Developing an airport sustainability evaluation index through composite indicator approach

Reference:
Jia Xibei, Buyle Sven, Macário Rosário.- Developing an airport sustainability evaluation index through composite indicator approach
Journal of air transport management - ISSN 1873-2089 - 113(2023), 102469
Full text (Publisher's DOI): https://doi.org/10.1016/J.JAIRTRAMAN.2023.102469
To cite this reference: https://hdl.handle.net/10067/1994450151162165141

Developing an airport sustainability evaluation index through composite indicator approach

Abstract

This paper proposes a holistic definition of airport sustainability comprising social, economic, operational, and environmental dimensions. Methodologically, a composite indicator approach is applied to build the Airport Sustainability Evaluation Index (ASEI), which aims to benchmark airports' sustainability performance across the four dimensions. To justify the issue of subjectivity in composite indicator building, two different methods are used in each of the normalization, weighting, and aggregation processes. Consequently, this forms eight composite indicator building schemes. Then, a variance-based sensitivity analysis, average shift in ranking (ASR), and Cronbach's alpha test are performed to examine the sensitivity and reliability among the eight schemes. Schiphol airport is selected as a demonstration to validate the ASEI with its data from 2012 to 2021 as inputs. The results reveal a significant consensus among the eight schemes in identifying the outstanding and bottom performers across the analyzed period. Additionally, weighting is found to be the most influential composite indicator building process. Further, the scheme with the most significant contribution to the result reliability found in this paper is re-scaling as the normalization method, Benefit-of-the-doubt (BoD) as the weighting method, and Non-compensatory multi-criteria approach (NCMC) as the aggregation method.

Keywords: Airport sustainability, Sustainability evaluation, Composite indicator, Sensitivity analysis.

1. Introduction

Airports are vital energy- and capital-intensive investments facilitating intermodal transmission for both people and goods. Therefore, airports are endowed with dual obligations to maximize socio-economic initiatives while minimizing negative impacts on the environment. From this point of view, airport sustainability is an essential topic in the world transportation network which involves striking a balance between these obligations. Furthermore, due to the strong dependence of the air transport industry on scarce resources, even if current equipment and techniques were fully optimized, this would not make European air transport sustainable in the longer term without significant changes in technology or supporting framework.

Prior to the outbreak of Covid-19, the aviation and aerospace industry in the EU employed an estimated 408,000 people directly in 2019, representing 0.2% of total employment (Eurostat, 2020). According to European Commission (2021, p.1), the aviation and aerospace industry contributed to the EU's GDP by an estimated 2.1% in 2017. On the environmental front, CO2 emissions from commercial aviation in the EU increased 30 percent between 2013 and 2019 to 151.8 million metric tons. Of this total, 81.6 million metric tons were attributable to international flights (Graver, Rutherford and Zheng, 2020). The aviation industry is a leader in using advanced technology and is vulnerable to environmental impacts. For this reason, the aviation industry is closely monitored by authorities and green organizations.

At the policy level, aviation authorities and related bodies worldwide are launching actions and guidance documents to promote sustainable airport development. The Federal Aviation Administration (FAA, 2022) has launched the Sustainability Master Plan and Airport Sustainability Plans, in which 44 airports in the US are involved in a way that will reduce environmental impacts, achieve economic benefits, and improve community relations. In Australia, Sydney Kingsford Smith Airport (SYD, 2020), Adelaide Airport (2019), and Melbourne Airport (2019) are developing their sustainability strategies, recognizing that the success of the airport itself can be boosted by trading in an environmentally, socially and economically responsible manner. The Copenhagen Airport is taking a forerunner role in terms of airport sustainability development (AECOM, 2020, p.7). Concerning the environment, Copenhagen Airport's Climate Strategy has set a goal of eliminating all carbon emissions from the airport, including surface access by 2050 (Copenhagen Airports, 2019). The ACI EUROPE (2022) launched a Sustainability Strategy for Airports in 2019, acting as a systematic approach to airport sustainability and practical guidance on achieving it. In line with its Sustainable and Smart Mobility Strategy, the European Commission (2020) reiterates the urgency of the transition to zero-emission airports, stating that "*the best practices followed by the most sustainable airports must become the new normal and enable more sustainable forms of connectivity*".

Therefore, sustainability evaluations have the ambitious mandate of identifying, measuring, and evaluating the potential impacts of alternatives for sustainability (Devuyst, 2000, pp.67-78). However, sustainability evaluation is still not a mature framework like Environmental Impact Assessment (EIA). The methods commonly used for sustainability evaluation fall into three main categories: monetary tools, biophysical models, and sustainability indicators/composite indexes (Gasparatos, El-Haram, and Horner, 2008). This study chooses the composite indicator approach as the principal methodology for its enhanced accuracy in assessing different sustainability issues, as they do not need to be transformed into other metrics, such as monetary. Further, it allows for a comprehensive assessment that simultaneously covers all aspects required for the sustainability of an airport.

The overall objective of this paper is to develop a holistic and transferable approach to evaluate airport sustainability across multi-dimensional indicators. Amsterdam Airport Schiphol (hereafter referred to as Schiphol) will be taken as a demonstration to validate this approach. In the pre-modeling phase, sustainability indicators will be screened from the literature and then collected separately from economic, environmental, social, and operational dimensions via the Schiphol Group's annual reports and other disclosed information. The different combinations of normalization, weighting, and aggregation methods will be used to construct the Airport Sustainability Evaluation Index (ASEI). In the end, by applying sensitivity and reliability analysis, how changes in the final composite indicator scores are linked qualitatively or quantitatively to various combination schemes will be identified. Theoretically, the ASEI proposed in this research can enrich the evaluation content and methods on airport sustainability and promote extended research. Practically, the index can facilitate airport managers to intuitively recognize the sustainability performance of the airport over a period of time and provide theoretical evidence for subsequent policy decisions. The paper is organized as follows: Section 2 presents a comprehensive definition of airport sustainability and outlines an overview on existing airport sustainability assessment studies. Section 3 depicts the methodology for constructing the ASEI and performing the sensitivity and reliability analyses. Section 4 elaborates on the analytical findings from Schiphol and further highlights the approach's transferability and policy implications. The final conclusions are drawn in Section 5.

2. Literature review

2.1. Sustainability, sustainable development and airport sustainability

It is crucial to distinguish between the concepts of sustainable development and sustainability. As stated by Lubk (2017, p. 14), the former addresses a process, whereas the latter describes a state or condition at the end of the process. Consequently, to achieve sustainability, it is essential to ensure development to be sustainable, making sustainability a long-term vision and sustainable development a means of achieving it.

The notion of sustainable development was shaped in 1987 from the World Commission on Environment and Development (WCED), which states that "*sustainable development is the development that meets the needs of the present without compromising the ability of future generations to meet their own needs*" (Brundtland and Khalid, 1987, p.37). Over the following decade, the WCED definition was widely accepted and paraphrased for use in alternative domains. In the transport sector, for example, Black (1996) defined sustainable transport as "*satisfying current transport and mobility needs without compromising the ability of future generations to meet these needs*". In 2000, the Millennium Development Goals (MDGs), proposed at the Millennium Summit, brought further clarity to the connotations of sustainable development. The MDGs present a historic and effective approach to global mobilization for sustainable development, by transforming it into an easily understandable set of eight goals for the year 2015 and emphasizing three areas: human capital, infrastructure, and human rights, with the aim of improving living standards (Sachs, 2012; United Nations, 2015a). The second major theoretical evolution on sustainable development occurred in 2015 with the launch of the 2030 Agenda (United Nations, 2015b). Within the Agenda, 17 Sustainable Development Goals (SDGs) are introduced, which place inequality at the heart of the Agenda through strong inclusivity for the poorest and most vulnerable, together with the guiding principle that global development should "*leave no-one behind*". The 169 targets and 230 indicators that come packaged with the SDGs also provide the basis for quantifying sustainability in terms of monitoring, evaluation, and impact accountability.

Very often, sustainability is represented as integrating three pillars – economic development, environmental protection, and social progress (Banister et al, 2000, p.119; Upham, 2001a). The three pillars are interdependent and thus equally important (Lubk, 2017, p.32). In Europe, the European Commission (2002) recognizes the multi-dimensionality in the definition of sustainability, claiming that the social, environmental, and economic dimensions must be dealt with together. However, the prominence in research and practice is unbalanced, with a stronger focus on environmental sustainability than economic and social sustainability (Walker and Cook, 2009).

Sustainability has become a prominent topic in the air transport field as the sector is developing at an accelerated pace. As of now, the definition of airport sustainability is contested, and no unified definition has been developed (Graham and Halpern, 2018, p. 299; Wan et al, 2020; Janic, 2007, p.3; Upham, 2001a; Knudsen, 2002, p.11). However, some organizations have their own definitions and interpretations for different purposes.

The FAA (2022) defined airport sustainability as "*reducing environmental impacts, helping maintain high, stable levels of economic growth, and assisting social progress. Also, it is a broad set of actions that ensure organizational goals are achieved in a way consistent with the needs and values of the local community*." The Airports Council International-North America (ACI-NA, 2010) stated airport sustainability as a holistic airport management approach to ensure the integrity of economic viability, operational efficiency, natural resource conservation, and social responsibility (EONS). Alongside the classical Triple Bottom Line model (Elkington, 1999), which encompasses economic, social and environmental aspects, the operational aspect of airport sustainability is also being progressively recognized (Gu, 2019). Many US airports have adopted EONS as their sustainability model or applied some adaptations based on it (Martin-Nagle and Klauber, 2015). Such adaptations can be explained by the unique characteristics and priorities of individual airports. Factors such as an airport's location, size, traffic volume, ownership structure, and regional regulations can influence how sustainability is defined and implemented. Also, an airport's stakeholders, including airlines, passengers, local communities, and government entities, may have varying expectations and preferences regarding sustainability (Amaeshi and Crane, 2006).

Up till now, although the definitions of airport sustainability proposed by various institutions have nuances and in specific contexts rather different interpretations of the priority and scope, it is worth noting that all endeavors in this field are pointing toward the same direction. A pragmatic approach to sustainable development can strike a careful balance between the aspects discussed above while maintaining a focus on the mission of the airport (Flouris and Yılmaz, 2011, p.185). Additionally, a sound and successful concept of airport sustainability must include minimum requirements and demonstrate possible ways to achieve and maintain them, giving policymakers and all participating bodies the freedom to set their priorities in order to achieve sustainability.

In conclusion, this paper defines airport sustainability in two facets: on the one hand, we recognize the uniformity of airport sustainability as a holistic vision embracing the promotion of social responsibility, operational efficiency, economic feasibility, and environmental friendliness; on the other hand, we endorse the specificity of airport sustainability derives from the external impacts and internal dynamics faced by airports. Fig. 1 gives a further explanation of the definition. This definition provides a broad canvas for the sustainability of the airport, and it also leaves space for the relevant stakeholders to tailor their priorities based on the characteristics of the airport.

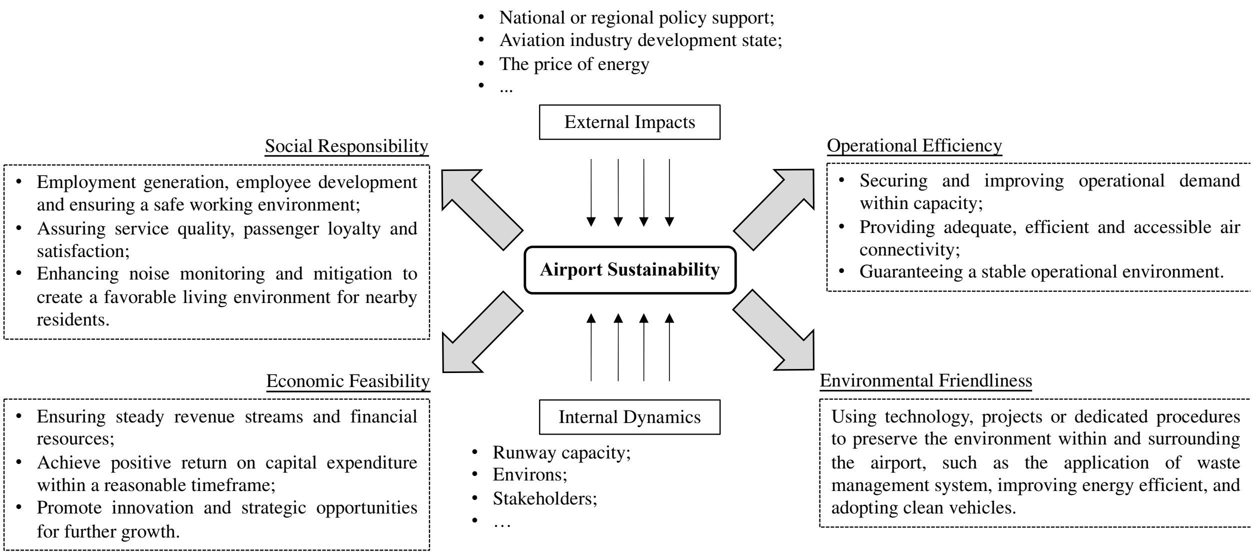

Fig. 1. Airport sustainability (source: own development)

2.2. *Review of airport sustainability evaluation studies*

Some of the previous studies on airport sustainability have focused on one particular aspect to tackle the airport's practical problems, such as airport energy management (Baxter, Srisaeng and Wild, 2018; Uysal and Sogut, 2017), aviation biofuel (Baxter, Srisaeng and Wild, 2020), airport waste management (Santos et al, 2020), and airport water management (Somerville et al, 2015; Baxter, Srisaeng and Wild, 2019). While these studies are tailored from a practical point of view and relatively easy to carry out, it lacks a holistic perspective.

In this paper, we will mainly focus on the integral sustainability assessment of the airports, and an overview of such studies is summarized in Table 1. According to Table 1, Upham (2001b) was the first pioneer who commenced the use of indicators to quantify airport sustainability. In his publication entitled "*Selecting indicators for a decision support tool for airport sustainability"*, he proposes an airport sustainability model comprising nine indicators (number of surface access vehicles, aircraft movements, static power consumption, gaseous pollutant emissions, aircraft noise emissions, terminal passengers, surface access passengers, water consumption and solid waste). In 2015, he adapted land take and biodiversity to the framework (Upham & Mills, 2005). Under Upham's initial theory, airport sustainability incorporates two dimensions: environmental and operational. In 2010, Janic (2010) established the theoretical basis for quantifying airport sustainability via the identification of "*effects-benefits*" and "*impact-externalities*" associated with airport activities. Meanwhile, he extended the dimensions of airport sustainability into four.

However, up until 2010, there was no complete case study to validate the indicators and frameworks developed for airport sustainability. Adler et al. (2013) benchmarked the performance of 85 European regional airports through DEA. Despite being titled as sustainability performance benchmarking by the authors, the selected input and output indicators are by no means different from those formerly utilized in DEA for airport operational efficiency assessment. The airport sustainability evaluation has de facto only been supported by practical examples since 2016. From 2016 to date, a total of 16 journal articles have proposed indicators and/or framework for airport sustainability evaluation, 14 of which have used one or more airport cases to validate them.

The primary approach to the selection and establishment of sustainability indicators to date remains through reviewing industrial standards and relevant research papers. Participatory approaches were used by Lu et al. (2018) and Orkomy and Sharbatdar (2021), who involved experts' interviews and questionnaires during the identification and weighting of sustainability indicators. Additionally, Fuzzy Delphi method and SWARA have also been used to retrieve experts' opinions (Chao et al., 2017; Kaya & Erginel, 2020).

In terms of evaluation methodology, the existing approaches mainly fall into two categories. The first consists of traditional decision-making techniques, including CBA, MCDM, DEA and its extended forms. The second comprises sustainability rating systems, during which the sustainability index will be developed. In spite of the fact that those

Table 1. An overview of airport sustainability evaluation studies.

Citation	Number of indicators	Dimensions of indicators * EC	OP	EN	SO	Participatory method is used in indicators selection	Indicators are validated by case studies (Evaluation method) (Airport samples)
Upham (2001b)	9		√	√		No	No
Upham and Mills (2005)	10		√	√		No	No
Janic (2010)	12	√	√	√	√	No	No
Adler et al. (2013)	7	√	√			No	Yes (Data Envelopment Analysis (DEA)) (85 European regional airports)
Ferrulli (2016)	22			√		No	No
Kılkış and Kılkış (2016)	25		√	√		No	Yes (Composite indicator approach) (9 world busiest and best airports)
Li and Loo (2016)	29	√		√	√	No	Yes (Cost-Benefit Analysis (CBA)) (Hong Kong International Airport, HK SAR)
Olfat et al. (2016)	9	√	√	√	√	No	Yes (Fuzzy dynamic network-DEA model) (28 Iranian airports)
Chao et al. (2017)	16			√	√	Yes (Fuzzy Delphi method)	Yes (Simple additive weighting (SAW)) (5 international airports)
Carlucci et al. (2018)	9	√	√			No	Yes (DEA) (34 Italian airports)
Lu et al. (2018)	15	√		√	√	Yes (Experts' interviews, and brain storming)	Yes (Multiple-criteria decision-making (MCDM)) (3 Taiwanese airports)
Kaya and Erginel (2020)	15			√	√	Yes (Stepwise Weight Assessment Ratio Analysis (SWARA))	Yes (SWARA) (Ankara Esenboga Airport, Turkey)
Kumar et al. (2020)	43		√	√	√	No	Yes (Best worst method (BWM) and VIKOR method) (5 Indian airports)
Wan et al. (2020)	55	√	√	√	√	No	Yes (Synthetic evaluation model) (Guangzhou Baiyun International Airport, China)
Wang and Song (2020)	7	√	√			No	Yes (Network DEA) (8 Chinese airports and 4 Asian airports)
Kucukvar et al. (2021)	10	√	√	√	√	No	Yes (DEA) (30 major international airports)
Orkomy and Sharbatdar (2021)	45	√		√	√	Yes (Experts' questionnaire)	Yes (SAW) (Zahedan International Airport, Iran)
Yangmin et al. (2021)	18	√	√	√	√	No	Yes (Synergy measure model) (Zhengzhou International Airport, China)
Dimitriou and Karagkouni (2022)	25			√		No	Yes (Multi-objective method) (20 airports from Europe, United States and Asia)
Ramakrishnan et al. (2022)	31			√	√	No	No

* : EC: Economic dimension; OP: Operational dimension; EN: Environmental dimension; SO: Social dimension.

decision-making techniques were not initially designed for sustainability evaluation, they have shown practical applications in airports and other transport segments. However, there are still deficiencies in the effectiveness of the decision-making techniques, as they leave the issue of airport sustainability incomplete, either by failing to encompass the various stages along the life cycle or by excluding specific sustainability dimensions. Among all the traditional decision-making techniques, DEA has been the most used one. However, a significant limitation in the application of DEA arises when dealing with a large number of input and output indicators. The *"curse of dimensionality"*—a term coined to describe the decrease in discriminatory power as the number of inputs and outputs increases relative to the Decision-Making Units (DMUs)—suggests that a higher proportion of DMUs may be mistakenly identified as efficient (Charles et al., 2019; Cook et al., 2014). Furthermore, Wong (2021) pointed out that an increased number of inputs and outputs can restrict the weights assigned to these variables, leading to less discernible analysis results. Collectively, these factors suggest that the effective incorporation and balancing of all sustainability dimensions pose a challenge to DEA, especially when the index is high (Mo et al., 2018).

To overcome the effectiveness deficit, some authors have adopted the sustainability rating approach to conduct both ex-ante and ex-post sustainability evaluations across various dimensions. The composite indicator was first introduced by Kılkış and Kılkış (2016) to evaluate the environmental and operational sustainability of airports. Although only two dimensions were talked in this study, it demonstrated the feasibility and applicability of the composite indicator for evaluating airport sustainability. Building on this, Wan et al. (2020) and Yangmin et al. (2021) extended the composite indicator approach to cover the four dimensions of airport sustainability. Nevertheless, within these research efforts, the issue of subjectivity is invariably overlooked when selecting methods in the composite indicator building process.

As a trailblazing attempt to explore the subjectivity issue of composite indicators applied to the airport sustainability field, this paper will use two alternative methods in normalization, weighting, and aggregation processes, respectively. The combination of eight composite indicator building schemes will then be tested by the sensitivity and reliability analyses to examine the effects of normalization, weighting and aggregation on the final evaluation results and to investigate the scheme with the most significant contribution to the result reliability.

2.3. Application of composite indicator for sustainability evaluation

Indicators and composite indicators are gaining increasing recognition as useful tools for policy-making and public communication, conveying information on the sustainability performance of countries, industries or corporates in areas ranging from environmental, economic, social or operational development (Singh et al., 2009). According to the European Commission's first state-of-the-art report (Saisana and Tarantola, 2002, p.5), "*Composite indicators are based on sub-indicators that have no common meaningful unit of measurement, and there is no obvious way of weighting these sub-indicators*". Gasparatos et al. (2008) defined the composite indicator as an aggregation of different indicators according to a well-developed and pre-determined methodology. Composite indicators can be classified into two types depending on the priorities considered during their building process: Data-driven (or bottom-up) and Theory-driven (or top-down) (Niemeijer, 2002). Data-driven approaches are favored when data availability is a core issue during the indicator construction; Theory-driven approaches are taken when selecting the optimal possible indicators to fit in a composite indicator from a theoretical perspective, whilst data availability is only one of the many aspects affected.

The composite indicator approach has been progressively implemented since 2000 for sustainability evaluation in sectors including agriculture (Zinck et al., 2004; Gómez-Limón et al., 2010), steel (Singh et al., 2007), urban planning (Ciegis et al., 2011; Yigitcanlar et al., 2015), transport and tourism (Perez et al., 2013; Blancas et al., 2016). In academia, based on the terminal use of the composite indicators, current studies can be divided into two main categories: sustainable decision-making facilitating and sustainable performance benchmarking.

For the first category of studies on decision-facilitating, they will construct a criterion system via the composite indicators and select the optimal alternative based on the computational results. In simple terms, they digitize the decision-making process so that it is easier to determine what to choose and why to choose. Due to the direct correspondence between the decision rationale and the indicators selected, the outcome of such studies enables information communication and policy-making support to work simultaneously. For example, Dobos and Vörösmarty (2014) have developed a composite indicator system used for green suppliers' selection. These indicators were selected based on two criteria: managerial and environmental. The findings in the study were used to foster supplier management and purchasing decisions. Similarly, Sun et al. (2020) used the Analytic Hierarchy Process (AHP) to construct composite indicators for choosing the most sustainable wastewater management options.

For another type of study addressing benchmarking, the aim of its indicators is to capture a holistic view of the sustainability performance. The selection of indicators therefore relies heavily on the examination of existing outputs from industry or academia to make the indicator system as comprehensive as possible. Participatory methods which incorporate decision-making processes are rarely used by this type. We also found that although composite indicators have advantages of tackling the multi-dimensional nature of sustainability and making this complex issue more readily communicated to the public, majority of the authors failed to suggest further policy changes based on the performance assessment results. This phenomenon can be explained by the "*use*" and "*influence*" of composite indicators. According to Sébastien and Bauler (2013), *"'use' traces back the original intentions pursued by actors handling the indicator, 'influence' enables one to identify the ways in which indicators interact with policymaking"*. The application of composite indicators does not automatically lead to an impact on policy (Henry and Mark, 2003), whereas influence emerges through dialogue and argumentation (Valovirta, 2002). This suggests that in practical sustainability performance studies, supplementary analysis should be conducted on the composite indicator results to support policy decisions. Such complementary analysis is well demonstrated by Dizdaroglu and Yigitcanlar (2016), who developed an Urban Ecosystem Sustainability Index and applied it to the Gold Coast city (Australia). The index consists of six main sustainability goals and 14 associated sub-indicators. After obtaining scores for the six goals, Dizdaroglu and Yigitcanlar (2016) looked into the relevant policies in the Gold Coast City targeting each goal, and further clarified what policies work, what should be followed or vice versa. Another complementary analysis was carried out by purely mathematical models to investigate the determinants of sustainability performance among 81 first-grade olive oil mills in Andalusia (Spain) (Vicario-Modroño et al., 2022). Through the application of truncated regression analysis and bootstrapping techniques on the composite indicator results, Vicario-Modroño et al. (2022) identified factors, including quality commitment and manager training, as key drivers to the sustainability of olive oil mills. These results provide a direct reference for company managers to design and adapt their sustainability policies. In airport sector, upon establishing an airport environmental sustainability ranking index, Kılkış & Kılkış (2016) suggested policy recommendations to the airport operators for upgrading each environmental sustainability dimension. Depending on the controllability of the airport, these measures were classified as direct controllable by the airport, guidable and influenceable.

This paper falls into the second category of research. To achieve public communication, complex sustainability issues are transferred to a series of simple, easily understood and highly aggregated indicators. This implies a process of moving from the micro to the macro level. However, in order to make indicators a direct assistance to decision-making, either participatory or mathematical models are needed to examine potential micro factors that can contribute to the macro picture of sustainability. As the focus of this paper is on the composite indicator building and its subjective exploration, and given the space constraints, we will not provide a complete and detailed account of policy aspect. However, we will enumerate the directions in which the ASEI constructed in this paper may derive managerial and policy insights within the airport sector and beyond.

3. Index construction

3.1. Overview of procedures

A typical composite indicator construction involves three key phases: data normalization, data weighting and data aggregation (Gasparatos et al., 2008; Nardo et al, 2005; Saisana and Tarantola, 2002). This paper will also follow such construction principle. However, there is subjectivity embedded in the choice of methods during each phase, and the subjective choices can manipulate the results. To investigate the impact of the normalization, weighting and aggregation on the final composite indicator results, this paper will use two different methods to construct the ASEI in each of the three phases.

An illustration of the ASEI construction procedure is outlined in Fig. 2: the building process starts with indicator selection and data extraction for the four sustainability dimensions, which is explained in Section 3.1; next, the raw data will be normalized through Re-scaling and Distance to a reference, as described in Section 3.2; Section 3.3 addresses how weights are assigned through Equal weightings (EWs) and Benefit-of-the-doubt (BoD); Section 3.4 will specify the aggregation process through Linear Aggregation (LIN) and Non-compensatory multi-criteria approach (NCMC). This process will lead to 8 combination schemes to build the composite indicator with 8 sets of distinct output results. Finally, the sensitivity and reliability analyses will be conducted in Section 3.5 to test and determine how changes in outputs are related qualitatively and quantitatively to the different combinations of subjective decisions made.

To support an intuitive illustration and display of the ASEI construction, Schiphol is selected as a demonstrator of the full process.

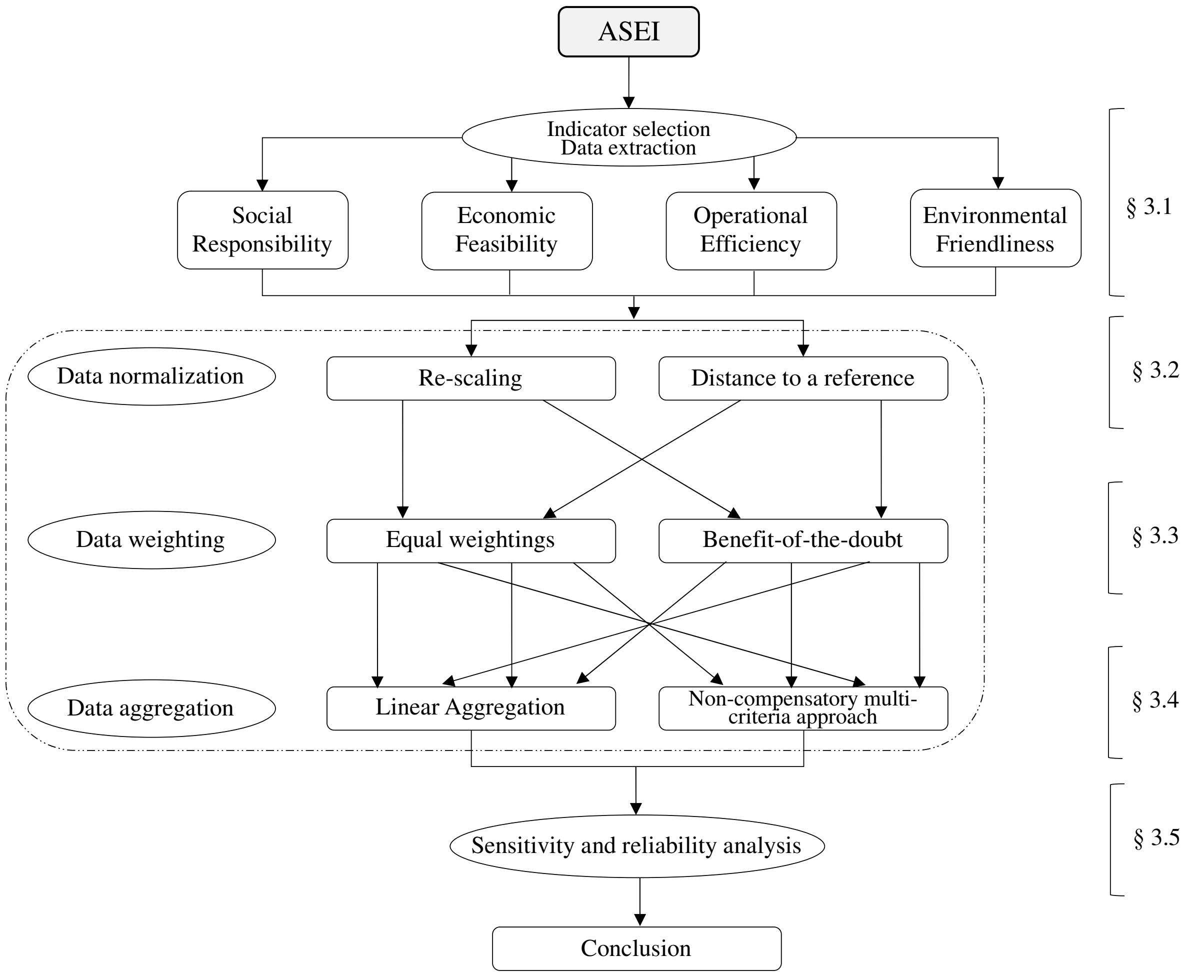

Fig. 2. ASEI building section (source: own development)

3.2. Indicators selection and data extraction

This paper follows a theory-driven approach in developing the composite indicator system, with the theoretical basis relying on the definition and scope of airport sustainability described in Section 2.1. The ASEI therefore incorporates four dimensions: Social Responsibility (D_1), Economic Feasibility (D_2), Operational Efficiency (D_3) and Environmental Friendliness (D_4).

To enable this indicator system be practical and transferable, the sub-indicators under each dimension ought to be accessible and easy-to-understand. In the selection and filtering process, we referred to existing literatures on indicators used for airport sustainability evaluation. Given the limited number of such studies, notably in the social dimension, both quantity and depth are significantly lower compared to the other three. We have supplemented this by drawing from sustainability evaluation studies outside of the airport domain. To this end, 28 sub-indicators across the four sustainability dimensions were obtained. The code, name, selection criteria, unit, impact characteristic, and reference source of these indicators are presented in Table 2.

Table 2. ASEI indicators set

Dimension (D_i)	Selection criteria	Indicator ($I_{i,j}$)	Unit	Impact on airport sustainability (P/N) [a]	Literature review (*) [b]
Social Responsibility (D_1)	Employment generation, employee development and safe working environment	$I_{1,1}$ Number of full-time employees	Dimensionless	P	Janic, 2010; Červinka, 2017; Kılkış and Kılkış, 2016; Kucukvar et al., 2021; Wan et al., 2020; Yangmin et al., 2021
		$I_{1,2}$ Employee benefits	Million Euro	P	Azapagic, 2004 *; Lee et al., 2021 *; Sreenath et al., 2021; Wang et al., 2018 *
		$I_{1,3}$ Lost Time Injury Frequency (LTIF) (excluding fire brigade) [c]	Dimensionless	N	Kucukvar et al., 2021
	Service quality, passenger loyalty and satisfaction	$I_{1,4}$ Skytrax world's airport ranking [d]	Dimensionless	N	Wang and Song, 2020
		$I_{1,5}$ Passengers' net promoter score (NPS) [e]	Dimensionless	P	Olfat et al., 2016; Sreenath et al., 2021; Wan et al., 2020
		$I_{1,6}$ Annual number of complaints	Dimensionless	N	Krajnc and Glavič, 2005 *; Veleva and Ellenbecker, 2001 *; Wan et al, 2020
	Noise monitoring and mitigation	$I_{1,7}$ Number of people who experience severe noise disturbance	Dimensionless	N	Yangmin et al., 2021
Economic Feasibility (D_2)	Revenue streams	$I_{2,1}$ Annual aviation revenue	Million Euro	P	Červinka, 2017; Kılkış and Kılkış, 2016; Kucukvar et al., 2021; Wan et al., 2020; Wang and Song, 2020
		$I_{2,2}$ Annual non-aviation revenue	Million Euro	P	Adler et al., 2013; Lu et al., 2018; Olfat et al., 2016; Wan et al, 2020; Wang and Song, 2020
	Financial resources	$I_{2,3}$ Annual operating expenses	Million Euro	N	Adler et al., 2013; Červinka, 2017; Wan et al, 2020
		$I_{2,4}$ Total assets	Million Euro	P	López et al., 2007 *; Wan et al, 2020
		$I_{2,5}$ Annual investment	Million Euro	P	Yangmin et al., 2021
	Profitability	$I_{2,6}$ Annual return on equity (ROE)	Dimensionless	P	López et al., 2007 *; Spicka et al., 2019 *; Wan et al, 2020
		$I_{2,7}$ Earnings before interest, taxes, depreciation and amortization (EBITDA)	Dimensionless	P	Červinka, 2017; Mukhambetov et al., 2020 *
Operational Efficiency (D_3)	Demand	$I_{3,1}$ Annual cargo load handled	Million Tonnes	P	Adler et al., 2013; Červinka, 2017; Janic, 2010; Kılkış and Kılkış, 2016; Wan et al, 2020; Wang and Song, 2020; Yangmin et al., 2021
		$I_{3,2}$ Annual passenger traffic (PAX)	Dimensionless	P	Adler et al., 2013; Červinka, 2017; Janic, 2010; Kılkış and Kılkış, 2016; Kucukvar et al., 2021; Wan et al., 2020; Wang and Song, 2020; Yangmin et al., 2021
		$I_{3,3}$ Annual air transport movements	Dimensionless	P	Adler et al., 2013; Janic, 2010; Upham and Mills, 2005; Wang and Song, 2020; Yangmin et al., 2021
	Connectivity	$I_{3,4}$ Number of direct destinations	Dimensionless	P	Dimitriou and Sartzetaki, 2020; Sartzetaki, M., 2019
		$I_{3,5}$ Percentage of transfer passengers	Dimensionless	P	Dimitriou and Sartzetaki, 2020; Sartzetaki, M., 2019
	Operational stability	$I_{3,6}$ Airline on-time performance	Dimensionless	P	Li and Loo, 2016; Wan et al., 2020
		$I_{3,7}$ Baggage handling incidence rate (IR)	Dimensionless	P	Sreenath et al., 2021

Table 2. (Continue)

Environmental Friendliness (D_4)	Emissions	$I_{4.1}$ Total Scope 1 and 2 CO2 emissions	Tonnes	N	Kılkış and Kılkış, 2016; Kucukvar et al., 2021; Li and Loo, 2016; Ramakrishnan et al., 2022; Sreenath et al., 2021; Upham and Mills, 2005; Wan et al., 2020
		$I_{4.2}$ Total Scope 3 CO2 emissions	Tonnes	N	Kucukvar et al., 2021; Li and Loo, 2016; Ramakrishnan et al., 2022; Upham and Mills, 2005; Wan et al., 2020
		$I_{4.3}$ Average annual mean PM2.5 concentration	microgram/m^3	N	Kılkış and Kılkış, 2016; Li and Loo, 2016; Wan et al., 2020
		$I_{4.4}$ Average annual mean NOx concentration	microgram/m^3	N	Ramakrishnan et al., 2022; Li and Loo, 2016; Sreenath et al., 2021; Upham and Mills, 2005; Wan et al., 2020
	Waste management	$I_{4.5}$ Waste separation percentage [f]	Dimensionless	P	Janic, 2010; Ramakrishnan et al., 2022; Sreenath et al., 2021
	Energy sourcing	$I_{4.6}$ Energy efficiency level	Dimensionless	P	Ramakrishnan et al., 2022; Sreenath et al., 2021
		$I_{4.7}$ Number of electric and biogas-powered vehicles (buses, vans and taxis) in use	Dimensionless	P	Ramakrishnan et al., 2022; Sreenath et al., 2021

[a] P: Positive; N: Negative.
[b] References marked with an " * " indicate that this sustainability evaluation study is outside the airport domain.
[c] The LTIF rate at Schiphol represents the number of work-related accidents resulting in absenteeism per million hours worked.
[d] The Skytrax world's airport ranking is developed through a survey assessing passengers' experience of different airport services, including check-in, arrival, transfer, shopping, security, and immigration. through to departure from the gate.
[e] The NPS is a passenger loyalty indicator that measures how likely passengers are to recommend Schiphol airport to friends, family and colleagues (on a scale of 0 to 10).
[f] The sources of waste here include terminals, offices and aircrafts.

The statistical data for indicators $I_{4.3}$ (Average annual mean PM2.5 concentration) and $I_{4.4}$ (Average annual mean NOx concentration) were collected from the website of the Dutch National Institute for Public Health and the Environment (RIVM, 2022). The values used in the analysis are the mean values of three monitoring stations (Badhoevedorp-Sloterweg, Hoofddorp-Hoofdweg and Oude Meer-Aalsmeerderdijk) around Schiphol. The location of the three monitoring stations is depicted in Fig. 3. The data for the remaining indicators were collected from the Royal Schiphol Group's annual reports (Royal Schiphol Group, 2022) for the years 2012 to 2021.

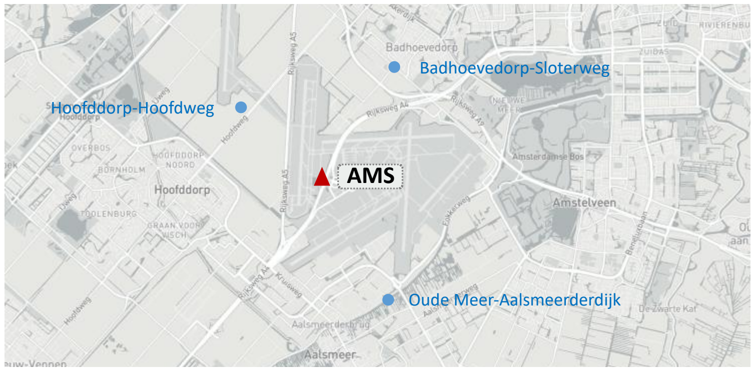

Fig. 3. Air quality monitoring stations around Schiphol
(source: own development based on the geographical information from RIVM (2022))

3.3. Data normalization

After the ASEI has been constructed, the indicators are usually not commensurate with each other or have different units of measurement. A normalization process is needed to bring these indicators up to the same standard by converting them into pure, dimensionless numbers so that they can be compared and aggregated in the later phases. Common normalization methods include Re-scaling, Distance to a reference and Z-scores (Nardo et al., 2005; OECD, 2008; Saisana and Tarantola, 2002). If successive years of indicator data are available, it is also feasible to employ the *Percentage of annual differences over consecutive years* which refers to a process used to normalize time-series data by expressing the difference between consecutive data points as a percentage of the value in the previous year. The choice of normalization method depends predominantly on the characteristics of the data sample itself. If an ill-fitting normalization method is chosen, it can cause the data being over-normalized. This will consequently lead to a loss of information and affect the results.

The data sample in this paper does not respect the normal distribution. If using the Z-scores method, the data will be transformed into a normality pattern, in which case the original distributional characteristics of the data would be disrupted. In addition, the existence of zero values in the sample also precludes the Percentage of annual differences over consecutive years. Therefore, we apply the Re-scaling and Distance to a reference methods for normalization.

3.3.1. Re-scaling (Min-Max)

This method is driven based on the dataset's range of values for each indicator. The indicators with a positive impact on the airport sustainability, $I^{+}_{i,j,t}$, are converted into a normalized form by equation (1):

$$I^{+}_{N_{i,j,t}} = \frac{I^{+}_{i,j,t} - I^{+,Min}_{i,j}}{I^{+,Max}_{i,j} - I^{+,Min}_{i,j}} \tag{1}$$

While the indicators with a negative impact on the airport sustainability, $I^{-}_{i,j,t}$, are normalized according to equation (2):

$$I^{-}_{N_{i,j,t}} = \frac{I^{-,Max}_{i,j} - I^{-}_{i,j,t}}{I^{-,Max}_{i,j} - I^{-,Min}_{i,j}} \tag{2}$$

Where $I^{+}_{i,j,t}$ and $I^{-}_{i,j,t}$ are values for indicator $I_{i,j}$ in year t with positive and negative impact on airport sustainability, respectively; $I^{+}_{N_{i,j,t}}$ and $I^{-}_{N_{i,j,t}}$ are their normalized forms, respectively. $I^{+,Max}_{i,j}$ represents the highest value of the positive indicator $I_{i,j}$ over the entire period analyzed, while for the negative indicator, it is denoted as $I^{-,Max}_{i,j}$. On the

contrary, the lowest value for indicator $I_{i,j}$ with positive and negative impact are denoted as $I_{i,j}^{+,Min}$ and $I_{i,j}^{-,Min}$, separately.
In this case, all normalized data will drop into the interval from 0 to 1. It is noticeable that the normalization method is time-dependent, which implies that the data will need to be recalculated when data for a new time-point becomes available, as the minimum or maximum values of some indicators may have changed.

3.3.2. Distance to a reference

When applying the distance to a reference normalization, the normalized value is calculated as the ratio between the indicator and a reference value. The reference can be a target to be achieved within a specific timescale, the best performer within the dataset (in this case, the method can also be named as "*distance to the best performer*"), or a universal reference baseline.
In this paper, we choose the reference as the best performer within the whole analyzed period and normalize each indicator according to equations (3) and (4):

$$I_{N_{i,j,t}}^{+} = \frac{I_{i,j,t}^{+}}{I_{i,j}^{Benchmark}} \tag{3}$$

$$I_{N_{i,j,t}}^{-} = \frac{I_{i,j}^{Benchmark}}{I_{i,j,t}^{-}} \tag{4}$$

where $I_{i,j}^{Benchmark}$ is the best performer for indicator *j* from the *i*th dimension, and its value is set to 1. The normalized values from this method reflect the percentage away from the leader.

3.4. Data weighting

Central to constructing a composite index is to combine different dimensions measured at different scales in a meaningful manner. This means deciding which weighting model to use, and which procedure to follow in order to aggregate the information.
The weights of indicators can be obtained either based on statistical models such as Principal Component Analysis (PCA), DEA, and BoD; or according to participatory methods such as Analytic Hierarchy Process (AHP), Budget Allocation Process (BAP), and Conjoint Analysis (CA).
The weighting methods employed in this paper are statistical based: EWs and BoD.

3.4.1. Equal weightings

EWs entails a recognition of equal status for all indicators. It can also be used as an alternative measure when there is no or limited information on the relative importance of indicators (Nardo et al, 2005, p.55). To date, the majority of composite indicator systems are weighted by EWs. As one of the most widely used and well-known composite indicator systems, the Human Development Index (HDI), developed by the United Nations Development Programme (UNDP, 2014), uses the EWs to give equal weight to three dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. In any case, EWs does not indicate that no weights are assigned but impliedly that the weights are equal. Accordingly, we consider the four sustainability dimensions and the seven sub-indicators under each dimension to be of equal importance, with the weights of each dimension and sub-indicator being one-quarter and 1/7 respectively.
Despite the simplicity and transparency, EWs may results in a loss of information, particularly in cases where some indicators are more important than others in reflecting the underlying concept that the composite indicator system is designed to measure.

3.4.2. Benefit-of-the-doubt

The BoD approach is an adaptation of DEA and was initially developed to assess macroeconomic performance (Melyn and Moesen, 1991). It was introduced into index theory since 2000 (Cherchye and Kuosmanen, 2002; Cherchye, Moesen and Puyenbroeck, 2004).
The BoD stems from a fundamental conceptual starting point of DEA that the information on specific weighting schemes for national performance benchmarking can be extracted from the countries' own data. The underlying idea

is that a country's good relative performance on a given dimension indicates that the country perceives its corresponding policy dimension to be relatively important.

The weighting scheme applied in this paper will follow a two-layer approach.

First layer: Weighting of airport sustainability sub-indexes

Under the definition of airport sustainability given in Section 2, the equal importance of the four dimensions will lead to the four airport sustainability sub-indexes $I_{AS_{i,t}}$ have an equivalent weight w_i (social responsibility dimension: $i = 1$; economic feasibility dimension: $i = 2$; operational efficiency dimension: $i = 3$; environmental friendliness dimension: $i = 4$) equal to 0.25. Therefore, the overall airport sustainability performance in year t can be expressed by equation (5):

$$CI_{AS_t} = \sum_i I_{AS_{i,t}} \cdot w_i \tag{5}$$

Second layer: Weighting of airport sustainability indicators under each sub-index

Following the BoD approach, the airport sustainability sub-index $I_{AS_{i,t}}$ in year t can be extracted through a linear programming as illustrated in equation (6):

$$I^*_{AS_{i,t}} = \underset{w_{i,j}}{\arg max}\left(\sum_j I^+_{N_{i,j,t}} \cdot w_{i,j} + \sum_j I^-_{N_{i,j,t}} \cdot w_{i,j}\right) \tag{6}$$

$$s.t. \quad I^{Benchmark}_{i,j} \cdot w_{i,j} \leq 1$$

$$\sum_i w_{i,j} = 1$$

$$w_{i,j} \geq 0$$

Where $w_{i,j}$ denotes the weight of indicator j from the i^{th} dimension, reflecting the importance assigned to this indicator during the airport sustainability assessment.

In the above case, the BoD model allows weights to be estimated freely in order to optimize the relative score of the evaluated airport, except for two restrictions on the weighting (the weight must be nonnegative, and it does not lead to a final score above the upper boundary of 1). This flexibility has the benefit of making it difficult for airports to argue that it is the weights putting them in a detrimental position. However, the full flexibility also has drawbacks. In particular, it may allow an airport to outperform in a way that is difficult to justify. For example, if some zero weights are assigned and no prior information is available to back up this possibility, then some achievement indicators will fail to contribute to an airport's composite metric (Cherchye et al, 2008). As a result, weight restrictions are introduced to prevent unrealistic evaluation results due to under- or over-weighting. In existing literature, there are not yet unified values for the weighting limits. In this paper, the lower and upper limits are set at 2% and 35% according to the values suggested by Verbunt and Rogge (2018):

$$L_r \leq w_{i,j} \leq U_r \tag{7}$$

3.5. Data Aggregation

As stated by the OECD (2008), when individual indicators are compiled into a single index, a composite indicator is established upon the underlying model of the measured multi-dimensional concept. The quality and robustness of a composite indicator system rely crucially on the baseline construction scheme, of which data aggregation is a key step (Zhou, Fan and Zhou, 2010).

The sequence of steps used in this study is to firstly group the selected indicators, $I_{i,j}$,into airport sustainability sub-indexes, $I_{AS_{i,t}}$. The $I_{AS_{i,t}}$ of each dimension is then merged into ASEI, with the final evaluation score represented by CI_{AS_t}. In this paper, LIN and NCMC are applied to implement such aggregation processes.

3.5.1. Linear Aggregation

The LIN is by far the most broadly adopted aggregation method that sums up the normalized and weighted sub-indexes (Nardo et al., 2005, pp. 74-75):

$$CI_{AS_t} = \sum_i I_{AS_{i,t}} \cdot w_i \tag{8}$$

$$\sum_i w_i = 1$$

$$w_i \geq 0$$

However, an undesirable feature of LIN is its full compensatory, which implies that poor performance on some indicators can be compensated against by sufficiently high values from other indicators.

The issue of compensation can be partly solved by Geometric Aggregation (GME), also known as the weighted geometric mean. In a GME process, the individual indicator scores are first multiplied together and then raised to a power equal to the weight of each indicator. This aggregation method can partially solve the compensation problem by considering trade-offs between indicators, as a reduction in one indicator will result in a reduced overall composite score, regardless of other indicators' values. The choice of GME is also advised when non-comparable and strictly positive indicators are expressed on a different scale of ratios (Nardo et al, 2005). However, this aggregation method is not applicable when a data sample contains zero or negative values, leaving it excluded from our analysis. In addition, a substantially lower value for one indicator can significantly reduce the final composite score, even if other indicators have higher values. This may lead to a more rigorous or conservative evaluation of the final composite score.

3.5.2. Non-compensatory multi-criteria approach

When using an additive or a multiplicative aggregation rule, especially in our sample where most indicators are expressed in terms of intensities (e.g. in million euros or tones) rather than qualities (e.g. good, bad, medium) or rankings, the substitution rates are equal to the weights of the variables up to a multiplicative coefficient. The weights in these aggregation schemes inevitably have the meaning of substitution rates and do not signify the absolute importance of the indicator concerned (OECD, 2008, p.112; Nardo et al, 2005, p.76). Although the data are normalized in the former step by transforming the intensities of indicators into a common scale. However, normalization alone does not eliminate the compensatory issue. The weights assigned to each normalized indicator still reflect the relative importance of the indicators in terms of substitution, i.e. the trade-offs between the indicators, rather than their absolute importance.

To overcome the compensatory issue, we introduce the NCMC method proposed by the Joint Research Center (JRC) of the European Commission (Nardo et al., 2005). This approach aims to eliminate the compensatory issue by treating indicators as criteria that cannot be compensated for by other indicators. This method is based on the principle that "*a good is a good"* and does not consider trade-off between criteria. There are two main steps regarding the NCMC: (1) pairwise comparisons of the sustainability indicators' values throughout the analysis period; and (2) ranking the concerned years according to the sustainability performance they achieved. These steps will run for two rounds to aggregate the $I_{AS_{i,t}}$ and CI_{AS_t} respectively.

For illustration purposes, the normalized data from re-scaling and the weights extracted from EWs are selected as the inputs for NCMC. A step-by-step procedure for aggregation over NCMC is described in Fig. 4 below:

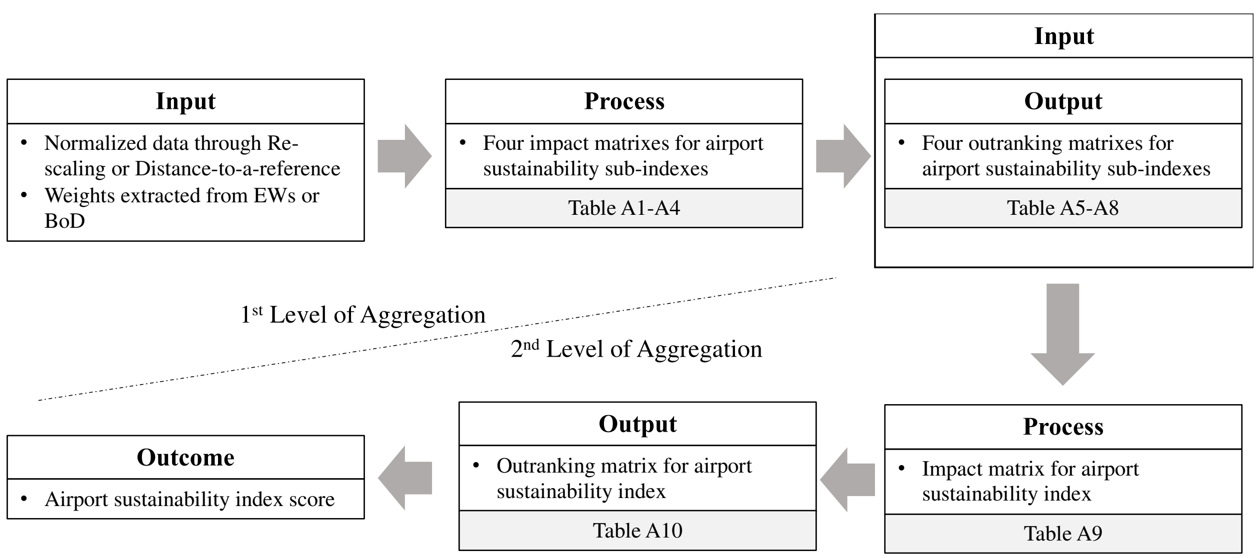

Fig. 4. Aggregation procedures through NCMC (source: own development)

Step 1: First level of aggregation – aggregating airport sustainability indicators

Based on the inputs, four impact matrixes are built corresponding to the four airport sustainability dimensions, respectively (see Table A1-A4 in Appendix A). Then, the pairwise comparisons are conducted within each impact matrix to construct an outranking impact matrix for $I_{AS_{i,t}}$.

During the pairwise comparison process, the score allocated for each year t is the sum of the weights for indicators that have performed better in year t. Since all indicator data is normalized, a larger indicator value represents better performance for both positive and negative indicators. If both years have the same indicator value, the former year is considered the better performer. Taking the social responsibility dimension as an example, for the year 2012, by comparison with 2013, the score is equal to 2/7, because in 2012 Schiphol had higher values on indicators $I_{1.1}$ and $I_{1.7}$. Accordingly, the score of 2013 is equal to 5/7 compared to 2012.

Further, the scores within each airport sustainability dimension will be ranked. If two or more years achieved the same score, the ranking from the earliest year would be higher as no improvements in sustainability development were achieved in the following years. However, this ranking rule does not necessarily entail that maintaining the same level of sustainability performance over the years corresponds to a step back in sustainable development. This interpretation depends on the specific objectives and expectations of the airport operator. If an airport operator has ambitious sustainability goals and wants to achieve them within a limited time frame, then not making progress on these goals can indeed be considered as a step back. In our case, Schiphol has set several sustainability targets, one of which is to be carbon neutral by 2030. Further, Schiphol has always been considered one of the best airports globally and holds a high reputation. Therefore, the ranking rule we have defined for the Schiphol case is less conservative.

The results of the pairwise comparisons were then combined into ranking matrixes for the four airport sustainability dimensions, as described in Tables A5-A8 in Appendix A.

<u>Step 2: Second level of aggregation – aggregating airport sustainability sub-indexes</u>

After obtaining the ranks of the airport sustainability sub-index by aggregating individual indicators, the ranks over the whole analysis period can be summarized in a new impact matrix (see Table A9 in Appendix A) for the second level of aggregation. The outranking matrix on the overall airport sustainability composite index CI_{AS_t} can be built through a similar pairwise comparison scheme, as described in Table A10. The rankings can be converted into the composite index by dividing the highest possible permutation (Nardo et al., 2005; OECD, 2008). The sample here contains data for ten years, thus the highest possible pairwise comparative permutation is equal to 9.

Consequently, the Schiphol achieved the highest sustainability rank in 2017 in this example, with a CI of 0.72.

3.6. Sensitivity and reliability analysis

Sensitivity and reliability analysis is invariably overlooked in studies using the composite indicator to solve practical problems. However, the choice of normalization, weighting and aggregation methods all gives rise to uncertainty when building a composite indicator system. Two primary questions therefore emerged from the building process: (1) to what extent do the three steps of normalization, weighting and aggregation contribute to the final composite indicator results; and (2) to what extent do the results from the eight normalization-weighting-aggregation combination schemes differ, and how reliable are they as a group.

To answer the above questions, this study adopts variance-based sensitivity analysis, average shift in rankings (ASR), and Cronbach's alpha test to investigate the qualitative and quantitative associations between the results and the combination schemes.

3.6.1. Variance-based sensitivity analysis

Variance-based sensitivity analysis (or the Sobol method) is a form of global sensitivity analysis (Sobol, 2001). For a model $Y = f(X_1, X_2, \dots, X_p)$, where the model output Y is a scalar and the inputs $X_1, X_2, \dots, X_p$ are considered to be independent random variables characterized by known probability distributions. The core idea of this method is to decompose the variance of a model output into fractions that can be attributed to each input factor.

In order to quantify the importance of an input factor X_i on the variance of Y, the "true value" of X_i is assumed to be X_i^*. The change in the variance of Y due to this assumption can be described as a conditional variance $V_{X_{-i}}(Y|X_i = X_i^*)$. The variance is taken over the $(p-1)$-dimensional parameter space X_{-i}, consisting of all factors but X_i. Because the value of X_i^* is unknown, the variance of the resulting function of X_i^* are taken over all possible X_i^* values: $E_{X_i}(V_{X_{-i}}(Y|X_i))$. Then, by applying the law of total variance (Sobol, 1993), equation (9) can be established:

$$V(Y) = V_{X_i}\left(E_{X_{-i}}(Y|X_i)\right) + E_{X_i}(V_{X_{-i}}(Y|X_i)) \tag{9}$$

Through normalizing, the equation (9) can be further transformed into equation (10):

$$1 = \frac{V_{X_i}\left(E_{X_{-i}}(Y|X_i)\right)}{V(Y)} + \frac{E_{X_i}(V_{X_{-i}}(Y|X_i))}{V(Y)} \tag{10}$$

Where the first term in (10) represents the first-order sensitivity index (or main effect index) for factor X_i :

$$S_i = \frac{V_{X_i}\left(E_{X_{-i}}(Y|X_i)\right)}{V(Y)} \tag{11}$$

This sensitivity index indicates the contribution of a single input factor X_i to the variance of the model output (Saisana, Saltelli and Tarantola, 2005). The value of S_i must always not exceed 1 according to equation (10).
By a similar approach, the conditional variances corresponding to more than one factor can also be calculated. Using the two input quantities X_i and X_j as an example, a second-order term variance contribution can therefore be written as:

$$V_{ij} = V_{X_i X_j}(E_{X_{-ij}}(Y|X_i, X_j)) - V_{X_i}\left(E_{X_{-i}}(Y|X_i)\right) - V_{X_j}\left(E_{X_{-j}}(Y|X_j)\right) \quad (12)$$

The second-order sensitivity index, S_{ij} , representing the amount of variance for Y explained by the interaction of the two input factors X_i and X_j is outlined by equation (13):

$$S_{ij} = \frac{V_{ij}}{V(Y)} = \frac{V_{X_i X_j}(E_{X_{-ij}}(Y|X_i, X_j))}{V(Y)} - S_i - S_j \quad (13)$$

Accordingly, for p input factors, there will be $2^p - 1$ sensitivity indexes. However, this can lead to high computational costs for calculating all higher-order terms. Homma and Saltelli (1996) introduced a total-order sensitivity index, S_{Ti} , into the Sobol method. S_{Ti} accounts for the total contributions to output variation due to the input factor X_i , including the first-order effect and other higher-order interactions. The S_{Ti} is given by Equation (14):

$$S_{Ti} = \frac{E_{X_{-i}}(V_{X_i}(Y|X_{-i}))}{V(Y)} \quad (14)$$

The additivity of a model can be concluded simply from a comparison between S_i and S_{Ti}. For additive models, $S_i = S_{Ti}$; whereas for non-additive models, $S_i < S_{Ti}$. In addition, the value of $S_{Ti} - S_i$ denotes the level of involvement for X_i in any interaction with other input factors.
When building up a composite indicator system, uncertainty arises in each of the three steps regarding the choice of normalization, weighting and aggregation methods. A Monte Carlo approach is used to simulate this process, which involves performing multiple evaluations of the model with three selected model input factors corresponding to each of the three steps (Nardo et al, 2005, p.89; Saisana, Saltelli and Tarantola, 2005).
The three uncertainty factors are grouped in Table 3 with their associated probability density functions (PDFs).

Table 3: The three uncertainty factors in composite indicator building

Input Factor	PDF	Range
X_1 (selection of normalization method)	Uniform	[0,1] Where [0, 0.5] = Distance to a reference (0.5, 1] = Re-scaling
X_2 (selection of weighting method)	Uniform	[0,1] Where [0, 0.5] = Equal weighting (0.5, 1] = Benefit of the doubt
X_3 (selection of aggregation method)	Uniform	[0,1] Where [0, 0.5] = Linear aggregation (0.5, 1] = Non-compensatory multi-criteria approach

X_1, X_2 and X_3 are all discrete random variables. In the Monte Carlo approach, they are generated by drawing a random number ζ uniformly distributed in the interval [0,1] and then applying the Russian roulette algorithm. Taking the first uncertainty factor – selection of normalization method, X_1, as an example: if $\zeta \in [0, 0.5]$, the distance to a reference method will be used for normalization; if $\zeta \in (0.5, 1]$, the re-scaling will be used. Similar procedures are followed by the second and third uncertainty factors. In this paper, a quasi-random sampling scheme (Sobol, 1967) is used for generating ζ. The sample size n an vary in the 100–1000 range (Saisana et al., 2005). We hereby apply the most commonly adopted base sample size of $n = 512$.
Upon operating the variance-based sensitivity analysis, the independent and interactive effects of the normalization, weighting and aggregation processes on the final composite indicator results will be determined.

3.6.2. Average shift in ranking

ASR is a well-used technique designed to test the robustness of composite indicators, notably where multiple schemes are used for composite indicator building (Liew, Che Ros and Harun, 2019; Hudrliková, 2013; De Montis et al, 2020). The ASR is given by equation (15):

$$\overline{R}_s = \frac{1}{M}\sum_{c=1}^{M}\left|Rank_{ref}(CI_c) - Rank(CI_c)\right| \quad (15)$$

$\overline{R}_s$ stands for the average of the absolute differences in rankings relative to the reference ranking across all M years. This value has the significance of encompassing the relative positional changes among the whole analyzed years into a single number. The median rank is taken as the reference ranking during most ASR analyses. A $\overline{R}_s$ value closer to zero implies that the rank is closer to the median.

3.6.3. Cronbach's alpha test

The Cronbach's alpha was developed initially by Crobach (1951) as a measure of reliability and is used here to test the internal consistency of the rankings given by different composite indicator building schemes.
The formula for Cronbach's alpha is given in equation (16):

$$\alpha = \frac{N\overline{c}}{\overline{v}+(N-1)\overline{c}} \quad (16)$$

where N is the number of composite indicator building schemes, $\overline{v}$ is the average variance, and $\overline{c}$ is the average inter-item covariance.
An α value above 0.9 indicates significant consistency between ranking schemes, whereas a value below 0.7 implies insufficient consistency.

4. Index implementation and extensions

4.1. Indicators processing and analytical results

Following the composite indicator building procedures elaborated in Section 3, the raw data collected for the 28 airport sustainability indicators across four dimensions will be normalized, weighted, and aggregated subsequently. The combination schemes during this process are organized in Table 4:

Table 4. Combination schemes used for composite indicator building

Combination scheme	Normalization method	Weighting method	Aggregation method
DEL	Distance to a reference	Equal weightings	Linear aggregation
DEN	Distance to a reference	Equal weightings	Non-compensatory multi-criteria approach
DBL	Distance to a reference	Benefit of the doubt	Linear aggregation
REL	Re-scaling	Equal weightings	Linear aggregation
REN	Re-scaling	Equal weightings	Non-compensatory multi-criteria approach
RBL	Re-scaling	Benefit of the doubt	Linear aggregation
DBN	Distance to a reference	Benefit of the doubt	Non-compensatory multi-criteria approach
RBN	Re-scaling	Benefit of the doubt	Non-compensatory multi-criteria approach

Therefore, we can obtain eight sets of composite index scores and rankings for the four sustainability dimensions. Fig. 5 portrays their uncertainty analysis results with the highest rank, lowest rank, median rank, and mean rank labeled. It can be concluded that:

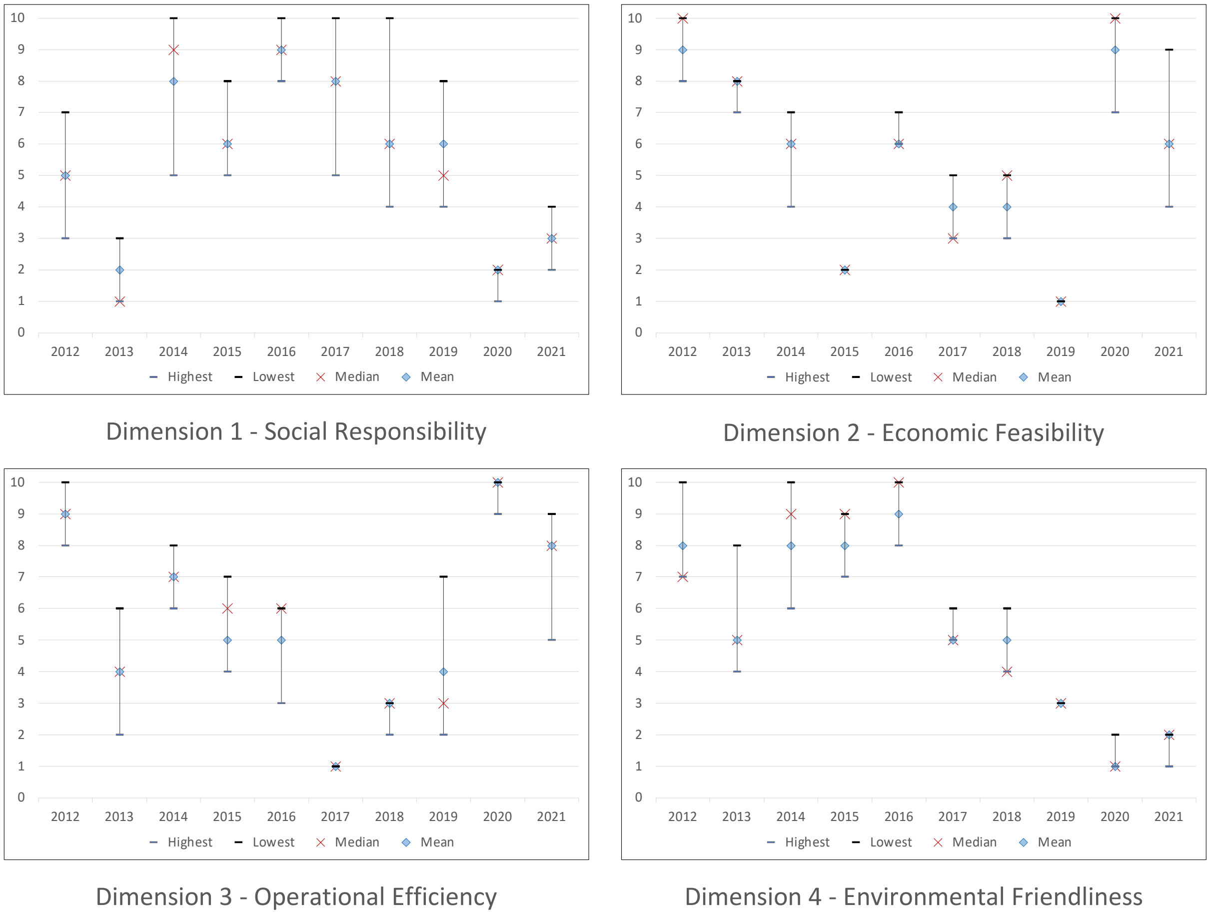

Fig. 5. Uncertainty analysis results of four airport sustainability dimensions

- The composite index rankings given by the different building schemes can vary to a certain extent: the largest difference in ranking is 5 places, whilst nearly sixty percent of the yearly ranking positions have a maximum difference of two places or less. However, the median and mean values among the eight sets of results are comparable, with the difference all equal to or less than one place.
- A prominent tipping point among the four graphs occurs in 2020, which is also the time point when Covid-19 struck globally. The economic dimension has suffered the most from this pandemic, with its sustainability index plummeting from the highest score in 2019 to the lowest level in the ten years analyzed, comparable to the score achieved in 2012. From 2020 to 2021, there was a slight rebound, but its sustainability level was still lower than that from 2014 to 2019. The situation is similar for the operational dimension, emerging at the bottom sustainability level in 2020 and experiencing a modest recovery for the year after. The sustainability performance for the social and environmental dimensions has steadily improved since 2016 and reached their highest levels in 2020. From 2020 to 2021, the sustainability ranking for both dimensions dropped by only one place. This indicates that the impact of the pandemic on the social and environmental sustainability of Schiphol is relatively mild and reflects the more resilient sustainability capability of Schiphol in these two dimensions.
- Another noteworthy tipping point came in 2013, where significant sustainability improvements were made in terms of social, operational, and environmental dimensions. The underlying reason can be attributed to Schiphol starting to implement its Master Plan (MP) in 2013. The MP plan spans four years with the ambitious goal of creating one of the most sustainable and high-quality airports in the world. However, with

the exception of economic feasibility, which saw sustainability improvement from 2013 to 2014, all three dimensions showed a downward trend, with social in particular. Besides, these three dimensions have not regained their sustainability levels observed in 2013 during the remaining two years of the MP strategy.

Synthesizing all four dimensions of the airport sustainability index scores, the overall airport sustainability ranking is given in Fig. 6. According to Fig. 6, the overall ranking derived from the eight composite indicator building schemes exhibits more variation than the amount within each sustainability dimension. The reason is that the sustainability sub-indexes will be aggregated into the overall sustainability index during this process, in turn creating an increase in the overall uncertainty.

However, the eight building schemes have reached a high degree of consensus on identifying the outstanding and worst performers. The year 2019 was the only year in which all schemes assigned a ranking in the top three positions, and therefore 2019 was considered the outstanding performer over the ten years analyzed. Among the schemes, five identified it as the best performer, two as the second-best performer, and one as the third best. Oppositely, the rankings for 2012, 2014, and 2016 are all concentrated in the bottom four positions and are classified as the worst performers.

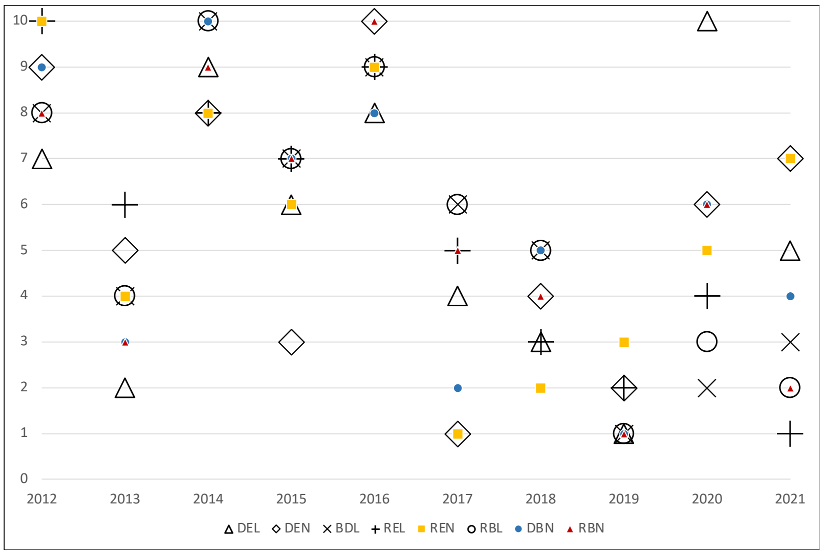

Fig. 6. Uncertainty analysis result of the overall composite indicator rankings (source: own development)

Fig. 7. Breakdown of airport sustainability scores for the years 2013, 2019 and 2020 (source: own development)

Taking three typical years from the above analysis: the Sustainable MP strategy implementation year (2013), the outstanding performer year (2019), and the Covid-19 outbreaking year (2020). These three years' average scores across the four sustainability dimensions are collapsed, as shown in Fig. 7. In 2013, the four dimensions of Schiphol's sustainability scores were connected in a rather regular diamond shape: the social and operational dimension scores are similar and outperform than environmental and economic dimensions. By 2019, Schiphol has maintained the same level of sustainability in the operational dimension as in 2013, with significant improvements in the economic and environmental dimensions, whereas a clear drop in the social dimension. The four sustainability dimensions exhibited the greatest variation in 2020. In addition, the Covid-19 has reshaped the Schiphol's sustainability pattern, switching from business- and operations-oriented in 2019 to environment- and society-oriented in 2020.

The four dimensions and their overall composite index scores obtained earlier in this Section demonstrate that different composite indicator building schemes can make a significant difference to the results. Therefore, the variance-based sensitivity analysis is performed to decompose the variance of the output and investigate key drivers of the variance. The first-order sensitivity indexes outlined in Fig. 8 illustrate the amount of contribution made by the normalization, weighting, and aggregation processes separately to the final results.

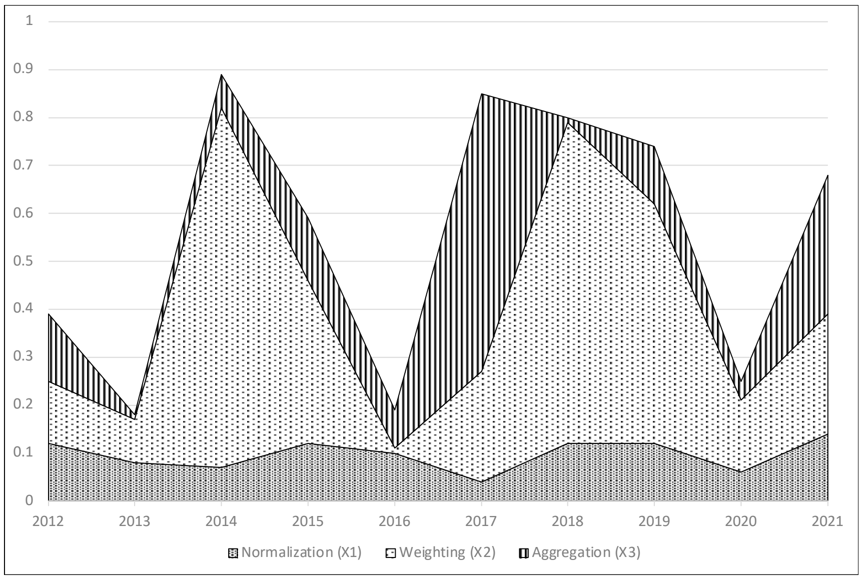

Fig. 8. Sensitivity analysis results based on first-order sensitivity index (source: own development)

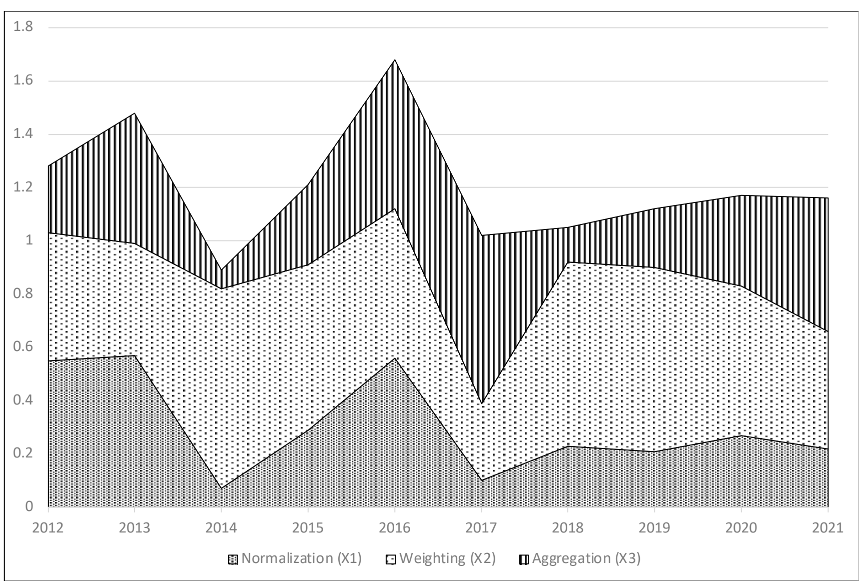

Fig. 9. Sensitivity analysis results based on total-order sensitivity index (source: own development)

According to the findings, all three processes have contributed to the final results, with the total variance differing each year. Nevertheless, the weighting process occupies the largest area in Fig. 8, with a total contribution to the output variance equal to 56.03%, and is therefore considered the most influential process in ASEI construction. Next is the aggregation process, which accounts for 26.49% of the output variance. Normalization is the least affected of the three processes, with a contribution of 17.48%.

The total contribution of the ASEI construction process (individually and by interactions) is measured by the total-order sensitivity indexes and is sketched in Fig. 9. By comparison with Fig. 8, the sum of the three processes' total-order sensitivity indexes each year is greater than the sum of the first-order sensitivity indexes that year, which is $S_{Ti} > S_i$. This verifies that all combination schemes are non-additive. Additionally, the gap between the three processes' total-order sensitivity indexes is narrowed compared to the first-order sensitivity index. Weighting remains the most influential process, explaining 45.60% of the total variance. Normalization and aggregation contributed 25.46% and 28.94% of the variance, respectively. In conclusion, weighting is the key process in the ASEI construction, with its individual and interactive contribution to output variance significantly higher than the others.

To examine the reliability of the eight composite indicator building schemes as a group, the Cronbach's alpha reliability test and ASR were used, and the results are presented in Tables 5 and 6, respectively. According to Table 5, the calculated value of Cronbach's alpha is 0.952, representing that the eight combination schemes exhibited excellent internal consistency. Table 5 also gives the Cronbach's alpha value after removing one scheme from the eight combinations. Based on the results, the DBN and RBN schemes would drop the most in group reliability if they were removed.

Table 5. The Cronbach's alpha reliability test

Cronbach's Alpha	Cronbach's alpha if an item deleted							
	DEL	DEN	DBL	REL	REN	RBL	DBN	RBN
0.952	0.952	0.952	0.946	0.946	0.949	0.944	0.939	0.939

Table 6. The average shift in rankings

	DEL	DEN	DBL	REL	REN	RBL	DBN	RBN
$\overline{R}_s$	1.3	1.6	0.8	1.1	1.5	0.8	0.8	0.5

Further, the RBN has the lowest value on $\overline{R}_s$, equalling 0.5, implying that the sustainability ranking given by the RBN is consistently situated close to the median position throughout the analysis period.

4.2. Transferability and policy insights

This paper selects Schiphol as a demonstration for building the airport sustainability composite indicators. The indicators and methodology can also be applicable to other hub and regional airports. During the transfer process, it's worth noting that although these indicators are theory-driven based on the principles of sustainability, in practice, they can vary in relation to the airport's own operational focus, its position in the air transport network and the data availability. In this context, the same sustainability criteria can be reflected from different indicators. For example, according to existing studies, the airport's service quality can be captured from Skytrax rankings, Google reviews or through surveys or interviews with passengers conducted by the airport operators. Similarly, the same sustainability issue can be considered by different criteria. A common example is that airport noise can be represented in the environmental dimension as a form of pollution, however, it can also be reflected in the social context through the number of noise complaints from local communities, as ensuring pleasant living conditions is also a social responsibility of the airport. Furthermore, for some regional airports with limited resources, they may not or not be able to strike a balance between the different dimensions. For addressing the above issues, when localizing the sustainability composite indicators to other airports, it is essential to start with a stakeholder communication to better understand on the role of the concerned airport expects to play in sustainable development. Likewise, these indicators and methodology can be adapted to assess the sustainability performance of other aviation components, such as airlines or air cargo carriers.

As a standard and complete framework, starting with normalization, weighting, aggregation through to the final sensitivity and reliability tests, this process can also be considered as a reference for sustainability assessments beyond the aviation industry. Although the choice of indicators can vary considerably due to the unique characteristics of the industries, the procedures for establishing and analyzing the sustainability composite indicators are identical and can be easily followed by other industries.

As discussed in Section 2.3, it is a common pitfall in composite indicator benchmarking studies that the policy insights are often overlooked or under-emphasized due to the potential gap between the *"use"* and *"influence"* of indicators. Whilst this paper is methodologically oriented on the process of constructing composite indicators, we also elaborate here on how this framework can be complemented to drive policy change and facilitate decision-makings:

- Use of sensitivity or regression analysis to investigate the individual influence of the indicators on the final composite indicator score. By highlighting the most and least critical influencing factors on airport sustainability, policymakers can identify the areas where they should focus their efforts and resources to improve sustainability performance.
- Inclusion of participatory methods during the composite indicator building. The selection and weighting of indicators can be achieved through the engagement of airport stakeholders or relevant experts. The composite indicator system created through participatory methods can be cross validated with that created through pure mathematical models and thus it can reduce the subjectivity and reflect the perspectives of all stakeholders. Meanwhile, participatory methods can facilitate decision-makers better understand the values, priorities, and concerns of different groups, thereby enhancing the legitimacy and effectiveness of policies.
- Diagnose the effectiveness of prior or ongoing airport sustainability policies. The effectiveness of an airport sustainability policy can be reflected by the sustainability performance of the airport at two different time points, such as before and after the policy is implemented. However, sustainable composite indicators do not provide detailed information on the specific factors that contribute to or inhibit sustainability performance. We can supplement this by using both the mathematical and participatory approaches mentioned in the first two points.

5. Conclusion

Although the term *"airport sustainability"* is increasingly mentioned by aviation policymakers and researchers, its definition is still contested, and no mature methodology has been formed to quantify it.

Conceptually, this study proposes a holistic definition of airport sustainability established on the basis of uniformity and specificity. On top of that, airport sustainability is not a constant value and may change in response to social and technological developments. The concrete initiatives for airport sustainability also vary from airport to airport, as they are subject to diverse internal dynamics and external factors. However, the development direction towards sustainability is consistent among all airports - to foster coherent developments on social responsibility, operational efficiency, economic feasibility, and environmental friendliness.

Methodologically, the composite indicator approach was applied to benchmark the airport's sustainability performance across four dimensions. In this study, the sustainability indicators were first aggregated into sustainability sub-indexes and then into the overall composite index. To explore the issue of subjectivity in composite indicator building, two different methods were used in each of the normalization, weighting, and aggregation processes. This resulted in eight combination schemes to construct the ASEI. In the end, the sensitivity and reliability analyses were performed to investigate the quantitative contribution of the three composite indicator building processes to the final results and the reliability amongst the eight combination schemes.

This study used Schiphol as an apply case to validate the ASEI system with its data from 2012 to 2021 as inputs. According to the results, there is a consensus among the eight combination schemes in identifying the outstanding (the year 2019) and bottom (the year 2012, 2014, and 2016) performers across the analyzed ten years. However, a high level of variation was observed in ranking the medium performers. The Covid-19 also has a sharp impact on Schiphol's sustainability performance by reshaping its sustainability pattern from business- and operations-oriented in 2019 to environment- and society-oriented in 2020.

The variance-based sensitivity analysis identified weighting as the most influential process in composite indicator building, followed by aggregation. Normalization made the least contribution to the result variance. The eight combination schemes also revealed an excellent internal consistency judging from the Cronbach's alpha test. Specifically, the DBN and RBN schemes had the most significant impact on reliability with respect to all schemes. In addition, the ASR indicated that the scheme RBN is consistently situated close to the median position throughout the analysis period. Therefore, the scheme with the most significant contribution to the result reliability found in this paper is RBN, which denotes re-scaling as the normalization method, BoD as the weighting method, and NCMC as the aggregation method.

As a forerunner in introducing the composite indicator approach to airport sustainability evaluation, this study demonstrates various aspects of sustainability issues and employs alternative normalization-weighting-aggregation schemes to build the airport sustainability index. The study also fulfils the content in sensitivity and reliability analyses of composite indicators, emphasizing the variations in the final results arising from different combinations of normalization, weighting, and aggregation methods.

In practice, the ASEI proposed in this study provides a holistic framework for sustainability evaluation, assisting airport operators to identify their sustainability status and trend regarding social responsibility, economic feasibility, operational efficiency, and environmental friendliness. The ASEI can also serve as a guideline methodology for sustainability assessment within the air transport sector (e.g., airlines and air cargo) and beyond (e.g., railway stations). In advancing this study, two complementary paths will be pursued. First, methodological refinement will be undertaken by integrating participatory methods such as Budget Allocation (BAL) and Multi-Criteria Decision Analysis (MCDA) into the weighting process. This inclusion will serve to augment our methodology and establish a mechanism for cross-validation against the pure mathematical methods utilized in this paper. Simultaneously, a distinct line of inquiry will be dedicated to the practical transposition of these refined methodologies. This endeavor will result in a separate publication delineating the transition from theoretical constructs to practical applications suitable for diverse airport environments. Moreover, the geographical ambit of the study will be extended to incorporate representative airports on both regional and global scales, thereby enabling cross-sectional comparisons of sustainability. Through this bifurcated approach, the research aims to contribute both to the theoretical discourse on airport sustainability and to the provision of actionable strategies for the global aviation industry.

The authors would like to thank the European Union project TULIPS (Grant agreement ID: 101036996) for funding this research.

Appendix A

Table A1: Impact matrix for Dimension 1 - Social Responsibility

	Social Responsibility indicators						
	$I_{1.1}$	$I_{1.2}$	$I_{1.3}$	$I_{1.4}$	$I_{1.5}$	$I_{1.6}$	$I_{1.7}$
2012	0.174	0.000	0.000	0.909	0.000	0.914	0.414
2013	0.063	0.067	1.000	1.000	0.118	1.000	0.283
2014	0.062	0.117	0.250	0.818	0.176	0.625	0.273
2015	0.000	0.050	0.750	0.455	0.059	0.886	0.313
2016	0.111	0.050	0.313	0.091	0.294	0.825	0.116
2017	0.257	0.517	0.375	0.273	0.412	0.493	0.010
2018	0.464	0.783	0.813	0.182	0.412	0.110	0.000
2019	0.739	1.000	0.250	0.000	0.529	0.000	0.081
2020	1.000	0.300	0.813	0.455	0.706	0.643	1.000
2021	0.447	0.517	0.438	0.182	1.000	0.599	0.913
Weight	1/7	1/7	1/7	1/7	1/7	1/7	1/7

Table A2: Impact matrix for Dimension 2 - Economic Feasibility

	Economic Feasibility indicators						
	$I_{2.1}$	$I_{2.2}$	$I_{2.3}$	$I_{2.4}$	$I_{2.5}$	$I_{2.6}$	$I_{2.7}$
2012	0.697	0.598	0.759	0.021	0.000	0.826	0.786
2013	0.729	0.570	0.792	0.000	0.021	0.860	0.826
2014	0.798	0.755	0.655	0.031	0.174	0.901	0.894
2015	0.801	0.604	0.765	0.168	0.251	1.000	1.000
2016	0.790	0.658	0.622	0.173	0.009	0.909	0.918
2017	0.766	1.000	0.401	0.228	0.342	0.868	0.880
2018	0.787	0.875	0.189	0.399	0.504	0.860	0.894
2019	1.000	0.781	0.000	0.500	1.000	0.913	0.951
2020	0.000	0.000	0.681	0.854	0.870	0.000	0.000
2021	0.142	0.365	1.000	1.000	0.270	0.694	0.383
Weight	1/7	1/7	1/7	1/7	1/7	1/7	1/7

Table A3: Impact matrix for Dimension 3 - Operational Efficiency

	Operational Efficiency indicators						
	$I_{3.1}$	$I_{3.2}$	$I_{3.3}$	$I_{3.4}$	$I_{3.5}$	$I_{3.6}$	$I_{3.7}$
2012	0.133	0.592	0.721	0.583	0.645	0.862	0.571
2013	0.288	0.623	0.729	0.750	0.776	0.917	0.857
2014	0.616	0.671	0.775	0.639	0.513	0.807	0.714
2015	0.577	0.736	0.821	0.722	0.447	0.691	0.857
2016	0.710	0.841	0.924	0.722	0.224	0.661	0.714
2017	1.000	0.937	0.990	0.833	0.105	0.364	1.000
2018	0.884	0.987	1.000	0.861	0.066	0.226	0.857
2019	0.414	1.000	0.990	1.000	0.000	0.000	0.857
2020	0.000	0.000	0.000	0.556	0.789	1.000	0.000
2021	0.726	0.091	0.146	0.000	1.000	0.558	0.857
Weight	1/7	1/7	1/7	1/7	1/7	1/7	1/7

Table A4: Impact matrix for Dimension 4 - Environmental Friendliness

	Environmental Friendliness indicators						
	$I_{4.1}$	$I_{4.2}$	$I_{4.3}$	$I_{4.4}$	$I_{4.5}$	$I_{4.6}$	$I_{4.7}$
2012	0.324	0.918	0.200	0.000	0.373	0.353	0.000
2013	0.464	1.000	0.000	0.026	0.414	0.627	0.026
2014	0.686	0.445	0.400	0.158	0.000	0.000	0.158
2015	0.655	0.095	0.400	0.192	0.102	0.275	0.192
2016	0.547	0.019	0.200	0.220	0.344	0.396	0.220
2017	0.529	0.000	0.600	0.454	0.672	0.375	0.454
2018	0.118	0.052	0.400	0.803	0.766	0.659	0.803
2019	0.000	0.104	0.600	1.000	0.762	0.851	1.000
2020	1.000	0.705	1.000	0.860	1.000	0.843	0.860
2021	0.957	0.297	1.000	0.798	0.934	1.000	0.798
Weight	1/7	1/7	1/7	1/7	1/7	1/7	1/7

Table A5: Outranking matrix for Dimension 1 - Social Responsibility

	Scoreboard										Sum	Rank
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021		
2012	0	2/7	4/7	4/7	4/7	3/7	3/7	3/7	2/7	2/7	3.86	7
2013	5/7	0	5/7	6/7	5/7	4/7	4/7	4/7	3/7	3/7	5.57	2
2014	3/7	2/7	0	4/7	3/7	3/7	3/7	4/7	1/7	2/7	3.57	9
2015	3/7	1/7	3/7	0	5/7	4/7	3/7	4/7	2/7	3/7	4.00	6
2016	3/7	2/7	4/7	2/7	0	2/7	2/7	4/7	1/7	1/7	3.00	10
2017	4/7	3/7	4/7	3/7	5/7	0	4/7	3/7	1/7	2/7	4.14	5
2018	4/7	3/7	4/7	4/7	5/7	3/7	0	3/7	2/7	4/7	4.57	4
2019	4/7	3/7	3/7	3/7	3/7	4/7	4/7	0	1/7	2/7	3.86	8
2020	5/7	4/7	6/7	5/7	6/7	6/7	5/7	6/7	0	5/7	6.86	1
2021	5/7	4/7	5/7	4/7	6/7	5/7	3/7	5/7	2/7	0	5.57	3

Table A6: Outranking matrix for Dimension 2 - Economic Feasibility

	Scoreboard										Sum	Rank
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021		
2012	0	2/7	1/7	0	1/7	1/7	1/7	1/7	5/7	4/7	2.29	10
2013	5/7	0	1/7	1/7	2/7	1/7	2/7	1/7	5/7	4/7	3.14	8
2014	6/7	6/7	0	1/7	4/7	4/7	4/7	1/7	4/7	4/7	4.86	5
2015	1	6/7	6/7	0	5/7	4/7	4/7	3/7	5/7	4/7	6.29	2
2016	6/7	5/7	3/7	2/7	0	4/7	4/7	1/7	4/7	4/7	4.71	6
2017	6/7	6/7	3/7	3/7	3/7	0	3/7	2/7	4/7	5/7	5.00	3
2018	6/7	5/7	3/7	3/7	3/7	4/7	0	2/7	4/7	5/7	5.00	4
2019	6/7	6/7	6/7	4/7	6/7	5/7	5/7	0	5/7	5/7	6.86	1
2020	2/7	2/7	3/7	2/7	3/7	3/7	3/7	2/7	0	1	3.86	7
2021	3/7	3/7	3/7	3/7	3/7	2/7	2/7	2/7	0	0	3.00	9

Table A7: Outranking matrix for Dimension 3 - Operational Efficiency

	Scoreboard										Sum	Rank
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021		
2012	0	0	2/7	2/7	2/7	2/7	2/7	2/7	5/7	4/7	3.00	9
2013	1	0	4/7	4/7	4/7	2/7	3/7	3/7	5/7	5/7	5.29	3
2014	5/7	3/7	0	3/7	3/7	2/7	2/7	3/7	5/7	4/7	4.29	7
2015	5/7	3/7	4/7	0	4/7	2/7	3/7	4/7	5/7	5/7	5.00	4
2016	5/7	3/7	4/7	3/7	0	2/7	2/7	3/7	5/7	4/7	4.43	6
2017	5/7	5/7	5/7	5/7	5/7	0	4/7	4/7	5/7	5/7	6.14	1
2018	5/7	4/7	5/7	4/7	5/7	3/7	0	5/7	5/7	5/7	5.86	2
2019	5/7	4/7	4/7	3/7	4/7	3/7	2/7	0	5/7	4/7	4.86	5
2020	2/7	2/7	2/7	2/7	2/7	2/7	2/7	2/7	0	2/7	2.57	10
2021	3/7	2/7	3/7	2/7	3/7	2/7	2/7	3/7	5/7	0	3.57	8

Table A8: Outranking matrix for Dimension 4 - Environmental Friendliness

	Scoreboard										Sum	Rank
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021		
2012	0	1/7	3/7	3/7	3/7	1/7	2/7	2/7	1/7	1/7	2.43	10
2013	6/7	0	3/7	3/7	3/7	2/7	2/7	2/7	1/7	1/7	3.29	6
2014	4/7	4/7	0	3/7	3/7	2/7	3/7	2/7	0	1/7	3.14	7
2015	4/7	4/7	4/7	0	3/7	2/7	3/7	1/7	0	0	3.00	8
2016	4/7	4/7	4/7	4/7	0	3/7	1/7	1/7	0	0	3.00	9
2017	6/7	5/7	5/7	5/7	4/7	0	2/7	2/7	0	0	4.14	5
2018	5/7	5/7	4/7	4/7	6/7	5/7	0	2/7	0	2/7	4.71	4
2019	5/7	5/7	5/7	6/7	6/7	5/7	5/7	0	3/7	2/7	6.00	3
2020	6/7	6/7	1	1	1	1	1	4/7	0	6/7	8.14	1
2021	6/7	6/7	6/7	1	1	1	5/7	5/7	1/7	0	7.14	2

Table A9: Impact matrix for the second level of aggregation

	Ranking of airport sustainability sub-index			
	D1	D2	D3	D4
2012	7	10	9	10
2013	2	8	3	6
2014	9	5	7	7
2015	6	2	4	8
2016	10	6	6	9
2017	5	3	1	5
2018	4	4	2	4
2019	8	1	5	3
2020	1	7	10	1
2021	3	9	8	2
Weight	1/4	1/4	1/4	1/4

Table A10: Outranking matrix for airport sustainability composite index

	Scoreboard										Sum	Rank	CI
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021			
2012	0	0	1/4	0	1/4	0	0	1/4	1/4	0	1.00	10	0.11
2013	1	0	3/4	3/4	3/4	1/4	1/4	1/2	1/4	3/4	5.25	4	0.58
2014	3/4	1/4	0	1/4	3/4	0	0	0	1/2	1/2	3.00	8	0.33
2015	1	1/4	3/4	0	1	1/4	1/4	1/2	1/2	1/2	5.00	6	0.56
2016	3/4	1/4	1/4	0	0	0	0	0	1/2	1/2	2.25	9	0.25
2017	1	3/4	1	3/4	1	0	1/2	1/2	1/2	1/2	6.50	1	0.72
2018	1	3/4	1	3/4	1	1/2	0	1/2	1/2	1/2	6.50	2	0.72
2019	3/4	1/2	1	1/2	1	1/2	1/2	0	1/2	1/2	5.75	3	0.64
2020	3/4	3/4	1/2	1/2	1/2	1/2	1/2	1/2	0	3/4	5.25	5	0.58
2021	1	1/4	1/2	1/2	1/2	1/2	1/2	1/2	1/4	0	4.50	7	0.50

Reference

ACI EUROPE. (2022). ACI EUROPE Sustainability Strategy - In Action. Retrieved March 23, 2022, from https://www.aci-europe.org/industry-topics/industry-topics/28-airport-sustainability.html

ACI-NA (2010). Airport Sustainability: A Holistic Approach to Effective Airport Management. Retrieved March 23, 2022, from https://crp.trb.org/acrpwebresource4/airport-sustainability-a-holistic-approach-to-effective-airport-management/

Adelaide Airport. (2019). Retrieved March 23, 2022, from https://www.adelaideairport.com.au/corporate/environment/

Adler, N., Ülkü, T., & Yazhemsky, E. (2013). Small regional airport sustainability: Lessons from benchmarking. Journal of Air Transport Management, 33, 22-31. https://doi.org/10.1016/j.jairtraman.2013.06.007

AECOM. (2020). Clean energy and infrastructure: Pathway to airport sustainability. Retrieved March 23, 2022, from https://www.cefc.com.au/media/402343/cefc-pathway-to-airport-sustainability.pdf.

Amaeshi, K. M., & Crane, A. (2006). Stakeholder engagement: a mechanism for sustainable aviation. Corporate social responsibility and environmental management, 13(5), 245-260. https://doi.org/10.1002/csr.108

Azapagic, A. (2004). Developing a framework for sustainable development indicators for the mining and minerals industry. Journal of cleaner production, 12(6), 639-662. https://doi.org/10.1016/S0959-6526(03)00075-1

Banister, D., Stead, D., Steen, P., & Nijkamp, P. (2000). European Transport Policy and Sustainable Mobility. https://doi.org/10.4324/9780203857816

Baxter, G., Srisaeng, P., & Wild, G. (2018). An Assessment of Airport Sustainability, Part 2—Energy Management at Copenhagen Airport. Resources, 7, 32. https://doi.org/10.3390/resources7020032

Baxter, G., Srisaeng, P., & Wild, G. (2019). An Assessment of Airport Sustainability: Part 3—Water Management at Copenhagen Airport. Resources. https://doi.org/10.3390/resources8030135

Baxter, G., Srisaeng, P., & Wild, G. (2020). The Use of Aviation Biofuels as an Airport Environmental Sustainability Measure: The Case of Oslo Gardermoen Airport. https://doi.org/10.14311/MAD.2020.01.01

Black, W.R. (1996). Sustainable transportation: a US perspective. Journal of Transport Geography, 4, 151-159. https://doi.org/10.1016/0966-6923(96)00020-8

Blancas, F. J., Lozano-Oyola, M., González, M., & Caballero, R. (2016). Sustainable tourism composite indicators: A dynamic evaluation to manage changes in sustainability. Journal of Sustainable Tourism, 24(10), 1403-1424. https://doi.org/10.1080/09669582.2015.1122014

Brundtland, G.H. and Khalid, M. (1987). World Commission on Environment and Development: Our Common Future. Oxford University Press, New York. https://doi.org/10.1002/jid.3380010208

Carlucci, F., Cirà, A., & Coccorese, P. (2018). Measuring and explaining airport efficiency and sustainability: Evidence from Italy. Sustainability, 10(2), 400. https://doi.org/10.3390/su10020400

Červinka, M. (2017). Small regional airport performance and Low cost carrier operations. Transportation Research Procedia, 28, 51-58. https://doi.org/10.1016/j.trpro.2017.12.168

Chao, C. C., Lirn, T. C., & Lin, H. C. (2017). Indicators and evaluation model for analyzing environmental protection performance of airports. Journal of Air Transport Management, 63, 61-70.

Charles, V., Aparicio, J., & Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. European Journal of Operational Research, 279(3), 929-940. https://doi.org/10.1016/j.ejor.2019.06.025

Cherchye L., and Kuosmanen T., (2002), Benchmarking sustainable development: a synthetic meta-index approach, EconWPA Working Papers.

Cherchye, L., Moesen, W. and Van Puyenbroeck, T., (2004), "Legitimately Diverse, yet Comparable: on Synthesizing Social Inclusion Performance in the EU", Journal of Common Market Studies 42, 919-955. https://doi.org/10.1111/j.0021-9886.2004.00535.x

Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T (2008) Benefit of the doubt composite indicators. Workpackage 5. Deliverable 5.3. KEI-project supported by European Commission by funding from the Sixth Framework Programme for Research

Ciegis, R., Kliucininkas, L., & Ramanauskiene, J. (2011). Assessment of state and tendencies of sustainable development in Lithuania. Management of Environmental Quality: An International Journal, 22(6), 757-768. https://doi.org/10.1108/14777831111170858

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. Omega, 44, 1-4. https://doi.org/10.1016/j.omega.2013.09.004

Copenhagen Airports. (2019). Copenhagen Airports Group Annual Report 2018 - Take off for a sustainable future. 1904-2558 https://cph-prod-cdn.azureedge.net/495470/globalassets/8.-om-cph/6.-investor/arsrapporter/2018/cph-annual-report-2018_uk_web.pdf

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334. https://doi.org/10.1007/BF02310555

De Montis, A., Serra, V., Ganciu, A., & Ledda, A. (2020). Assessing landscape fragmentation: A composite indicator. Sustainability, 12(22), 9632. https://doi.org/10.3390/su12229632

Devuyst, D. (2000). Linking impact assessment and sustainable development at the local level: the introduction of sustainability assessment systems. Sustainable development, 8(2), 67-78. https://doi.org/10.1002/(SICI)1099-1719(200005)8:2<67::AID-SD131>3.0.CO;2-X

Dimitriou, D., & Karagkouni, A. (2022). Assortment of Airports' Sustainability Strategy: A Comprehensiveness Analysis Framework. Sustainability, 14(7), 4217. https://doi.org/10.3390/su14074217

Dimitriou, D., & Sartzetaki, M. F. (2020). Social dimensions of aviation on sustainable development. Sustainable Aviation: Greening the Flight Path, 173-191. https://doi.org/10.1007/978-3-030-28661-3_9

Dizdaroglu, D., & Yigitcanlar, T. (2016). Integrating urban ecosystem sustainability assessment into policy-making: Insights from the Gold Coast City. Journal of environmental planning and management, 59(11), 1982-2006. https://doi.org/10.1080/09640568.2015.1103211

Dobos, I., & Vörösmarty, G. (2014). Green supplier selection and evaluation using DEA-type composite indicators. International Journal of Production Economics, 157, 273-278. https://doi.org/10.1016/j.ijpe.2014.09.026

Elkington, J. (1999). Triple bottom-line reporting: Looking for balance. Australian CPA, 69, 18- 21.

European Commission. (2021). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on ensuring a level playing field for sustainable air transport. Brussel https://ec.europa.eu/info/sites/default/files/refueleu_aviation_-_sustainable_aviation_fuels.pdf

European Commission. (2020). Sustainable and Smart Mobility Strategy – putting European transport on track for the future. COM(2020) 789. Brussel https://www.eumonitor.eu/9353000/1/j4nvke1fm2yd1u0_j9vvik7m1c3gyxp/vlehqjsi6jzv/v=s7z/f=/com(2020)789_en.pdf

European Commission. (2001). Communication from the Commission: A Sustainable Europe for a Better World: A European Union Strategy for 'Sustainable Development'. Brussel.

European Commission - DG ENTR (2001) European Innovation Scoreboard, Brussels.

Eurostat. (2020). How many people work in Air Transport? How many people work in air transport? Products Eurostat News - Eurostat. Retrieved March 22, 2022, from https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200525-1

FAA. (2022). Airport Sustainability – Airports. Federal Aviation Administration (FAA), Washington, DC, USA

Ferrulli, P. (2016). Green airport design evaluation (GrADE)–methods and tools improving infrastructure planning. Transportation Research Procedia, 14, 3781-3790.

Flouris, T.G., & Yılmaz, A.K. (2011). Risk Management and Corporate Sustainability in Aviation. https://doi.org/10.4324/9781315606507

Gasparatos, A., El-Haram, M.A., & Horner, M. (2008). A critical review of reductionist approaches for assessing the progress towards sustainability. Environmental Impact Assessment Review, 28, 286-311. https://doi.org/10.1016/j.eiar.2007.09.002

Global Reporting Initiative (2011). Sustainability Reporting Guidelines & Airport Operators Sector Supplement; Global Reporting Initiative: Amsterdam, The Netherlands.

Gómez-Limón, J. A., & Sanchez-Fernandez, G. (2010). Empirical evaluation of agricultural sustainability using composite indicators. Ecological economics, 69(5), 1062-1075. https://doi.org/10.1016/j.ecolecon.2009.11.027

Graham, A., & Halpern, N. (2018). The Routledge Companion to Air Transport Management. https://doi.org/10.4324/9781315630540

Graver, B., Rutherford, D., & Zheng, S. (2020). CO_2 Emissions From Commercial Aviation: 2013, 2018, and 2019.

Gu, Y. (2019). Defining Airport Operational Sustainability for the U.S. General Aviation Airports (Version 1). Purdue University Graduate School.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. American journal of Evaluation, 24(3), 293-314. https://doi.org/10.1177/109821400302400302

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52, 1-17. https://doi.org/10.1016/0951-8320(96)00002-6

Hudrliková, L. (2013). Composite indicators as a useful tool for international comparison: The Europe 2020 example. Prague economic papers, 22(4), 459-473.

Janic, M. (2007). The Sustainability of Air Transportation: A Quantitative Analysis and Assessment. https://doi.org/10.4324/9781315236889
Janic, M. (2010). Developing an indicator system for monitoring, analyzing, and assessing airport sustainability. European Journal of Transport and Infrastructure Research, 10(3).
Kaya, S. K., & Erginel, N. (2020). Futuristic airport: A sustainable airport design by integrating hesitant fuzzy SWARA and hesitant fuzzy sustainable quality function deployment. Journal of Cleaner Production, 275, 123880. https://doi.org/10.1016/j.jclepro.2020.123880
Kılkış, Ş., & Kılkış, Ş. (2016). Benchmarking airports based on a sustainability ranking index. Journal of Cleaner Production, 130, 248-259. https://doi.org/10.1016/j.jclepro.2015.09.031
Knudsen, F. B. (2002). Defining sustainability in the aviation sector. development, 25, 83.
Krajnc, D., & Glavič, P. (2005). A model for integrated assessment of sustainable development. Resources, conservation and recycling, 43(2), 189-208. https://doi.org/10.1016/j.resconrec.2004.06.002
Kucukvar, M., Alawi, K. A., Abdella, G. M., Bulak, M. E., Onat, N. C., Bulu, M., & Yalçıntaş, M. (2021). A frontier-based managerial approach for relative sustainability performance assessment of the world's airports. Sustainable Development, 29(1), 89-107. https://doi.org/10.1002/sd.2134
Kumar, A., Aswin, A., & Gupta, H. (2020). Evaluating green performance of the airports using hybrid BWM and VIKOR methodology. Tourism Management, 76, 103941. https://doi.org/10.1016/j.tourman.2019.06.016
Lee, T. H., Jan, F. H., & Liu, J. T. (2021). Developing an indicator framework for assessing sustainable tourism: Evidence from a Taiwan ecological resort. Ecological indicators, 125, 107596. https://doi.org/10.1016/j.ecolind.2021.107596
Liew, D. Y. C., Che Ros, F., & Harun, A. N. (2019). Developing composite indicators for flood vulnerability assessment: Effect of weight and aggregation techniques. Int. J. Adv. Trends Comput. Sci. Eng, 8, 383-392.
Li, L., & Loo, B. P. (2016). Impact analysis of airport infrastructure within a sustainability framework: Case studies on Hong Kong International Airport. International Journal of Sustainable Transportation, 10(9), 781-793. https://doi.org/10.1080/15568318.2016.1149647
López, M. V., Garcia, A., & Rodriguez, L. (2007). Sustainable development and corporate performance: A study based on the Dow Jones sustainability index. Journal of business ethics, 75, 285-300. https://doi.org/10.1007/s10551-006-9253-8
Lubk, C. (2017). The Concept of Sustainability and Its Application to Labor Market Policy.
Lu, M. T., Hsu, C. C., Liou, J. J., & Lo, H. W. (2018). A hybrid MCDM and sustainability-balanced scorecard model to establish sustainable performance evaluation for international airports. Journal of Air Transport Management, 71, 9-19. https://doi.org/10.1016/j.jairtraman.2018.05.008
Martin-Nagle, R., & Klauber, A. (2015). Lessons learned from airport sustainability plans. ACRP Synthesis 66. Transportation Research Board of the National Academies, Washington, D.C. DOI: 10.17226/22111.
Melbourne Airport. (2019). Environment at Melbourne Airport. Retrieved March 23, 2022, from https://www.melbourneairport.com.au/Corporate/About-us/Environment/Environment-at-Melbourne-Airport
Melyn W. and Moesen W. (1991): Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available, Public Economics Research Paper 17, CES, KU Leuven.
Mo, Y., Zhao, Y., Li, S., & Liu, X. (2018). Evaluation of Port Logistics Competitiveness Based on DEA. In IOP Conference Series: Earth and Environmental Science (Vol. 189, No. 6, p. 062041). IOP Publishing.
Mukhambetov, T., Yerdavletova, F., Kurbanova, K., Mukhametzhanova, Z., & Sadvakassova, K. (2020). Analysis of financial indicators used to assess the sustainability of companies. In E3S Web of Conferences (Vol. 208, p. 03049). EDP Sciences. https://doi.org/10.1051/e3sconf/202020803049
Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. European Comission, Ispra, 15(1), 19-20.
Niemeijer, D. (2002). Developing indicators for environmental policy: data-driven and theory-driven approaches examined by example. Environmental Science & Policy, 5(2), 91-103. https://doi.org/10.1016/S1462-9011(02)00026-6
OECD (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD: OECD Publishing.
Olfat, L., Amiri, M., Soufi, J. B., & Pishdar, M. (2016). A dynamic network efficiency measurement of airports performance considering sustainable development concept: A fuzzy dynamic network-DEA approach. Journal of Air Transport Management, 57, 272-290. https://doi.org/10.1016/j.jairtraman.2016.08.007
Orkomy, A., & Sharbatdar, M. K. (2021). Identifying Effective Sustainable Development Indicators for Airport Construction Projects: Zahedan International Airport in Iran as Case Study. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 45(1), 241-252.

Perez, V., Guerrero, F., González, M., Perez, F., & Caballero, R. (2013). Composite indicator for the assessment of sustainability: The case of Cuban nature-based tourism destinations. Ecological Indicators, 29, 316-324. https://doi.org/10.1016/j.ecolind.2012.12.027

Ramakrishnan, J., Liu, T., Yu, R., Seshadri, K., & Gou, Z. (2022). Towards greener airports: Development of an assessment framework by leveraging sustainability reports and rating tools. Environmental Impact Assessment Review, 93, 106740. https://doi.org/10.1016/j.eiar.2022.106740

RIVM. (2022). Rijksinstituut voor Volksgezondheid en Milieu. Retrieved from https://www.rivm.nl/landelijk-meetnet-luchtkwaliteit

Royal Schiphol Group. (2022). Schiphol Group's annual report. Retrieved from https://www.schiphol.nl/en/schiphol-group/page/annual-reports/

Sachs, J. D. (2012). From millennium development goals to sustainable development goals. The lancet, 379(9832), 2206-2211. https://doi.org/10.1016/S0140-6736(12)60685-0

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

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Santos, A.J., Mancini, S.D., Frutuoso Roveda, J.A., Ewbank, H., & Roveda, S.R. (2020). A fuzzy assessment method to airport waste management: A case study of Congonhas Airport. https://doi.org/10.1016/j.jairtraman.2020.101838

Sartzetaki, M. (2019). Airport enterprises management performance evaluation towards innovation and sustainable development. The International Journal of Business Management and Technology, 3(4), 108-114.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2007). Development of composite sustainability performance index for steel industry. Ecological Indicators, 7(3), 565-588. https://doi.org/10.1016/j.ecolind.2006.06.004

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

Sobol, I.M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. Ussr Computational Mathematics and Mathematical Physics, 7, 86-112. https://doi.org/10.1016/0041-5553(67)90144-9

Sobol I.M., Sensitivity estimates for nonlinear mathematical models, Math. Mod. and Comput. Exp., Vol 1, 407-414, 1993.

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. https://doi.org/10.1016/S0378-4754(00)00270-6

Somerville, A., Baxter, G., Richardson, S., & Wild, G. (2015). Sustainable water management at major Australian regional airports: the case of Mildura Airport. Aviation, 19, 83-89.

Spicka, J., Hlavsa, T., Soukupova, K., & Stolbova, M. (2019). Approaches to estimation the farm-level economic viability and sustainability in agriculture: A literature review. Agricultural Economics, 65(6), 289-297. https://doi.org/10.17221/269/2018-AGRICECON

Sreenath, S., Sudhakar, K., & Yusop, A. F. (2021). Sustainability at airports: Technologies and best practices from ASEAN countries. Journal of environmental management, 299, 113639. https://doi.org/10.1016/j.jenvman.2021.113639

Sun, Y., Garrido-Baserba, M., Molinos-Senante, M., Donikian, N. A., Poch, M., & Rosso, D. (2020). A composite indicator approach to assess the sustainability and resilience of wastewater management alternatives. Science of The Total Environment, 725, 138286. https://doi.org/10.1016/j.scitotenv.2020.138286

SYD (2020). Sustainability. Sydney Kingsford Smith Airport (SYD). Retrieved March 23, 2022, from https://www.sydneyairport.com.au/corporate/sustainability/investor-sustainability

United Nations. (2015a). Millennium Dvelopment Goals and Beyond 2015. Retrieved from https://www.un.org/millenniumgoals/

United Nations. (2015b). Sustainable development goals: 17 goals to transform our world. Retrieved 2023, from http://www.un.org/sustainabledevelopment/

United Nations Development Programme (UNDP). Human Development Index (HDI) Report 2014; UNDP: New York, NY, USA, 2014.

Upham, P. J. (2001a). A comparison of sustainability theory with UK and European airports policy and practice. Journal of environmental management, 63 3, 237-48. https://doi.org/10.1006/jema.2001.0469

Upham, P. J. (2001b). SELECTING INDICATORS FOR A DECISION SUPPORT TOOL FOR AIRPORT SUSTAINABILITY.

Upham, P. J., & Mills, J. N. (2005). Environmental and operational sustainability of airports: Core indicators and stakeholder communication. Benchmarking: An international journal.

Uysal, M., & Sogut, M.Z. (2017). An integrated research for architecture-based energy management in sustainable airports. *Energy, 140*, 1387-1397. https://doi.org/10.1016/j.energy.2017.05.199

Valovirta, V. (2002). Evaluation utilization as argumentation. Evaluation, 8(1), 60-80. https://doi.org/10.1177/1358902002008001487

Veleva, V., & Ellenbecker, M. (2001). Indicators of sustainable production: framework and methodology. Journal of cleaner production, 9(6), 519-549. https://doi.org/10.1016/S0959-6526(01)00010-5

Verbunt, P., & Rogge, N. (2018). Geometric composite indicators with compromise Benefit-of-the-Doubt weights. European Journal of Operational Research, 264(1), 388-401. https://doi.org/10.1016/j.ejor.2017.06.061

Vicario-Modroño, V., Gallardo-Cobos, R., & Sánchez-Zamora, P. (2022). Sustainability evaluation of olive oil mills in Andalusia (Spain): a study based on composite indicators. Environment, Development and Sustainability, 1-30. https://doi.org/10.1007/s10668-022-02307-5

Walker, S., & Cook, M. (2009). The contested concept of sustainable aviation. Sustainable Development, 17, 378-390. https://doi.org/10.1002/sd.400

Wan, L., Peng, Q., Wang, J., Tian, Y., & Xu, C. (2020). Evaluation of Airport Sustainability by the Synthetic Evaluation Method: A Case Study of Guangzhou Baiyun International Airport, China, from 2008 to 2017. Sustainability, 12, 3334. https://doi.org/10.3390/su12083334

Wang, C., Wang, L., & Dai, S. (2018). An indicator approach to industrial sustainability assessment: The case of China's Capital Economic Circle. Journal of Cleaner Production, 194, 473-482. https://doi.org/10.1016/j.jclepro.2018.05.125

Wang, Z., & Song, W. K. (2020). Sustainable airport development with performance evaluation forecasts: A case study of 12 Asian airports. Journal of Air Transport Management, 89, 101925.

Wong, W. P. (2021). A global search method for inputs and outputs in data envelopment analysis: Procedures and managerial perspectives. Symmetry, 13(7), 1155. https://doi.org/10.3390/sym13071155

Yangmin, B., Shaohong, F., & Yan, L. (2021). Assessing the synergy and sustainability of "Airport-Industry-City"(AIC) system in aerotropolis: Evidence from Zhengzhou Aerotropolis in China. Environmental Research, 195, 110886. https://doi.org/10.1016/j.envres.2021.110886

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

Zhou, P., Fan, L., & Zhou, D. (2010). Data aggregation in constructing composite indicators: A perspective of information loss. Expert Syst. Appl., 37, 360-365. https://doi.org/10.1016/j.eswa.2009.05.039

Zinck, J. A., Berroteran, J. L., Farshad, A., Moameni, A., Wokabi, S., & Ranst, E. V. (2004). Approaches to assessing sustainable agriculture. Journal of Sustainable Agriculture, 23(4), 87-109. https://doi.org/10.1300/J064v23n04_08