Md Yusof Ismail, S. Shuib, Dayangku Norfazidah, and K.L. Edwards. An aggregation technique for optimal decision-making in materials selection. *Materials Design*, 32(10):4918–4924, 2011.
- [2] Seyed Hadi Mousavi-Nasab and Alireza Sotoudeh-Anvari. A comprehensive mcdm-based approach using topsis, copras and dea as an auxiliary tool for material selection problems. *Materials Design*, 121:237–253, 2017.
- [3] Ali Alsamhan Syed Hammad Mian Lotfi Hidri Atef M. Ghaleb, Husam Kaid. Assessment and comparison of various mcdm approaches in the selection of manufacturing process. *Advances in Materials Science and Engineering*, 2020(10), 2020.
- [4] Mirko Stojčić, Edmundas Kazimieras Zavadskas, Dragan Pamučar, Željko Stević, and Abbas Mardani. Application of mcdm methods in sustainability engineering: A literature review 2008–2018. *Symmetry*, 11(3), 2019.
- [5] Honghao Zhang, Yong Peng, Guangdong Tian, Danqi Wang, and Pengpeng Xie. Green material selection for sustainability: A hybrid mcdm approach. *PLOS ONE*, 12(5):1–26, 05 2017.
- [6] Abhishek Kumar, Bikash Sah, Arvind R. Singh, Yan Deng, Xiangning He, Praveen Kumar, and R.C. Bansal. A review of multi criteria decision making (mcdm) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews*, 69:596–609, 2017.
- [7] Tao Li, Ang Li, and Yimiao Song. Development and utilization of renewable energy based on carbon emission reduction—evaluation of multiple mcdm methods. *Sustainability*, 13(17), 2021.
- [8] P. Chithambaranathan, Nachiappan Subramanian, Angappa Gunasekaran, and PL.K. Palaniappan. Service supply chain environmental performance evaluation using grey based hybrid mcdm approach. *International Journal of Production Economics*, 166:163–176, 2015.

- [9] Ananna Paul, Nagesh Shukla, Sanjoy Kumar Paul, and Andrea Trianni. Sustainable supply chain management and multi-criteria decision-making methods: A systematic review. *Sustainability*, 13(13), 2021.
- [10] Szymon Fierek and Jacek Zak. Planning of an integrated urban transportation system based on macro – simulation and mcdm/a methods. *Procedia - Social and Behavioral Sciences*, 54:567–579, 2012. Proceedings of EWGT2012 - 15th Meeting of the EURO Working Group on Transportation, September 2012, Paris.
- [11] K. Héberger. Sum of ranking differences compares methods or models fairly. *TrAC - Trends in Analytical Chemistry*, 29(1):101–109, 2010.
- [12] Károly Héberger and Klára Kollár-Hunek. Sum of ranking differences for method discrimination and its validation: Comparison of ranks with random numbers. *Journal of Chemometrics*, 25:151–158, 04 2011.
- [13] Károly Héberger. Method and model comparison by sum of ranking differences in cases of repeated observations (ties). *Chemometrics and Intelligent Laboratory Systems*, 127:139–146, 08 2013.
- [14] Argyro G. Fragkaki, Ieni G. Farmaki, Nikolaos S. Thomaidis, Anna Tsantili-Kakoulidou, Yiannis S. Angelis, Michael A. Koupparis, and Costas G. Georgakopoulos. Comparison of multiple linear regression, partial least squares and artificial neural networks for prediction of gas chromatographic relative retention times of trimethylsilylated anabolic androgenic steroids. *Journal of Chromatography A*, 1256:232–239, 2012.
- [15] V. Fernão Pires, A. Vieira Pombo, and João Miguel Lourenço. Multi-objective optimization with post-pareto optimality analysis for the integration of storage systems with reactive-power compensation in distribution networks. *Journal of Energy Storage*, 24:100769, 2019.
- [16] João Miguel Lourenço and Luiz Lebensztajn. Post-pareto optimality analysis with sum of ranking differences. *IEEE Transactions on Magnetics*, 54(8), 2018.
- [17] Attila Gere, Anita Rácz, Dávid Bajusz, and Károly Héberger. Multicriteria decision making for evergreen problems in food science by sum of ranking differences. *Food Chemistry*, 344:128617, 2021.
- [18] Robert A. van den Berg, Hub CJ Hoefsloot, Johan A. Westerhuis, Age K. Smilde, and Mariët J. van der Werf. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7:142, 2006.

- [19] Liping Yu, Yuntao Pan, and Yishan Wu. Research on data normalization methods in multi-attribute evaluation. In *2009 International Conference on Computational Intelligence and Software Engineering*, pages 1–5, 2009.
- [20] Balázs R. Sziklai, Máté Baranyi, and Károly Héberger. Testing cross-validation variants in ranking environments, 2021.
- [21] Edward J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.
- [22] Alfred Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.
- [23] Maurice Kendall. Rank correlation methods. 1970.
- [24] Zhiguang Zhou, Zhifei Ye, Jiajun Yu, and Weifeng Chen. Cluster-aware arrangement of the parallel coordinate plots. *Journal of Visual Languages Computing*, 46:43–52, 2018.
- [25] George A. F. (Ed.). Seber. Multivariate observations. *Wiley Series in Probability and Statistics*, pages 139–146, 1984.
- [26] Geoffrey Ellis and Alan Dix. Enabling automatic clutter reduction in parallel coordinate plots. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):717–724, 2006.
- [27] Lan Nguyen, Andrea Degasperi, Philip Cotter, and Boris Kholodenko. Dyvipac: An integrated analysis and visualisation framework to probe multi-dimensional biological networks. *Scientific reports*, 5:12569, 07 2015.
- [28] Omar Alminagorta, Charlie J.G. Loewen, Derrick T. de Kerckhove, Donald A. Jackson, and Cindy Chu. Exploratory analysis of multivariate data: Applications of parallel coordinates in ecology. *Ecological Informatics*, 64:101361, 2021.
- [29] Robert F. Dell and Mark H. Karwan. An interactive mcdm weight space reduction method utilizing a tchebycheff utility function. *Naval Research Logistics (NRL)*, 37(2):263–277, 1990.
- [30] Alexandre Bevilacqua Leoneti. Utility function for modeling group multicriteria decision making problems as games. *Operations Research Perspectives*, 3:21–26, 2016.
- [31] *Climate Watch*. Ghg emissions, 2020.

- [32] *Food and Agriculture Organization*. Food and agriculture statistics (*FAOSTAT*) emissions database, 2020.
- [33] *Organisation for Economic Co-operation and Development (OECD)/International Energy Agency (IEA)*. Co₂ emissions from fuel combustion, 2019.
- [34] *World Bank*. Gdp, population, urban population growth, rural population, gdp per capita growth indicators, 07 2021.
- [35] Goutam Chakraborty and Basabi Chakraborty. A novel normalization technique for unsupervised learning in ann. *IEEE Transactions on Neural Networks*, 11(1):253–257, 2000.
- [36] Aniko Szabo, Kenneth M. Boucher, William L. Carroll, Lev B. Klebanov, Alexander D. Tsodikov, and Andrei Yu Yakovlev. Variable selection and pattern recognition with gene expression data generated by the microarray technology. *Mathematical Biosciences*, 176(1):71–98, 2002.
- [37] Li-Chu Chien. A rank-based normalization method with the fully adjusted full-stage procedure in genetic association studies. *PLOS ONE*, 15(6):1–16, 06 2020.
- [38] George E. P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.
- [39] In-Kwon Yeo and Richard A. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 12 2000.
- [40] Changyong Feng, Hongyue Wang, Naiji Lu, and Xin M. Tu. Log transformation: application and interpretation in biomedical research. *Statistics in Medicine*, 32(2):230–239, 2013.
- [41] J. Tarrío-Saavedra, Jorge López Beceiro, Salvador Naya, Mario Francisco-Fernández, and Ramón Artiaga. Simulation study for generalized logistic function in thermal data modeling. *Journal of Thermal Analysis and Calorimetry*, 118:1253–1268, 11 2014.
- [42] E.C. Harrington. The desirability function. *Industrial Quality Control*, 21:494–498, 1965.
- [43] George Derringer and Ronald Suich. Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12(4):214–219, 1980.

- [44] Peter C. Fishburn and Ralph L. Keeney. Seven independence concepts and continuous multiattribute utility functions. *Journal of Mathematical Psychology*, 11(3):294–327, 1974.
- [45] Ralph L. Keeney. Multiplicative utility functions. *Operations Research*, 22(1):22–34, 1974.
- [46] Ralph L. Keeney and Howard Raiffa. *Decisions with Multiple Objectives*. Cambridge University Press, 1993.
- [47] József Dombi and Orsolya Csiszár. Explainable neural networks based on fuzzy logic and multi-criteria decision tools. *Springer Nature*, 408:14–18, 2021.
- [48] Harold D. Griffin. Graphic computation of tau as a coefficient of disarray. *Journal of the American Statistical Association*, 53(282):441–447, 1958.
- [49] Sandro Federici, Francesco N. Tubiello, Mirella Salvatore, Heather Jacobs, and Josef Schmidhuber. New estimates of co₂ forest emissions and removals: 1990–2015. *Forest Ecology and Management*, 352:89–98, 2015. Changes in Global Forest Resources from 1990 to 2015.
- [50] Song-Li Zhu. Comparison and analysis of co₂ emissions data for china. *Advances in Climate Change Research*, 5(1):17–27, 2014.
- [51] Wanqiu Hu, Jinping Tian, and Lujun Chen. Greenhouse gas emission by centralized wastewater treatment plants in chinese industrial parks: Inventory and mitigation measures. *Journal of Cleaner Production*, 225:883–897, 2019.
- [52] Thomas Hendrickson, Mi Nguyen, Marsha Sukardi, Alexandre Miot, Arpad Horvath, and Kara Nelson. Life-cycle energy use and greenhouse gas emissions of a building-scale wastewater treatment and nonpotable reuse system. *Environmental Science & Technology*, 49, 07 2015.
- [53] Jan Broucek. Production of methane emissions from ruminant husbandry: A review. *Journal of Environmental Protection*, 5:1482–1493, 01 2014.
- [54] Hanna Tuomisto and Joost Mattos. Environmental impacts of cultured meat production. *Environmental Science & Technology*, 45:6117–6123, 06 2011.
- [55] Ted C. Schroeder, Andrew P. Barkley, and Kathi C. Schroeder. Income growth and international meat consumption. *Journal of International Food & Agribusiness Marketing*, 7(3):15–30, 1996.

- [56] Marie-Claude Bonneville, Ian B. Strachan, Elyn R. Humphreys, and Nigel T. Roulet. Net ecosystem CO_2 exchange in a temperate cattail marsh in relation to biophysical properties. *Agricultural and Forest Meteorology*, 148(1):69–81, 2008.
- [57] Wei Ming Huang, Grace W.M. Lee, and Chih Cheng Wu. Ghg emissions, gdp growth and the kyoto protocol: A revisit of environmental kuznets curve hypothesis. *Energy Policy*, 36(1):239–247, 2008.
- [58] Benito E. Flores and D. Clay Whybark. Implementing multiple criteria abc analysis. *Journal of Operations Management*, 7(1-2):79–85, 1987.
- [59] Mikhail Sofiev, James Winebrake, Lasse Johansson, Edward Carr, Marje Prank, Joana Soares, Julius Vira, R. Kouznetsov, Jukka-Pekka Jalkanen, and James Corbett. Cleaner fuels for ships provide public health benefits with climate tradeoffs. *Nature Communications*, 9, 02 2018.
- [60] Derek R. Johnson, April N. Covington, and Nigel N. Clark. Methane emissions from leak and loss audits of natural gas compressor stations and storage facilities. *Environmental Science & Technology*, 49(13):8132–8138, 2015.
- [61] Ola Eriksson. *Environmental and Economic Assessment of Swedish Municipal Solid Waste Management in a Systems Perspective*. PhD thesis, Royal Institute of Technology, Stockholm (Sweden). Department of Chemical Engineering and Technology, 06 2002.
- [62] Uisung Lee, Jeongwoo Han, and Michael Wang. Evaluation of landfill gas emissions from municipal solid waste landfills for the life-cycle analysis of waste-to-energy pathways. *Journal of Cleaner Production*, 166:335–342, 2017.
- [63] Charlotte Scheutz, Peter Kjeldsen, Jean E. Bogner, Alex De Visscher, Julia Gebert, Helene A. Hilger, Marion Huber-Humer, and Kurt Spokas. Microbial methane oxidation processes and technologies for mitigation of landfill gas emissions. *Waste Management & Research*, 27(5):409–455, 2009.
- [64] Dave Chadwick, Sven Sommer, Rachel Thorman, David Fanguero, Laura Cardenas, Barbara Amon, and Tom Misselbrook. Manure management: Implications for greenhouse gas emissions. *Animal Feed Science and Technology*, 166-167:514–531, 2011. Special Issue: Greenhouse Gases in Animal Agriculture - Finding a Balance between Food and Emissions.
- [65] Selvakumar Dhandapani and Stephanie Evers. Oil palm ‘slash-and-burn’ practice increases post-fire greenhouse gas emissions and nutrient concentrations in burnt

- regions of an agricultural tropical peatland. *Science of The Total Environment*, 742:140648, 2020.
- [66] Clifford S. Snyder, Thomas W. Bruulsema, Tom L. Jensen, and Paul E. Fixen. Review of greenhouse gas emissions from crop production systems and fertilizer management effects. *Agriculture, Ecosystems Environment*, 133(3):247–266, 2009. Reactive nitrogen in agroecosystems: Integration with greenhouse gas interactions.
- [67] Tek Narayan Maraseni, Ravinesh C. Deo, Jiansheng Qu, Popular Gentle, and Prem Raj Neupane. An international comparison of rice consumption behaviours and greenhouse gas emissions from rice production. *Journal of Cleaner Production*, 172:2288–2300, 2018.
- [68] Carolina Voigt, Maija Marushchak, Benjamin Abbott, Christina Biasi, Bo Elberling, Steven Siciliano, Oliver Sonnentag, Katherine Stewart, Yuanhe Yang, and Pertti Martikainen. Nitrous oxide emissions from permafrost-affected soils. *Nature Reviews Earth Environment*, 1:1–15, 07 2020.
- [69] Alejandro Gallego-Schmid, Han-Mei Chen, Maria Sharmina, and Joan Manuel F. Mendoza. Links between circular economy and climate change mitigation in the built environment. *Journal of Cleaner Production*, 260:121115, 2020.
- [70] Gemma Durán-Romero, Ana M. López, Tatiana Beliaeva, Marcos Ferasso, Christophe Garonne, and Paul Jones. Bridging the gap between circular economy and climate change mitigation policies through eco-innovations and quintuple helix model. *Technological Forecasting and Social Change*, 160:120246, 2020.
- [71] Carine Lausset, Francesco Cherubini, Gabriel David Oreggioni, Gonzalo del Alamo Serrano, Michael Becidan, Xiangping Hu, Per Kr. Rørstad, and Anders Hammer Strømman. Norwegian waste-to-energy: Climate change, circular economy and carbon capture and storage. *Resources, Conservation and Recycling*, 126:50–61, 2017.
- [72] Miimu Airaksinen and Pellervo Matilainen. A carbon footprint of an office building. *Energies*, 4, 12 2011.
- [73] Mirela Mihai, Vladimir Tanasiev, Cristian Dinca, Adrian Badea, and Ruxandra Vidu. Passive house analysis in terms of energy performance. *Energy and Buildings*, 144:74–86, 2017.
- [74] André Stephan, Robert H. Crawford, and Kristel de Myttenaere. A comprehensive assessment of the life cycle energy demand of passive houses. *Applied Energy*, 112:23–34, 2013.

- [75] Qange Siphesihle and Mdoda Lelethu. Factors affecting subsistence farming in rural areas of nyandeni local municipality in the Eastern Cape Province. *South African Journal of Agricultural Extension*, 48:92 – 105, 00 2020.
- [76] Barbara Doric and Vlado Dimovski. Managing petroleum sector performance – a sustainable administrative design. *Economic Research-Ekonomska Istraživanja*, 31(1):119–138, 2018.
- [77] Michael Hendryx, Mohammad Saidul Islam, Guang-Hui Dong, and Gunther Paul. Air pollution emissions 2008–2018 from australian coal mining: Implications for public and occupational health. *International Journal of Environmental Research and Public Health*, 17(5), 2020.
- [78] Sarah Strode, Lyatt Jaeglé, and Noelle E. Selin. Impact of mercury emissions from historic gold and silver mining: Global modeling. *Atmospheric Environment*, 43(12):2012–2017, 2009.
- [79] Aditya Kumar Patra, Sneha Gautam, and Prashant Kumar. Emissions and human health impact of particulate matter from surface mining operation—a review. *Environmental Technology & Innovation*, 5:233–249, 2016.
- [80] Edward Olale, Thomas O. Ochuodho, Van Lantz, and Jafar El Armali. The environmental kuznets curve model for greenhouse gas emissions in canada. *Journal of Cleaner Production*, 184:859–868, 2018.