MASTER IN BUSINESS ENGINEERING PROFESSIONAL FOCUS IN DATA SCIENCE

Towards a methodology and implementation of a Composite Indicator for Air Navigation Service Provider Benchmarking

Remy, Aline

Award date:
2020

Awarding institution:
University of Namur

Towards a methodology and implementation of a Composite Indicator for Air Navigation Service Provider Benchmarking

Aline Remy

Directeur: Prof. C. Burnay

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en data science

ANNEE ACADEMIQUE 2019-2020

Université de Namur, ASBL
Faculté des Sciences économiques, sociales et de gestion – Département des Sciences de gestion

Rempart de la Vierge 8, B-5000 Namur, Belgique, Tel. +32 [0]81 72 48 41/49 58, Fax +32 [0]81 72 48 40

Acknowledgments

This master thesis is my last work as a student at the University of Namur. I was able to complete it thanks to the invaluable help provided by several people.

First of all, I would like to express my deep gratitude to my master thesis advisor, Professor Burnay, for proposing me this fascinating subject, for his support and feedback. He gave me the opportunity to discover the world of ANSP and to meet experts in the field during my internship at skeyes.

A big thank you to my internship supervisor, Bertrand Gallez, for his help with air traffic control acronyms, for his patience with all my questions and for his advice on who to contact.

This brings me to thank Jean-Claude Hustache, Economic Expert at the Performance Review Unit, who took the time to send me the data and his comments indispensable for me to undertake my analyses.

Finally, I would also like to thank my friends, my sister and, in particular, my parents who believed in me and motivated me from the very beginning. During the Great Quarantine, their words of encouragement gave me the strength to finalize my master thesis.

Table of content

Chapter 1: Introduction

1.1 Context

In 2017, more than 1 billion of passengers took a flight for business or for vacation in the European Union (Eurostat, 2018). This demand has been increasing over recent years. The safety of those travellers is ensured by the Air Navigation Service Providers (ANSP) which role is to “manage flight traffic on behalf of a company, a region or a country” (EUROCONTROL, 2020). The main mission of an ANSP is to deliver services for Air Traffic Management. This mainly consists in ensuring a safety distance between aircrafts, maximising capacity at the airports, optimising the flow of aircrafts and organising the airspace to meet the demand.

Each ANSP controls a predefined area. The *Figure 1-1* shows how Europe is fragmented according to the airspace managed by individual ANSP. The ANSP, which controls the air movements from 7500 and 24500 feet above Belgium and Luxembourg, is called since 2018, “skeyes”.

Figure 1-1: European Flight Information Regions (FIR) (Source: EUROCONTROL)

1.1.1 skeyes

Briefly, Belgocontrol (former name of skeyes) was created in 1998 with the aim of resuming the airspace management activities provided by the “Régie des Voies Aériennes”. As this company pursues an objective of public order, it has a particular status: “autonomous public enterprise”. It means that the State owns 100% of the shares of this enterprise but it conducts its operations independently. This autonomy has given it the flexibility to better meet the airspace users’ needs. In 2005, the main site of Belgocontrol was moved to Steenokkerzeel, just next to the runways at Brussels-Zaventem airport. Since 2010, Belgium has been part of a Functional Airspace Block (FAB) called FABEC allowing the optimisation of services in collaboration with the BENELUX countries, Germany, France and Switzerland. In 2018, Belgocontrol took advantage of its 20th anniversary to change its name by becoming skeyes (the eyes of the sky). skeyes is now active in 6 airports (Brussels, Ostend, Antwerp, Kortrijk, Charleroi and Liège), it managed 1 101 145 movements and had 872 employees in 2018 (skeyes, 2018).

Figure 1-2: Main facts of skeyes' history (Adapted from (skeyes, 2018))

1.1.2 Challenge

Due to the parcelling of the European airspace, pilots, who execute cross-border flights, have to communicate their flight plans to many ANSPs. To overcome this challenge, EUROCONTROL, a European Organisation for the Safety of Air Navigation, was created in 1960 to harmonize the air traffic management across Europe (skeyes, 2018) and to set up a unit, later called Performance Review Unit (PRU), in charge of evaluating the performance of European ANSPs. In relation to its main mission, EUROCONTROL has been tasked to support the implementation of the Single European Sky (SES) initiative launched in 2000 by the European Commission (SKYBRARY, 2019). The SES legislation spells out the responsibilities of the 38 ANSPs. It can be summarized as handling air traffic safely, efficiently, environmentally friendly, cost-effectively and in compliance with the regulations of the state where those ANSPs operate.

1.2 Research motivation

In line with the missions highlighted by SES initiative, the ANSPs have to improve their performance to provide good quality services, to reduce the costs and by extension, to decrease the charges requested by the ANSPs to the airline companies. To reach those objectives, it is necessary for each ANSP to evaluate its market position regarding the others. This means to conduct benchmarking analysis in order to identify the best practices among the most efficient ANSPs and to replicate them into others' processes.

Since 2003, financial performance benchmarking analysis has been subject to an international study released annually by the PRU. Through those annual reports, the PRU follows the evolution of many indicators including the “Financial Cost-Effectiveness Indicator”. This measure is computed from the data submitted by the ANSP members in compliance with the Decision N88 of the Permanent Commission of EUROCONTROL (Performance Review Unit, 2019).

Nevertheless, as mentioned by G.Nero & S. Portet (2007) and by PRU in the ACE Benchmarking Report (2019), this analysis only states facts without considering the elements outside the ANSPs’ control which influence their performance. The PRU identified a set of exogenous factors such as the size of the territory managed by the ANSP, the traffic complexity or the weather. For instance, the bigger the territory controlled by the ANSP is, the more it needs manpower, workstations and surveillance devices on the area to control the air movements. This situation leads to increasing costs. As a consequence, those aspects should be taken into account to carry out a fair comparison between ANSPs in terms of cost management.

1.3 Academic Motivation

1.3.1 Problem statement

The common financial indicators are, for example, Operating Cash Flow (total amount of money generated by the enterprise), Net Profit Margin (Net Profit/Revenue), … Those kinds of measures include at most two dimensions: money and outputs produced. The companies use them to have an overview of their performance and also to implement specific strategies with the aim of enhancing their overall process. A way to come up with new ideas for improvements is to conduct a benchmarking analysis. It implies to compare the enterprise’s performance with the others belonging to the same industry through a reference indicator, called benchmark. The use of regular financial indicators in the context of benchmarking is meaningful only under the hypothesis that those organisations operate in the same context. This is not the case for the ANSPs that manage different types of airspace. Indeed, they need to adapt their investments depending on those particularities (size, complexity, …) and on the regulations established by each country. Consequently, other dimensions should be considered to build a measurement that enables fair comparisons between ANSPs.

As defined by the Organisation for Economic Co-operation and Development (OECD), the indicator that “measures multi-dimensional concepts which cannot be captured by a single indicator” is called Composite Indicators (CI). They differ from the regular ones, quoted previously, in terms of complexity because they allow the combination of different units of measurement into a single index.

1.3.2 Contribution

The main objective of this master thesis is to build and implement a new financial "Complex Performance Indicator" through the use of Data Envelopment Analysis (DEA). This DEA is a non-parametric method based on the construction of a "best-practice" frontier used as a benchmark. Firstly, it will aim to help ANSPs to position themselves relative to the others in terms of costs performance. Secondly, it will attempt to meet the non-functional requirement of equity by considering the exogenous factors affecting the costs.

Therefore, the contribution will focus on two poles. On the one hand, from a theoretical point of view, we will complete a broad-based methodology conceived by OECD (the Statistics Directorate and the Directorate for Science, Technology and Industry) and JRC (Econometrics and Applied Statistics Unit of the Joint Research Centre established by the European Commission) with tools specifically used for DEA. On the other hand, from a practical perspective, we will apply this newly constructed framework to the case related to ANSPs.

1.4 Approach

Before developing the methodology and designing the CI, the concepts related to the airspace management will be clearly defined (*Chapter 2*). Then, we will position this master thesis into the literature by introducing the notions on performance management, indicators as well as fairness (*Chapter 3*) and by explaining the measurements currently used in the airspace management field (*Chapter 4*).

Then, the methodology of OECD and JRC will be introduced (*Chapter* 5), adapted to better suit this particular case study (*Chapter 6*) and then implemented (*Chapter 7*).

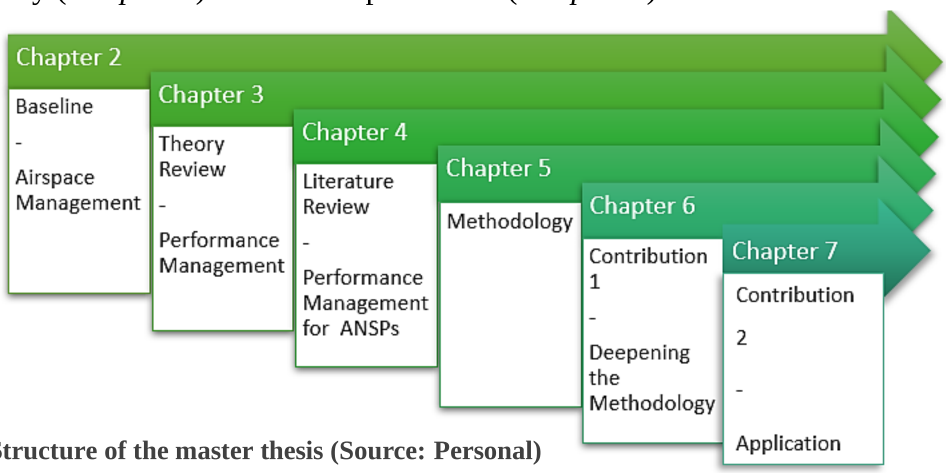
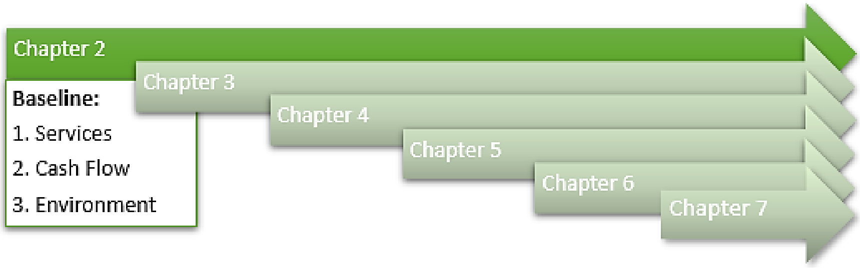


Figure 1-3: Structure of the master thesis (Source: Personal)

Chapter 2: Baseline

Given the complexity of the aviation field and the frequent use of acronyms (see *Appendix 1*), this section aims to provide a basic background to understand the role of an ANSP, the structure of the cash flow movements and also the characteristics of the airspace environment.



2.1 Services

The services provided by ANSPs are grouped in five categories:

1. Meteorological services for air navigation (MET)
2. Aeronautical Information Services (AIS)
3. Search and Rescue (SAR)
4. Air Traffic Management (ATM)
5. Communication, Navigation and Surveillance systems (CNS)

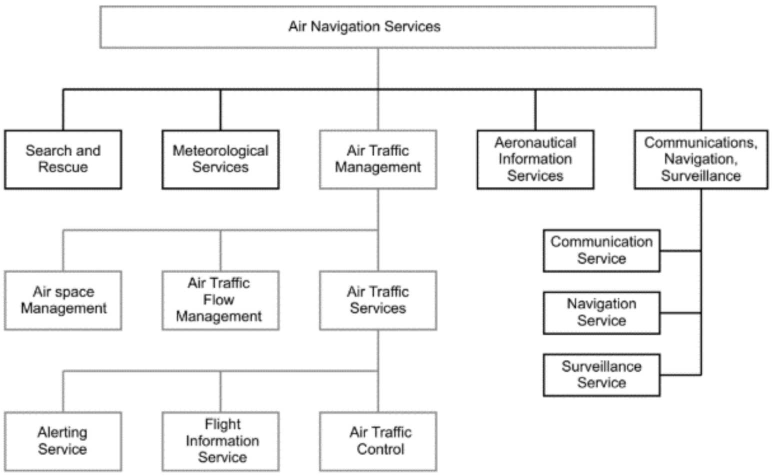


Figure 2-1: Overview of Air Navigation Services (Source: Arblaster, 2018)

The first three services are not always offered by ANSPs. The MET and AIS categories concern the transmission of meteorological information and aeronautical data. The third one, SAR, is related to the assistance of aircrafts in need.

The two main services proposed by all ANSPs are **ATM and CNS** which are related to the management of the flow of aircrafts in a predefined area. The ATM is composed of three functions: Air Traffic Control (ATC), Air Traffic Flow Management (ATFM) and Air Space Management (ASM). The aim of ATC is to ensure a safety distance between aircrafts to avoid collisions of planes during the take-off, the landing or the cruise phases of the flights. The ATFM focuses on maximising the capacity at the airports and optimizing the flow of aircrafts to reduce delays. Finally, the ASM aims to organise the airspace efficiently to meet the demand. In a complementary manner, the CNS systems support the ATM services. They are employed to allow effective communication between the ground-based staff and the pilots and to follow the position of aircrafts to guide the pilots safely through the airspace.

2.1.1 Type of ATM services and CNS

The types of ATM services delivered and CNS systems employed depend on the phase of a flight and the category of airspace (Arblaster, 2018).

The flight from one airport to another can be broken down into five steps: the take-off, the ascent, the cruise, the descent and the landing. Different stakeholders will interact at each stage to ensure airspace safety and efficient air flow.

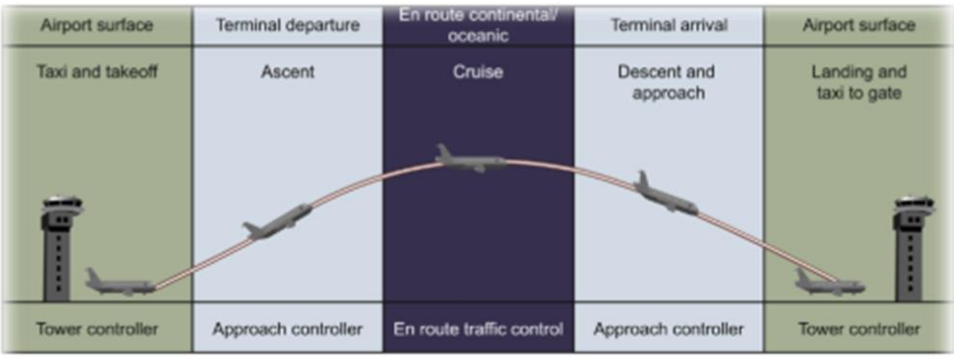


Figure 2-2: Phases of a flight requiring different types of air traffic control (Source: Arblaster, 2018)

2.1.1.1 Take off

To avoid collisions of aircrafts on the taxiways (routes between the gates and the runways) and the runways, the tower controllers (Tower ATCO) give instructions to the pilots to move their aircraft. At first, the pilots tune in to a specific radio frequency to communicate with the controllers and ask for a clearance to leave the parking area (Wright, 2013). Those controllers are located in a control tower usually based at the airport itself to conduct movements on the ground visually. Then, the pilots wait for instructions to take the taxiways and, afterwards, take off when they receive the permission.

2.1.1.2 The ascent

When the plane takes off, the tower controllers assign to the pilots the radio frequency of the terminal controllers (APP ATCO). They will give indications for maintaining separation between aircrafts during the ascent phase and for leading the pilots to its air corridor. Those airways can be seen as highways that aircrafts follow to join the destination airport.

This approach service is provided to manage the flow of aircrafts in the terminal area. To identify the position of airplanes in this area, the approach controllers use surveillance technologies such as radars because they are not able to do it visually unlike the tower controllers.

Once the plane reaches an altitude of 4500 feet (equivalent to 1371.6 meters), the terminal controllers instruct the pilots to communicate with an en-route controller by changing the radio frequency.

2.1.1.3 The cruise

The ascent and the descent phases are called **terminal navigation** and the cruise stage is part of the **en-route navigation**. The missions of en-route controllers are to insert the plane in the air traffic flow and to avoid conflicts by maintaining a separation distance vertically (5 nautical miles) and horizontally (1000 feet) between aircrafts. "Horizontally" means that two aircrafts at the same altitude have to keep a safety distance and "Vertically" concerns the separation between aircrafts flying at different altitudes (Arblaster, 2018). If this security zone is violated, there is an "**interaction**" which might lead to an accident.

At this stage, the ATM services will depend on the altitude at which the aircrafts fly. The airspace is composed of two layers: the lower and the upper altitudes. In general, the two levels are controlled by the same ANSP but for some countries, they are handled by different service providers. For example, in Belgium, the lower layer is managed by skeyes and the upper one is under the control of MUAC (Maastricht Upper Area Control Centre) (skeyes, 2018). Therefore, the intensities of en-route and terminal activities are different for skeyes and MUAC.

Each level of altitude is divided into sectors which can be combined or separated to handle the fluctuation of the demand (Arblaster, 2018). Those smaller air volumes are usually under the supervision of a team of two Air Traffic Control Officers (ATCO): a radar controller, responsible for the immediate airspace management, and a planner controller, who anticipates and resolves congestion. To achieve their missions, the controllers have at their disposal radar screens which display the position, the speed, the altitude and the direction of each aircraft.

Furthermore, the ATCO keep contact with the pilots through communication devices such as radio or text messages. When the aircrafts go outside the area under their control, they transmit data related to the flights to the neighbouring ANSP and assign a new radio frequency to the pilots. The efficiency of an ANSP has thus consequences on the ANSPs around (Neiva, 2014).

2.1.1.4 The descent

When the pilots plan to leave the airways and to begin their descent operation, they change their radio frequency to communicate with the terminal controllers. They organise the sequence of aircrafts ready to land. Once the pilots receive permission, they start a step-down operation.

2.1.1.5 The landing

Below a certain altitude, the terminal controllers give way to the tower controllers who assist the pilots for the landing and guide them to the gate where the passengers can leave the plane.

2.1.2 Services provided in Belgium

In Belgium, skeyes counts five control towers situated at the main airports which are Antwerp, Charleroi, Liège, Oostende and Brussels. The terminal area is managed by four Approach Control Units (APP) based in Brussels, Charleroi, Liege and Oostende. Furthermore, the Area Control Centre (ACC), known under the name of CANAC 2, handles the movements of aircrafts from 4500 ft to 24 500 ft. This Belgian lower airspace is subdivided into 7 sectors and those sections can be combined depending on the density of planes to control.

Finally, the particularity of skeyes is that the upper airspace above 24 500 feet is managed by another entity, MUAC. Since 1975, the territory controlled by MUAC extends over Belgium, Luxembourg, Netherland and north-west Germany. This non-profit organisation is regulated by those four States but belongs to EUROCONTROL, a European Organisation for the Safety of Air Navigation mainly created to harmonize air traffic management across Europe (skeyes, 2018). This consolidation of ATM and the sectorisation of the upper airspace allow a more efficient and smoother flow of aircrafts.

Beside the airspace management, skeyes provides meteorological data (MET) and aeronautical information (AIS) needed by the pilots to take off.

2.2 Cash Flow Structure

Like all companies, the ANSPs need revenues to cover the costs.

2.2.1 Revenues

The primary source of incomes comes from the charges asked by ANSPs to the users of Air Navigation Services (ANS) (Performance Review Unit, 2006). Those tariffs can be classified according to the services offered: en-route or terminal services. The en-route charge is computed from the multiplication of three elements: a unit rate, the aircraft weighted factor and a distance factor (Cogen, 2016). The unit rate depends on the charging zone, which has a single cost base and unit rate, spanned by the flights. In the same way, the terminal charges are calculated from the aircraft weighted factor and unit rate of the terminal (Castelli & Ranieri, 2007).

The other revenues are oceanic en-route charges, payments from Governments and charges from other services provided.

2.2.2 Costs

Regarding the expenses, the ANS incur costs belonging to five categories (Performance Review Unit, 2006):

1. MET
2. Payments to governmental or regulatory authorities
3. EUROCONTROL costs
4. Payments to other ANSPs for delegated services
5. ATM/CNS provision costs

The first four types of expenses can be considered as fixed costs. On the one hand, the MET costs depend if the ANSP provides the service internally or relies on national institutions to have access to meteorological data. On the other hand, the three other kinds of fees (2,3,4) are established by other organisms and, therefore, they are outside the control of the ANSPs.

Only the ATM/CNS provision costs can be influenced by actions undertaken by ANSPs because they are directly linked to the way they manage the air traffic flow. This last category gathers all expenditures related to the staff costs, non-staff operating costs, capital-related costs (depreciation and investments in capital) and exceptional items.

2.3 Environment

The ANSPs operate in different socio-economic, operational conditions and have to comply with various regulatory frameworks.

The socio-economic conditions gather all the concepts associated with the cost of living, market wage rate, working hours, retirement age, social security and pensions which vary from one ANSP to another.

The operational conditions concern the characteristics of the airspace managed by the ANSPs such as the size, the traffic variability, the type of airspace (oceanic or continental), the weather and the traffic complexity. The latter consists of three components:

1. **Density of traffic**: distribution of traffic throughout the airspace.
2. **Structural complexity**: number of interactions between the planes. The interactions might be of three types, as depicted in Table 2-1.
3. **Traffic Mix:** of types of aircrafts.

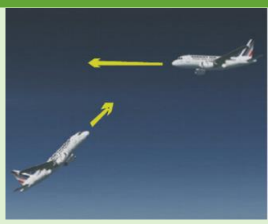
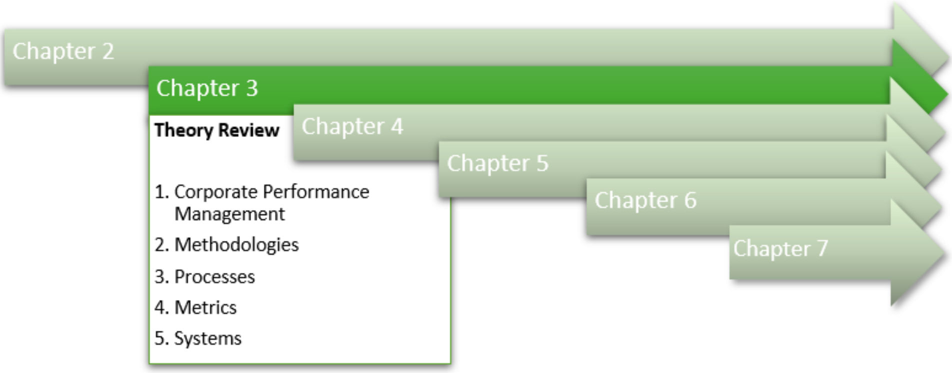
Horizontal	Vertical	Speed
		
Different headings	Different flight phases (cruise, climbing, descending)	Same headings, different speed.

Table 2-1: Type of interactions (Source: Performance Review Unit)

Finally, the ANSPs have different sorts of ownership (private, public or autonomous public) and are subject to special rules established by the national government or other institutions such as EUROCONTROL. For example, in Brussels, skeyes has to respect the "Preferential Runway System" that indicates which runways have to be used according to the time of the day. If the ATCOs need to employ another runway, they need to justify this change.

Chapter 3: Theory Review

The aim of this master thesis is to produce a Composite Indicator. To reach this objective, many theoretical concepts will be mobilized through the steps. They must be specified to put them to good use. Developing indicators is only one key element that helps the managers to improve their business' performance. They are created in the context of Corporate Performance Management.



3.1 Corporate Performance Management

As mentioned by Taouab & Issor (2019), the definition of performance has evolved through the years and depends on personal perceptions. Thus, a general specification of this term might be given by Collins dictionary in which the performance is defined as "how successful the companies are or how well they do something". All the **methodologies**, the **processes**, the **metrics** and **systems** employed to track the company's success are gathered under the concept of "Corporate Performance Management" (CPM) (Gartner, 2019), also referenced as Business (BPM) or Enterprise Performance Management (EPM). It aims to set goals in line with the company's strategies, to develop indicators trailing the completion of those goals and to take actions accordingly. This approach allows, on the one hand, for employees to understand the objectives they have to pursue and, on the other hand, for managers to focus their attention on specific business areas to improve, called Key Performance Areas (KPA).

3.2 Methodologies

The methodologies, developed in the field of CPM, provide frameworks to represent the organisation according to some KPAs, to structure the objectives in compliance with the strategy and finally, to enhance the global performance by following an improvement process. Plenty of methodologies have been created and focus on different aspects of the entity. Among the most famous ones, we might quote the Balanced Scorecard and the European Foundation for Quality Management Excellence Model.

3.2.1 Balanced Scorecard

The Balanced Scorecard (BSC) is a strategic management performance metric introduced by Kaplan and Norton in the early 90s. This methodology has evolved to better match the complexity of an organisation. However, its foundations are still the same. The BSC enables managers to have a global view of their business by examining it from at least four perspectives: Financial, Customers, Internal Business Process and Learning & Growth (Balanced Scorecard Institute, 2020; Kaplan & Norton, 2005). The management process associated with the BSC consists in formulating the strategy, defining strategic objectives, understanding the cause and effect chain between them through a complementary tool, the Strategic Map, and fixing the targets to achieve. For each objective, a measurement system needs to be put in place to track their completion and to take initiatives to reach the targets. Finally, the managers must stand back from details to observe the evolution of their business.

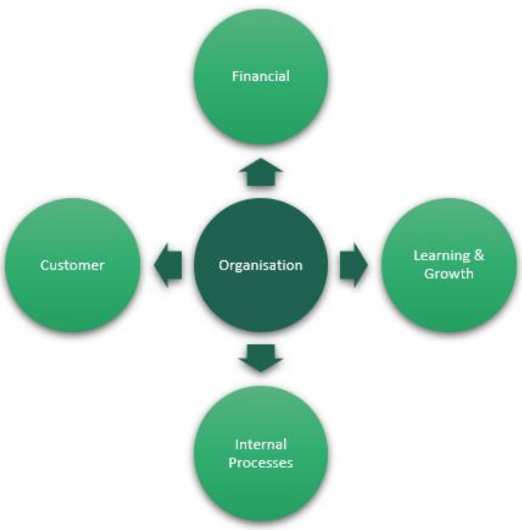


Figure 3-1: BSC perspectives (Adapted from Kaplan and Norton 2005)

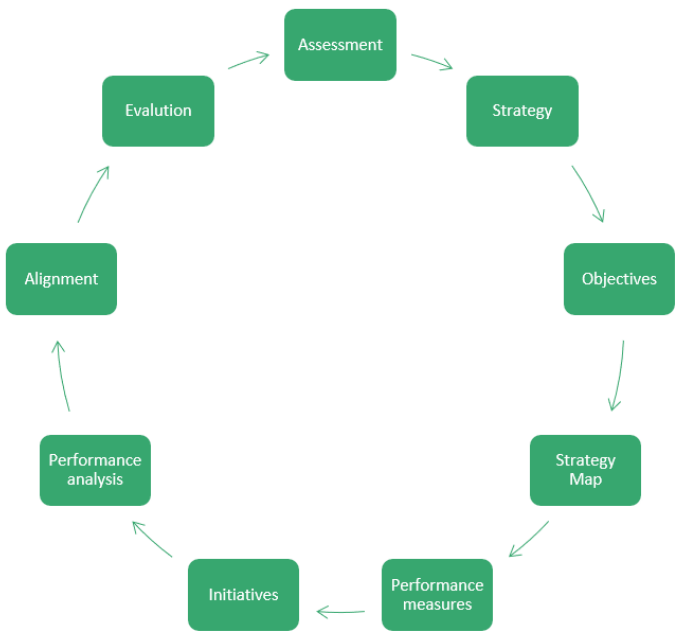


Figure 3-2: Performance Process (Adapted from Balanced Scorecard Institute 2020)

3.2.2 European Foundation for Quality Management Excellence Model

In contrary with the BSC, the European Foundation for Quality Management (EFQM) Excellence Model was created in 1988 initially as self-assessment quality tool before becoming a management tool. It is seen as a “cause and effect diagrams” (EFQM, 2018) which relies on nine criteria split into two categories: the Enablers which are related to how the organisation manages its operations and the Results which represent what was achieved. This model switched to a management framework, inter alia, with the integration of the RADAR logic made up of four broad-based steps: Determine the objectives to reach, plan approaches to achieve the targets, deploy those approaches and assess the results achieved.

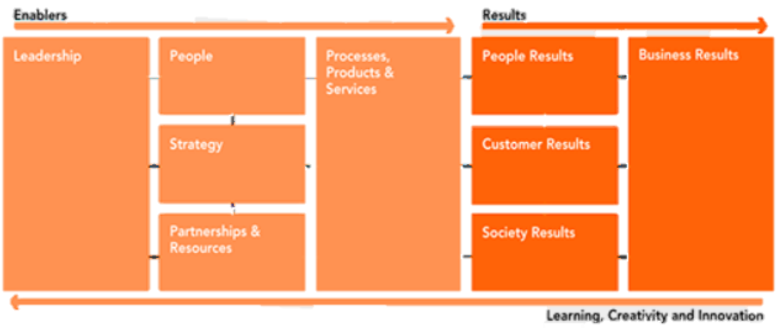


Figure 3-4: Nine criteria (Source: EFQM Company)

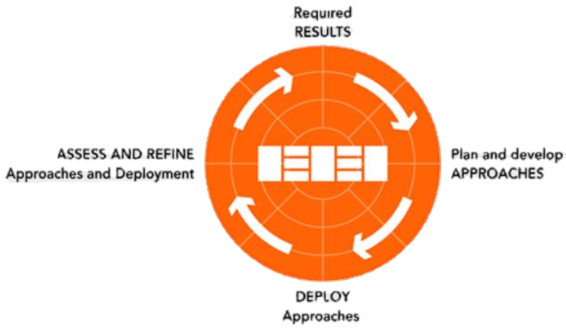


Figure 3-3: RADAR Logic (Source: EFQM Company)

3.3 Processes

Each methodology encompasses an improvement cycle composed of many steps such as the "Performance Process" in the BSC. However, those mechanisms only specify the global phases which must be executed by the managers through a series of actions. For instance, formulating the strategy is an analytical process which implies the observation of the organisational environment, the implementation of a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis, … All those processes hidden in the methodology are also part of the CPM put in place by the managers.

3.4 Metrics

One of the most important processes is the development of indicators which enable managers to follow the evolution of their business. An indicator might be generally defined as a quantitative or a qualitative measure that indicates the status of the company in a particular area (Barone, et al., 2011). Those measurements might be categorised according to different criteria such as the part of the process they refer to (Input/Process/Output/Outcome), their relationship with critical aspects of the organisation and with the temporality (Result/Performance), their utilization for comparison (Benchmark) and their level of complexity (CI).

3.4.1 Input/Process/Output/Outcome indicators

Every company is created for the purpose of producing products or providing services. To do so, they need resources and processes to generate the outputs. Thus, indicators might be computed for each stage to identify the area on which the managers should focus their efforts.

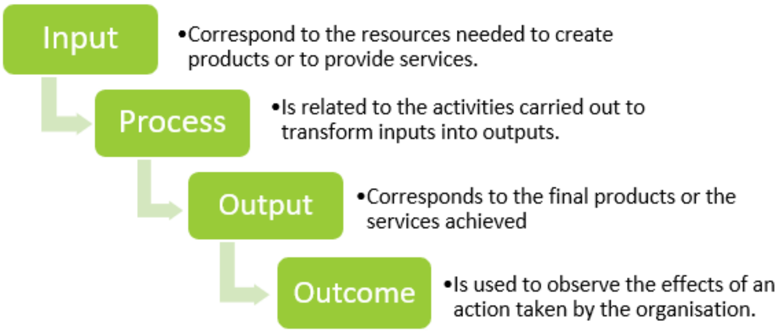


Figure 3-5: Types of indicators as part of a process (Source: Personal)

For instance, the World Health Organisation (2014) has proposed many indicators in the context of "Child health and development" program:

- *Input*: Funds needed to conduct a training course.
- *Process*: Number of training courses conducted.
- *Output:* Number of medical assistants trained.
- *Outcome:* Proportion of sick children correctly managed by the trained medical assistants.

3.4.2 Performance and Results Indicators

Another classification has been introduced by D. Parmenter (2015) who established four types of performance measures grouped into two classes: *result indicators* and *performance indicators.*

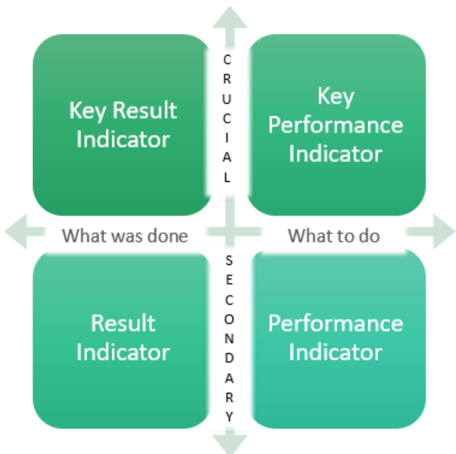


Figure 3-6: Type of performance measures (Source: Personal)

3.4.2.1 Result indicators

The *result indicators* reflect what was achieved as a "result" of actions undertaken by many teams. There is a distinction to be made between the Key Result Indicators (KRI) and the Result Indicators (RI). The RIs give only a quick overview of how certain teams combine their efforts to produce a result. However, the KRIs are more useful for the board because it provides an "overall summary of how the organisation is performing" (Parmenter, 2015). It indicates if the business steps in the right direction. For instance, the financial metrics, such as the net profit before tax or the Return On Capital, give an idea of how the business thrives. Nevertheless, by looking at those measurements, the managers are not capable of putting their finger on the sources of non-performance issues because the *result indicators* are not tied to a specific group of workers.

3.4.2.2 Performance indicators

On contrary, the *performance indicators* help managers to pinpoint the areas to analyse and improve by identifying the group responsible for bad performance. Therefore, they are used to align the teams' work with the strategy. In parallel with the *result indicators*, the Performance Indicators (PI) are judged less fundamental than the Key Performance Indicators (KPI). Indeed, these latter are related to the organisation's critical success factors, considered as crucial problems or aspects for the current and future company's wellbeing. In light of their importance, the KPIs should be simple, non-financial, frequently measured, constantly followed by the CEO and attached to a team. In a nutshell, those indicators allow managers to evaluate the company's success at reaching their goals.

3.4.3 Benchmark

Other indicators, called benchmarks, are developed with the aim of assessing a company's performance relative to others. They are implemented in the context of benchmarking. As defined by Andersen (1996), benchmarking is "a process of continuously measuring and comparing one's business processes against comparable processes in leading organisations to obtain information that will help the organisation to identify and implement improvements".

3.4.3.1 Types of Benchmarking

More generally, the benchmarking does not only focus on competitors' performance. Indeed, there exist different kinds of benchmarking depending on the angle chosen and also the object of comparison.

The possible angles to analyse are the process, the strategy or the performance. The first two types, process and strategy, allow the company, on the one hand, to compare methods used by best companies and, on the other hand, to understand the choices of the strategy of the others with the aim of refining their own long-term vision. The third sort of benchmarking concerns the performance measures which help to position the company relative to the others.

The objects of comparison can be internal, functional, generic and competitive. The two opposite benchmarking analyses are the internal and the competitive ones. The former aims to examine indicators or processes internally, between departments, countries or units. The latter puts emphasis on the comparison between the company itself and its direct competitors.

Thus, the performance competitive benchmarking allows constructing a ranking from a benchmark which emphasizes the performance gaps between those organisations. However, benchmarking does not only consist in creating benchmarks. It implies also other steps to understand the sources of deviation and to set feasible goals to achieve better performance levels.

3.4.3.2 Benchmarking process

The benchmarking process is made up of five phases: planification, analysis, integration, action and maturity (O'Rourke, 2012).

Figure 3-7: Benchmarking process (Adapted from (O'Rourke, 2012))

The planning stage implies to specify the subject on which the company will be compared, to determine the criterion of comparison, to choose the benchmarking partners, to define the performance measures and to collect data for their computation. The benchmarking differs from a simple competitive analysis by integrating a second step. This involves further researches to understand the deviation of the performance indicators between the organisations, to identify the best practices among the leaders and to predict future performance levels. The conclusions drawn from the analysis must be communicated to the employees to gain their acceptance of the coming changes. After informing the organisational members, the managers must derive concrete operational objectives from the global ones. Furthermore, actions must be taken to achieve those targets. Then, measurements must be set up to follow the signs of progress as well as the narrowing of performance gaps. Finally, the maturity phase is reached when the firm has attained the leader position. However, to maintain this ranking situation, the company should continue this process to further improve.

3.4.4 Composite indicators

In the benchmarking analysis, the object of comparison might grow in complexity by its multidimensionality. For instance, due to the wake-up call on global warming, the sustainability of corporate is not limited anymore to the capital provided to the owners. Two other pillars, the social and the ecological perspectives (Purvis, et al., 2019), must be taken into account to evaluate the ability of the company to continue their activities on the long term. This concept is made up of three dimensions which cannot be captured by a simple indicator such a count, a rate, a proportion or a ratio. It implies the development of a more complex measurement called Composite Indicators (CI).

3.4.4.1 Description

As defined by the Organisation for Economic Co-operation and Development (OECD), the CI is “formed when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured”. It has a relative value which means that it can only be interpreted through a comparison between entities. This type of indicator might be conceived according to different thematises and also for various kinds of organisations.

For instance, at the country level, the Technology Achievement Index (TAI) is a CI, created by the United Nations in 2001, which assesses the performance of countries regarding technological achievements. This index aggregates 8 indicators related to four dimensions: “Technologies Creation, Diffusion of Old Innovation, Diffusion of Recent Innovations and the Human Skills” (Incekara, et al., 2017).

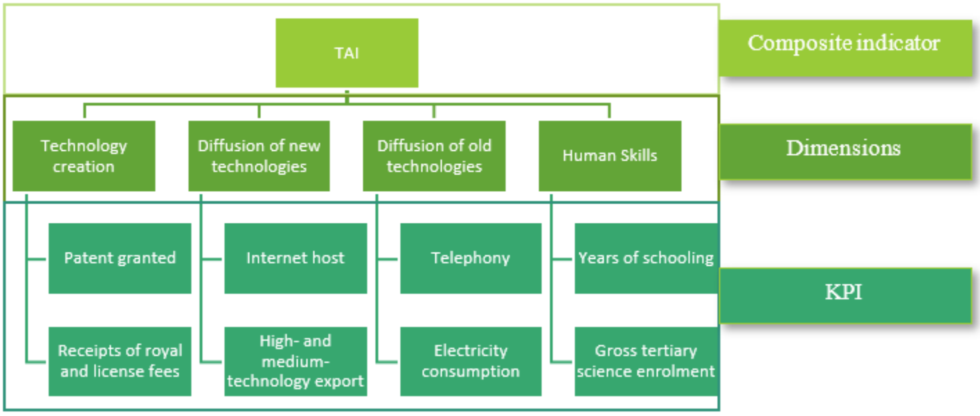


Figure 3-8: Structure of the Technology Achievement Index (TAI) (Source: Personal)

At the corporate level, a Complex Performance Indicator was proposed by Avláková Docekalová & Kocmanová (2016) to evaluate the corporate's sustainability. To build this indicator, the authors selected KPIs for each dimension (environmental, social, economic and corporate governance), assigned them weights and aggregated them into a single index.

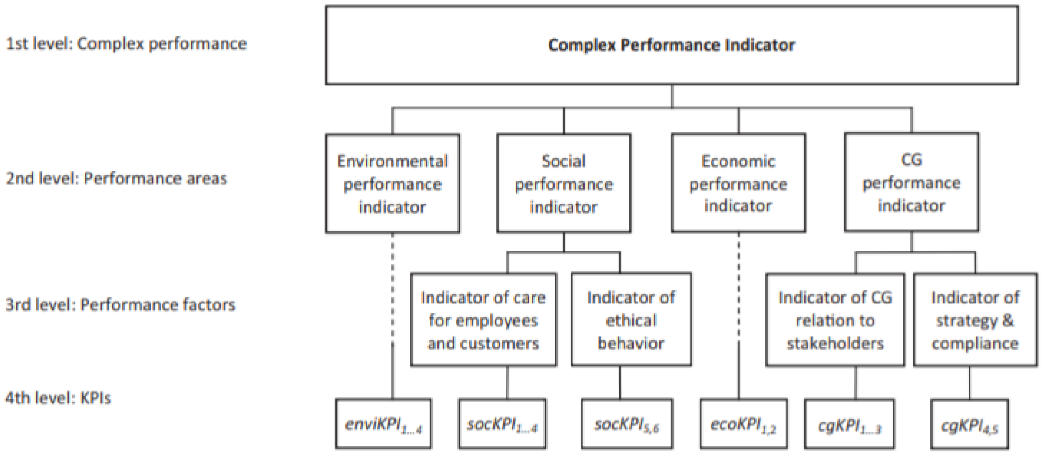


Figure 3-9: Structure of the Complex Performance Indicator (Source: Docekalová & Kocmanová, 2016)

3.4.4.2 Weighting and aggregating methods

Plenty of aggregating and weighting methods exist to construct the CI (Gan, et al., 2017; European Commission, 2019). The weighting tools enable to compute weights for the indicators which are combined through the aggregating technics into a single index.

In absence of adequate model, the equal weighting approach is usually applied. To avoid the problem of double-counting due to the correlation between variables, other methodologies might be employed based on statistical models or on public opinion.

Weighting methods	
Equal weights	The most famous technic consists in assigning the same weight to all the variables. It implies that those factors equally contribute to the general concept. This results from a lack of knowledge of the correct model or of the causal relationships between the variables and the main object to be measured. With high correlated indicators, it leads to a double-counting issue.
Weights assigned through statistical models	
Principal Components Analysis (PCA)	It allows to reduce "the dimensionality of the variable space by representing it with a few uncorrelated variables that capture most of its variability" (Tauler, et al., 2009). The index is thus based on statistical dimensions created from the combination of correlated indicators.
Factor Analysis (FA)	It is similar to PCA because it relies also on the reduction of the dimensionality by means of factors. The difference with the PCA lies in the construction of those statistical dimensions.
Data Envelopment Analysis (DEA)	It is a non-parametric method based on the estimation of an efficiency frontier used as a benchmark (Cooper, et al., 2007). A score is computed from the distance between the benchmark and the organisation's performance.
Regression Analysis	It is a multivariate technique which tends to estimate the relationship between the explanatory indicators and the dependent variable.
Unobserved Components Models (UCM)	The premise is that the indicators are imperfect signals of unobserved variables. The UCM aims to extract the information on the unobserved components from the indicators and conceive the best possible index.
Weights based on public/ expert opinion	
Budget allocation	The experts have at their disposal N points that they have to allocate to indicators in order to determine the weights. This step must be done multiple times to reach a consensus.
Public Opinion	The stakeholders participate by giving their preferences. The level of concern among the public opinion will determine the weights for each indicator.
Analytic Hierarchy Process (AHP)	It is used in the context of multi-attribute decision making. It relies on the decomposition of a complex problem into a structure of goals, criteria and attributes. Then, the experts compare those elements per pair and derive the relative importance of each variable.
Conjoint Analysis	It implies the evaluation of respondents to a set of alternative scenarios.

Table 3-1: Most commonly used methods for weighting (Adapted from Gan, et al. 2017)

Amongst the aggregating techniques, the most commonly applied are the additive ones. However, their use implies conditions which cannot be always met. Therefore, other methods, such as the geometric aggregation and the non-compensatory aggregations, have been developed to overcome those difficulties.

Aggregating methods	
Additive methods	They consist in linearly adding up weighted and normalised indicators. However, they are effective under two conditions. Firstly, they assume that there is no synergy between the indicators. Secondly, they are compensatory which means that a poor result for one indicator in the index can be compensated by a high performance from another. $\omega_1 I_1 + \omega_2 I_2 + \dots + \omega_m I_m = \sum_{i=1}^{m} \omega_i I_i$ **Figure 3-3-10: Additive methods (Source: Gan, et al., 2017)**
Geometric aggregation	Those technics are based on multiplicative aggregations such as the weighted geometric mean. In comparison with the additive methods, the geometric aggregations rely on a less strong condition because it is not a full compensatory approach. $I_1^{\omega_1} I_2^{\omega_2} \dots I_m^{\omega_m} = \prod_{i=1}^{m} I_i^{\omega_i}$ **Figure 3-3-11: Geometric aggregation (Source: Gan, et al. 2017)**
Non compensatory multi-criteria analysis	This type of methodology allows the interpretation of weights as "importance coefficient" thanks to the non-compensatory aggregation procedure which avoids the substitution between indicators. It is based on the construction of a ranking matrix where each indicator "votes" for a country and defines a ranking.

Table 3-2: Most commonly used methods for aggregating (Adapted from Gan, et al. 2017)

3.4.4.3 Pros and Cons

Given that the choices of the weighting and the aggregating technics are mostly subjective, some analysts are reluctant to use CI. Therefore, it is important to be aware of the strengths and the weaknesses of CI and to take them into account through the building process.

The main advantage of this type of indicator is the aggregation of many measures into a single one. It allows capturing the multi-dimensional characteristic of a concept. This specific measurement, thus, provides a "big picture" that eases the communication with the stakeholders, helps organisations to position themselves relative to the others and helps to identify the main areas to focus on (Joint Research Centre-European Commission, 2008).

Concerning the weaknesses, the construction of a CI firstly implies the judgment for some stages such as for the identification of indicators, the attribution of weights, the methods, … The errors due to this subjectivity in the process may lead to erroneous "big picture" on which the stakeholders will base their decisions. To mitigate this drawback, the choices made must be clearly specified, they should also rely on statistical principles as much as possible and sensitivity and robustness analyses should be carried out. Secondly, this indicator should be supported by sub-indicators to avoid stakeholders making simplistic conclusions.

3.4.4.4 Equity

The utilization of sub-indicators is useful to understand the performance gaps between the entities. Those deviations might be due to external factors on which the managers do not have control and which should be considered to avoid comparing apples with oranges. Usually, the comparability of the CI is taken into account thanks to the normalisation of the indicators which allows space comparisons (OECD, 2015, p. 75). However, normalisation is not always sufficient to undertake a cross-sectional benchmarking analysis because it assumes that the individuals operate under the same environmental conditions. Ideally, the CI should integrate those contextual variables to equitably compare the organisations' performance and to derive realistic targets. This notion of "equity", generally defined as "the quality of being fair" (Collins dictionary) by giving an equal treatment, has never been really treated in the context of the conception of indicators. This question deserves to be addressed but it is unfortunately not part of the scope of this master thesis.

3.5 Systems

Those indicators, once determined and specified, are created thanks to technologies which enable to turn unstructured data into valuable information. As defined by Gartner (2019), all the applications, the infrastructures and the tools, that allow the creation and the analysis of information to support the decision-making procedure, are gathered under the term "Business Intelligence". In other words, "From a technical point of view, Business Intelligence refers to the process of extracting, transforming, managing and analysing business data" (Negash & Gray, 2008). In a nutshell, it requires a system for which the generic architecture is made up of four levels of components shown in the *Figure 3-12*.

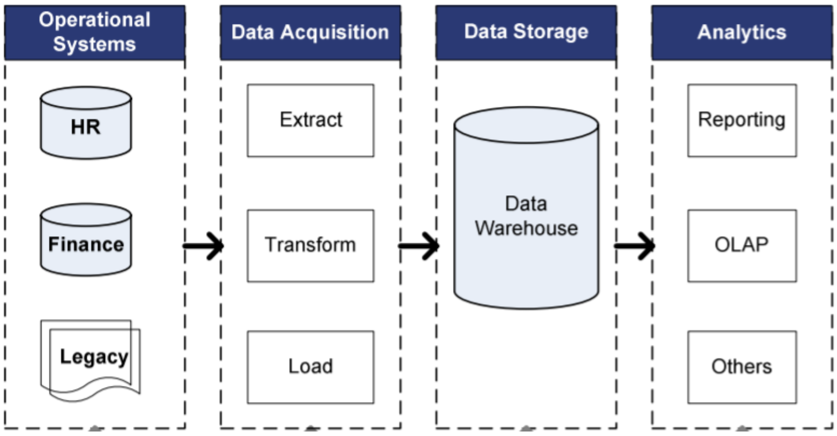


Figure 3-12: The General Architecture of Business Intelligence Systems (Source: Negash & Gray, 2008)

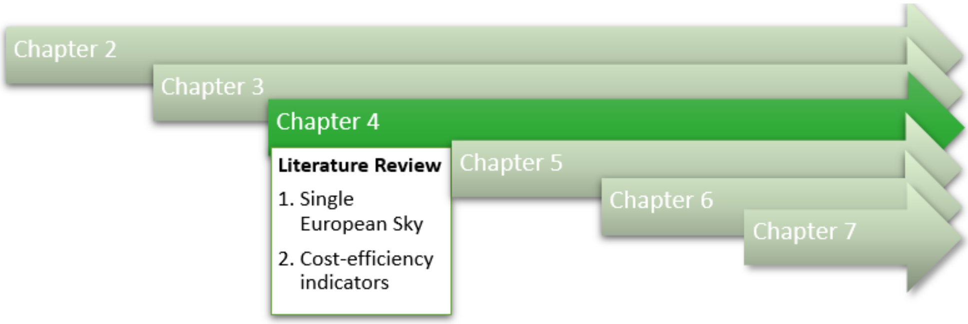
The "Operational Systems" layer corresponds to all the systems used for daily operations. Then, through the Extract-Transform-Load (ETL) process, data are retrieved from those structures, reorganised and pushed into a Data Warehouse. As defined by W.H. Inmon (2005), the Data Warehouse is a collection of data which is:

- **Subject oriented**: it is organised according to major subject areas of the organisation (i.e. customers, products, …).
- **Integrated:** it gathers data from multiple sources.
- **Time-Variant:** it keeps historical data and grows progressively.
- **Non-volatile:** Once the data are loaded, they stay and should not be modified.

Finally, this data structure is manipulated by many reporting applications to generate, inter alia, dashboards showing the indicators. Of course, the Business Intelligence is more complex than the briefing note provided above but our problem is independent of the architecture employed.

Chapter 4: Literature Review

Now that the theoretical concepts have been well established, we are more able to analyse their utilisation in the context of ANSPs. To grasp the importance of performance in the airspace management, it is necessary to go back in time until the adoption of the Single European Sky initiative in 2000.



4.1 Single European Sky

In response to challenges arising from the traffic growth, the European Commission (EC) launched the Single European Sky (SES) initiative in 2000. The overall plan is to enhance the collaboration between European ANSPs by integrating airspace management and introducing new regulations for ATM. Roughly speaking, the objectives of the SES initiative are to improve the safety of air transport in Europe, reduce delays, decrease costs by harmonizing airspace management, and better integrate the military systems into ATM systems (SKYBRARY, 2019). For this purpose, the EC adopted two legislative packages, one in 2004 and another in 2009, to mitigate the consequences of the fragmentation of the European airspace into small territories managed by each ANSP (European Commission, 2020). After the implementation of the first package, the capacity was still a challenge to handle, so the EC set up a second set of regulations, referred to as SES II. This reform relies on four pillars: "performance, single safety framework, new technologies and managing capacity on the ground" (Mendes De Leon & Calleja Crespo, 2011).

Concerning the performance, which is the focus of this master thesis, the SES II contributed to two major changes: the introduction of the Performance Scheme and the establishment of a Performance Review Body.

4.1.1 Performance Scheme

Before the reform, the ANSPs operated under the "full cost recovery system" which means that an increase in the costs was passed on to the airspace users (European Commission, 2018). This approach weakened the incentives for ANSPs to improve their global performance. Following this observation, the EC introduced the Performance Scheme which resulted in three improvements.

Firstly, it led to the transition from the "full cost recovery system" to the "determined cost mechanism". This new mechanism implies that the costs are established in advance for a certain time period, called Reference Period (RP – RP1: 2012-2014 and RP2: 2015-2019). If the actual costs are lower than determined, the ANSPs can keep the surplus (Performance Review Unit, 2019). Otherwise, the ANSPs have to bear the losses.

Secondly, the Performance Scheme entails the specification of PIs and KPIs which are related to four KPAs: Safety, Environment, Capacity and Cost-Efficiency. As mentioned in a report released by the EC (European Commission, 2018), "the plan was that the targets would lead to more direct routes (less fuel burn and less CO_2) and services delivered with fewer and shorter delays and in a more cost efficient manner". A sample of the KPIs used by the EC is reported in the **Table 2-1** and those indicators are accessible to the general public through the dashboards[1] implemented for each RP.

KPA	KPI	Description
Safety	Effectiveness of safety management	This score tends to capture the level of implementation and performance of the ANSPs regarding Safety.
Environment	Horizontal en-route flight efficiency	It provides a measure to compare the length of the actual flight trajectory and the shortest distance between its endpoints (EUROCONTROL, 2020).
Capacity	En-route ATFM delay per flight	Due to, for example, capacity problems in some sectors, the flights are "regulated" which means that the Network Manager determines a new slot to take off in order to regulate the flow. ATFM delay is "the duration between the last take-off time requested by the aircraft operator and the take-off slot allocated by the Network Manager following a regulation" (EUROCONTROL, 2016).
Cost Efficiency	Determined Unit Cost for en-route	For the Cost Efficiency, a comparison is carried out between the determined and the actual costs for each ANSP.

Table 4-1: Sample of KPI used in each KPA (Source: Personal)

[1] Available on https://www.eurocontrol.int/prudata/dashboard/vis/2019/

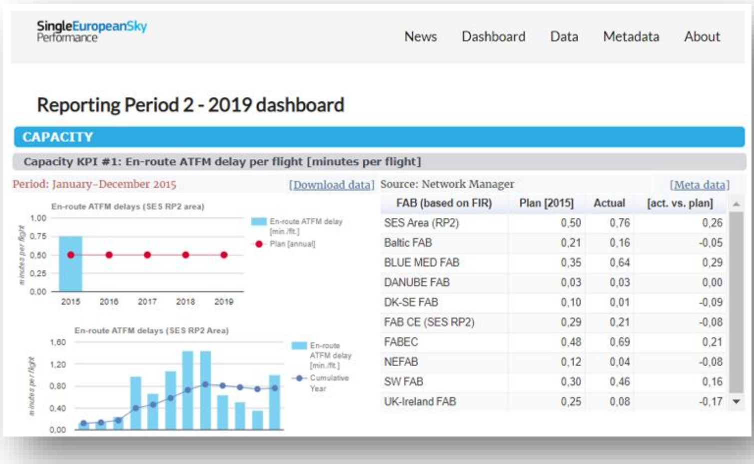


Figure 4-1: Dashboards (Source: PRU)

Lastly, the Performance Scheme includes also a "periodic review, monitoring and benchmarking of the performance of air navigation services and network functions" (EUROCONTROL, 2016).

4.1.2 Performance Review Body

The organisation, in charge of the performance monitoring and which advices the EC in the setting of targets, is called the Performance Review Body (PRB). This independent group of experts is appointed by the EC to support the implementation of the SES (PRC, 2018; PRB, 2020; European Commission, 2020). The Performance Review Commission (PRC), created in 1998 by EUROCONTROL, deals with complementary tasks and is backed up by the Performance Review Unit (PRU). To fulfil its mission, the PRU collects data amongst European ANSPs on an annual basis and assesses their performance. The analyses conducted by the PRU and the PRC are summarized in two reports: "Performance Review Report" (PRR) and "ATM Cost-Effectiveness Benchmarking Report" (ACE).

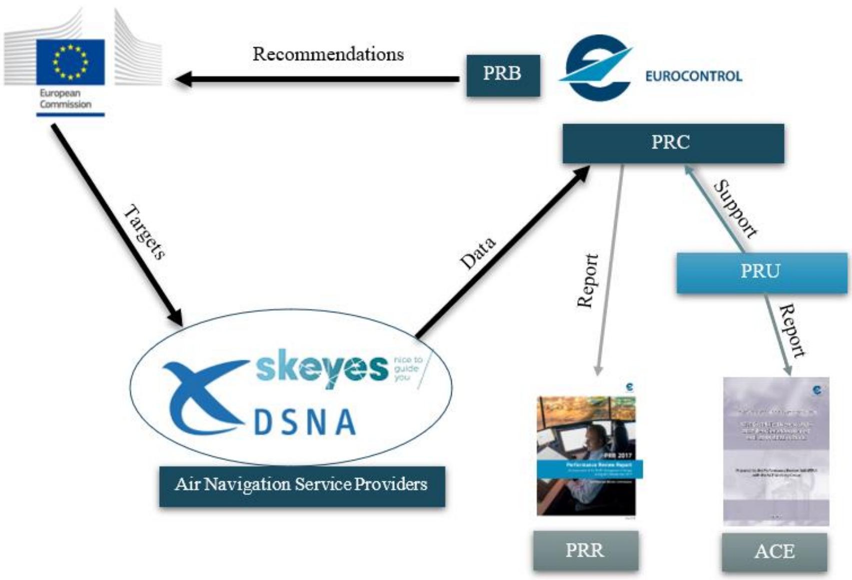


Figure 4-2: Performance Process (Source: Personal)

4.2 Cost-efficiency indicators

The KPA addressed by both reports is related to the cost management. The PRC and the PRU seem to measure the same concept by linking the inputs – in this case, the costs – to different measures of outputs. However, the notion is sometimes mentioned as "Cost-efficiency" (in PRR and on the EUROCONTROL website) or "Cost-effectiveness" (in ACE Report). Therefore, it seems essential to clarify the distinction between those two ideas which are frequently used interchangeably.

4.2.1 Efficiency vs Effectiveness

As mentioned by Asmild, et al. (2007), "the **efficiency** is often associated with performing activities as well as possible or "doing the things right" whereas the **effectiveness** is equated with the proper selection of the activities or "doing the right things"".

An entity is thus considered as cost-efficient if it produces a certain amount of outputs at minimal costs. Nevertheless, cost-efficiency does not imply cost-effectiveness. Indeed, an organisation might be cost-efficient but not effective when it manages well the costs, but it pursues the wrong goals.

The focus here is to measure how well the ANSPs manage their costs, in other words, how well they perform their activities at minimal costs. By consequent, "cost-efficiency" seems to be the most accurate term to employ throughout the master thesis. This notion is computed in several ways in the ATM field depending on the type of costs, the output variables, the quality variables as well as the technics (simple ratio or CI) used.

4.2.2 Simple indicator: Ratio

The cost-efficiency links the resources to the production. Therefore, the indicators monitored in the PRR and the ACE are simple ratios of costs and outputs.

4.2.2.1 Performance Review Report

In the PRR, two indicators were developed by dividing the costs, related to the en-route or terminal phase of the flight, by a result corresponding to an outcome measure.

Name	**Formula**	**Description**
En-route cost-efficiency performance	$\frac{en-route\ costs}{Total\ Service\ Units}$	• $\underline{En-route\ costs:}$ Costs generated by the en-route services provided.

		• *Total Service Units*: are computed for each flight and depend on two factors: the distance flown and the Maximum Take-Off Weight[2].
Terminal cost-efficiency performance	$\frac{Terminal\ ANS\ cost}{Terminal\ Navigation\ Service\ Units}$	• *Terminal ANS cost*: Costs generated by the terminal services provided. • *Terminal Navigation Service Units (TNSU)* depend only on the Maximum Take-Off Weight.

Table 4-2: Cost-efficiency indicators introduced by the PRR (Adapted from (PRC, 2018))

It is important to know that the Service Units (for en-route and terminal) are essentially computed to determine the charges requested from the airspace users for the services provided by the ANSPs (see *Chapter 2*). This role clarifies the use of "Maximum Take-Off Weight" in the formula. The logic is that a heavier aircraft belongs to a bigger company that is able to pay higher fees.

4.2.2.2 ATM Cost-Effectiveness Benchmarking Report

In the ACE report, the PRU reported two other measurements called "Financial Cost-Effectiveness" and "Economic Cost-Effectiveness". The "Economic Cost-Effectiveness" combines the "Financial Cost-Effectiveness" with a measure of quality, called the costs of ATFM delays. The reasoning behind it is that lower costs are not necessarily associated with better performance. Indeed, providing "a safer and more punctual, predictable and efficient service" (CANSO, 2018) might lead to increasing costs.

Financial Cost-Effectiveness

$$\frac{gate - to - gate\ ATM/CNS\ provision\ costs}{Composite\ Flight\ Hours}$$

- *ATM/CNS provision costs*:
  - Includes operating (staff, non-staff and exceptional costs) and capital related costs (depreciation and costs of capital) induced by en-route and terminal control.
  - Excludes "MET" costs, "EUROCONTROL Agency" costs, "Payments for delegation of ANS" and "Payment to governmental or regulatory authorities" which are out of ANSPs' control.
- *Composite gate-to-gate Flight Hours:* seeks to approximate the number of hours controlled by APP ATCOs in the terminal area and ACC ATCOs in the en-route zone.

[2] **Maximum Take-Off Weight**: corresponds to the maximum allowed mass of the aircraft by including the weight of the equipment, the fuel, the crew, the passengers and the cargo.

Economic Cost-Effectiveness
$$\frac{gate-to-gate\ ATM/CNS\ provision\ costs + costs\ of\ ATFM\ delays}{Composite\ Flight\ Hours}$$
• **<u>Aim</u>**: "to capture the trade-off between ATC capacity and costs" *(Performance Review Unit, 2019)* by combining the "Financial Cost-Effectiveness indicator" and "costs of ATFM delays per unit output". • *<u>Costs of ATFM delays[3]:</u>* According to the report published by the University of Westminster (Cook & Graham, 2011), the costs of one minute of ATFM delay is estimated at 102€ in 2017.
Composite Flight Hours (CFH)
$$En-route\ flight\ hours + (0.27\ \times IFR\ airport\ movements)$$
• *<u>En – route flight hours:</u>* "difference between the exit time and entry time of any given flight in the controlled airspace of an operational unit" (EUROCONTROL, 2019) • *<u>IFR[4] airport movements:</u>* gather all the movements of take-off and landing for IFR flights. It is considered as an output measure for terminal ANS. • *<u>0.27</u>:* the relative importance of terminal and en-route costs in the cost base (EUROCONTROL).

Table 4-3: Cost-efficiency indicators introduced by the PRU (Source: Personal)

4.2.3 Endogenous and exogenous factors

The problem of those ratios arises when the benchmarking analysis is carried out. Indeed, a comparison between ANSPs is meaningful only under the assumption that they operate under the same environmental conditions. However, as seen in *Chapter 2*, this postulate is not correct. Therefore, the analyses provided by the PRU are factual because those indicators do not reflect the external factors that influence the ANSPs' performance.

Aware of this issue, the PRU pinpointed elements impacting the ANSPs performance and classified them into three categories. The factors:

- Outside direct control of ANSPs (exogenous)
- Under influence of State and International institutions
- Under direct ANSPs' control (endogenous)

[3] ATFM delay: the duration between the last take-off time requested by the aircraft operator and the take-off slot allocated by the Central Flow Management Unit following a regulation communicated by the FMP (Flow Management Position), in relation to an airport (airport delay) or a sector (en-route delay) location (EUROCONTROL, 2016).
[4] IFR: Instrument Flight Rules "can operate in all weather conditions" in contrast with Visual Flight Rules (VFR) movements (Arblaster, 2018).

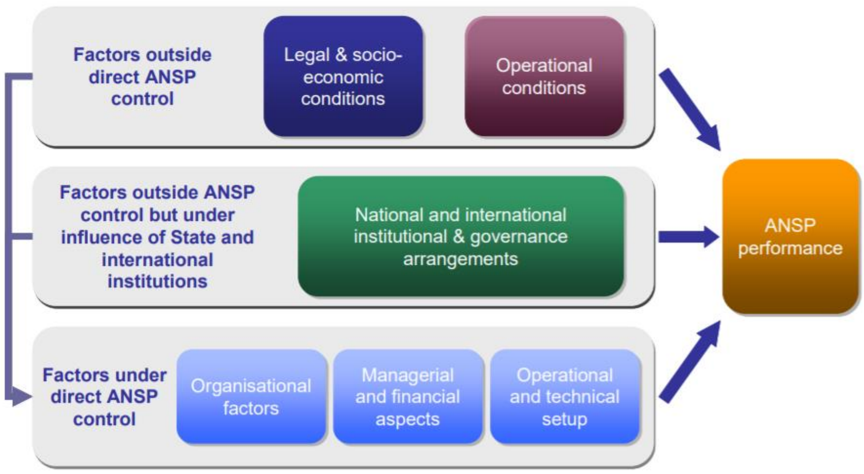


Figure 4-3: Factors affecting cost-effectiveness performance (Source: (Performance Review Unit, 2019))

Among the exogenous factors, there are three broad classifications depending on the nature of the drivers.

Legal & socio-economic conditions	Operational conditions	National and international institutional & governance arrangements
= Gather the conditions prevailing in each country and set by national policy makers.	= Are related to the characteristics of the airspace environment in which the ANSPs operate.	= Correspond to the arrangements established by regulations and aviation laws.

• Exchange & inflation rates
• Cost of living & market wage rates
• Political factors
• Taxes on turnover or profit
• Accounting standards
• Working hours
• Retirement age
• Social security and pensions
	• Size of the ANSP
	• Traffic complexity
	  ○ Density of traffic
	  ○ Structural complexity
	  ○ Traffic mix
	• Spatial and temporal traffic variability
	• Type of airspace under ANSP's responsibility
	• Weather
		• Information disclosure & independent benchmarking
		• Overall policy for "market access"
		• Degree of economic oversight/regulation
		• Institutional structures
		• Ownership and control structures
		• Civil/military arrangements

Table 4-4: Exogenous factors (Adapted from PRU 2019)

Up to now, the PRU proposed indexes for four exogenous factors: the size of ANSP, the adjusted density, the structural complexity and the seasonal traffic variability.

4.2.4 Composite indicators

By taking those parameters into account, the definition of cost-efficiency gets more complex and can be summarized as "a measure of how far an organisation's cost is from the best practice organisation's cost if both were to produce the same output under the same environmental conditions" (Isik & Hassan, 2002). Thus, the evaluation of cost-efficiency in a benchmarking analysis does not only depend on the total costs anymore but the particularities of the production process and the exogenous factors must be considered to compare apples with apples.

Many studies were carried out to address this issue. The researchers applied various **models**, on diverse explanatory **variables** and obtained different **results**.

4.2.4.1 Models

The cost-efficiency is often evaluated through frontier efficiency analyses. They consist in estimating a production frontier used as a benchmark and in computing the distance between this benchmark and the performance of the organisations.

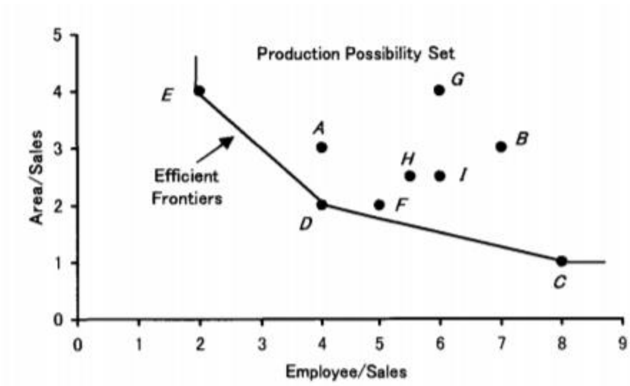


Figure 4-4: Example of an efficient frontier where all the companies (A,B, C, …) produce the same amount of output with different combination of inputs (Source: (Cooper, et al., 2007)**)**

Those "best practice" frontiers might be built by means of two different methods: the parametric, built from econometric, and the non-parametric, which relies on mathematical programming (Dong, et al., 2014). The Stochastic Frontier Analysis (SFA) is the most used parametric method whereas Data Envelopment Analysis (DEA) is famous as non-parametric technic. Those two models have been already applied in the ANSPs field. However, the SFA and DEA have each advantages and drawbacks which differ.

Stochastic Frontier Analysis (SFA)

The SFA is based on a regression analysis. The advantage of this technic is its capacity to separate the impact of statistical noise and the effect of inefficiency.

$$y_i = \alpha + x_i\beta + (v_i + u_i) \qquad i = 1, \dots, n$$

- y_i: total or unit cost of production of the ith firm
- α: a constant
- x_i: input prices and output of the ith firm
- β: vector of unknown parameters
- v_i: random variables assumed to account for the **statistical noise**
- u_i: non-negative random variables assumed to account for the cost of **inefficiency**

Figure 4-5: Example of SFA specification (Adapted from (Rowena, 2001))

However, one of the drawbacks is that it requires to know, a priori, the functional form of the relationship between the inputs and the outputs (i.e. Cobb Douglas or translog). Indeed, the econometric analysis allows estimating the parameters of that relation. Another disadvantage is that the probability distributions of the efficiency and the error term have to be also specified.

Despite those cons, this technic was investigated, inter alia, by four groups of researchers to handle the issue of heterogeneity between ANSPs' environmental conditions in the assessment of benchmarking cost-efficiency analysis.

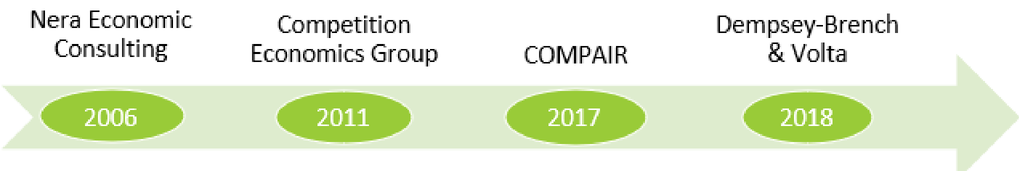


Figure 4-6: Studies carried out by using SFA (Source: Personal)

Three of those studies were requested by international institutions. Indeed, firstly, the PRU commissioned Nera Economic Consulting and Competition Economics Group (CEG) to tackle the problem. Later on, COMPAIR received funds from the EC to lead this project and tried to estimate the impact of ownership structure on ANSPs' performance. The last one was realized independently by Dempsey-Brench and Volta. Even though the authors adopted the same method, SFA, there are obvious differences between the analyses.

On the one hand, COMPAIR evaluated the inefficiency separately for en-route and terminal control services. This avoided the use of the Composite Flight Hours which is, according to them, artificially created. Indeed, the weighting factor is the same for all ANSPs even though the intensity of en-route and terminal control activities is different for each of them.

On the other hand, the SFA requires the specification of the shape of the frontier and the distribution of the error term. Those choices led to disparities between the results obtained.

Functional Form	
Cobb-Douglas	Supposes a log-linear relationship between the total costs, the output, the inputs and the exogeneous factors. $$\ln(Y_t) = \alpha + \beta \ln x + u_t$$ $$where\ \alpha = constant, x = explanatory\ variables, u = error\ term$$ **Equation 4-1: Cobb-Douglas functional form**
Translog	Is more flexible than the Cobb-Douglas. However, the number of parameters involved explodes (Pavescu, 2011) which necessitates a bigger dataset to assess the coefficients. $$\ln(Y) = \ln A + \sum_{i=1}^{n} \alpha_i \ln X_i + \frac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} \ln X_i \ lnX_j$$ **Equation 4-2: Translog functional form**
Error Term	
Decomposition	The error term is further decomposed into at least two components: 1. Random term which captures statistical noises. 2. A zero-bounded term which represents the inefficiency.
Differences	The difference between the reports is the assumption made on the variance of the inefficiency over time and with exogenous drivers. All the consulting firms tried different types of models.

Table 4-5: Functional Forms and error term considered by the studies (Source: Personal)

Data Envelopment Analysis (DEA)

The non-parametric method, DEA, is based on mathematical programming to measure the relative efficiency of several entities using multiple inputs to produce multiple outputs.

$$(FP) \qquad \max_{v,u} \theta = \frac{u_1 y_{1o} + u_2 y_{2o} + \cdots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}}$$

$$\text{Subject to} \qquad \frac{u_1 y_{1j} + u_2 y_{2j} + \cdots + u_s y_{sj}}{v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj}} \leq 1 \ (j = 1, \ldots, n)$$

$$v_1, v_2, \ldots, v_m \geq 0$$

$$u_1, u_2, \ldots, u_s \geq 0$$

Where

- $(y_{1o}, \ldots, y_{so})$: output vector of the firm o
- $(x_{1o}, \ldots, x_{mo})$: input vector of the firm o
- $(u_1, \ldots, u_s)$: weights for the outputs
- $(v_1, \ldots, v_m)$: weights for the inputs

Figure 4-7: Fractional Program introduced by Charnes, Cooper and Rhodes in 1978

In contrast with the SFA, the traditional DEA considers that all the deviation from the frontier is due to inefficiency. By consequent, the efficiency measures tend to be lower than those generated by the SFA. However, the advantages are that the DEA can be used in the case of multiple outputs and it does not require a priori assumptions on the shape of the production frontier as well as on the probability distribution of the efficiency.

The use of DEA was further examined by several researchers.

Figure 4-8: Studies carried out by using DEA (Source: Personal)

R. M. Arnaldo et al. applied four models of traditional basic DEA on different inputs and outputs. The result obtained is a ranking of the ANSPs for each model. On the contrary, K. Button and R. Neiva, V. Bilotkach et all got efficiency scores through the bootstrap DEA which provides confidence levels of the efficiency measures for each ANSP. The difference between those two pieces of research lies in the evaluation of the impact of environmental variables. K. Button and R. Neiva conducted a bootstrap regression with the bias-corrected bootstrapped DEA efficiency scores as the dependent variable and the environmental components as explanatory variables. V. Bilotkach et all did not consider the environmental conditions but they produced some insight on the changes in productivity between two periods by means of Malmquist Total Factor Productivity Index.

4.2.4.2 Variables

The common point of all those studies is the manipulation of data collected by the PRU. However, the Nera Economic Consulting and CEG also extracted data from International Monetary Fund Outlook Database (cost-of-living), from Eurostat (capital index) and from the Transparency International Database (Business environment quality variable). Even though the main dataset originates from the same source, it was exploited differently in each research because of the period considered and also the types, number as well as computations of the explanatory variables.

4.2.4.3 Results

Due to the disparities between the models and the variables employed, the results for each analysis are also different.

2006 For the first study, the **NERA Economic Consulting** (2006) failed to assess the level of inefficiency for each ANSP. Also, the coefficients for some contextual variables were surprisingly not significant.

2011 The **CEG** (2011) estimated four models but only two of them were retained. However, the results obtained were quite divergent: the Pitt & Lee model predicted 60% of inefficiency and for the True Random effects, it was 13%.

2014 **Rosa M. Arnaldo & all.** (2014) provided a table with the ranking of the ANSPs computed according to four approaches. Some of the results were quite different. For example, DFS had an efficiency score of 100 for model 1 but 38.67 for model 2.

2014 **K. Button & R. Neiva** (2014) measured efficiency scores for the ANSPs between 2002 and 2009 and evaluated the impact of environmental conditions on the performance. They concluded that some ANSPs maintained their efficiency score while others improved it over the period. However, a lot of ANSPs have low efficiency scores which can be partially explained by the operational environment in which they operate.

2015 **V. Bilotkach et all**. (2015) produced a range of efficiency scores for each ANSP. They mainly demonstrated that the overall productivity has increased over the time period and that the ANSPs use an inefficient mix of resources.

2017 **COMPAIR** (2017) succeeded to estimate an individual level of inefficiency for the en-route and terminal services. However, it is sometimes difficult for the ANSPs to separate their costs according to the types of services provided.

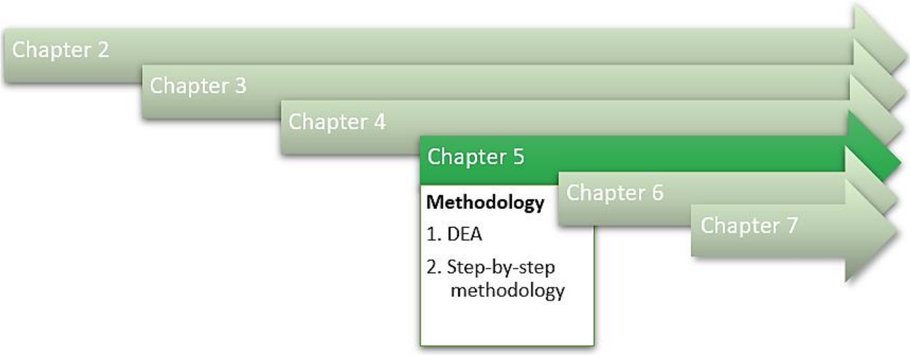
2018 The objective of **Dempsey-Brench & Volta** (2018) was to understand the impact of ownership structure on the ANSPs' cost performance. They discovered that this characteristic does not have a significative influence on the cost structure.

4.2.5 Conclusion

The differences in the variables, the models, … are linked to the choices made by the authors. Indeed, the construction of CI implies subjectivity, and this is one reason why the analysts are sometimes reluctant to use them. Therefore, as mentioned by OECD & JRC (2008), it is important to follow a complete methodology which helps to clearly specify the decisions made and which relies as much as possible on statistical principles. The consulting groups, which carried out projects at the behest of the PRU, explicitly justified each decision and tested several types of SFA. However, regarding the use of DEA, the researchers did not undertake their studies with as much rigour. This lack of precision might affect the very credibility of their work. Therefore, it is necessary to apply a methodology which helps the managers to understand and accurately interpret the results.

Chapter 5: Methodology

As seen previously, the construction of an efficient frontier, through DEA or SFA models, is a common practice to compute the cost-efficiency of organisations. In this master thesis, the choice fell on DEA to implement the cost-efficiency indicator for ANSPs. This model offers many advantages. Inter alia, in contrast with SFA, the specification of a functional form is not required, the inputs-outputs relationship is not obligatorily the same for all the entities and the inclusion of multiple outputs is possible without knowing a priori the weights (Cooper, et al., 2004). To get an idea of what is going to be developed in the next chapters, the basic concepts behind the DEA and the broad-based methodology followed to construct our CI are explained in this section.



5.1 DEA

The most basic DEA, referred to *CCR* model, was introduced by Charnes et al. (1978) and has been implemented in many fields. The main idea is that efficiency is expressed through the ratio between multiple types of resources (inputs), used by a company, and outputs produced. This principle can be thus applied to any production process. Within the framework of DEA, the entities responsible for converting inputs into outputs under environmental conditions are called Decision Making Units (DMUs) and their performance is evaluated. The frontier analysis allows to carry out a relative comparison between those DMUs to bring out the efficient ones and to encourage to uncover best practices among them.

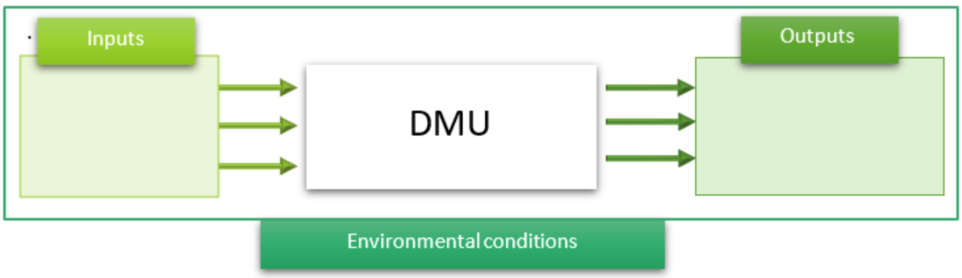


Figure 5-1: Production Process of a DMU (Source: Personal)

Each DMU needs a combination of inputs, x (such as labour, capital, …), to generate outputs (such as products or services), y. Therefore, the efficiency score for each DMU_j is assessed through a ratio of outputs and inputs combined together through weights.

$$\frac{Virtual\ Output}{Virtual\ Input} = \frac{u_1 y_{1j} + \cdots + u_s y_{sj}}{v_1 x_{1j} + \cdots + v_m x_{mj}}$$

Usually, the weights are fixed in advance such as the CFH ($= \text{En} - \text{route flight hours} + \mathbf{0.27} \times \text{IFR airport movements}$) where 0.27 is the same for all ANSPs even though the intensities of their en-route and terminal activities are not identical. The benefit of using DEA is that the weighting is derived from the data in a way that the best set of weights is assigned for each DMU to maximise their efficiency ratio (Cooper, et al., 2004).

For illustrative purpose, let's take a dataset containing n DMUs ($j = 1, \dots, n$). Those entities consume m different quantities of inputs, x_{ij} ($i = 1, \dots, m$) and deliver s various outputs, y_{rj} ($r = 1, \dots, s$). This problem is then expressed by this fractional program (FP) for the DMU_o, called *CCR* Model with reference to the authors Charnes, Cooper and Rhodes:

(FP)

$$\max \boldsymbol{\theta} = \frac{u_1 y_{1o} + u_2 y_{2o} + \cdots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}}$$

Subject to

$$\frac{u_1 y_{1j} + u_2 y_{2j} + \cdots + u_s y_{sj}}{v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj}} \leq \mathbf{1}\ (j = 1, \dots, n)$$ All efficiency scores ≤ 1

$$v_1, v_2, \dots, v_m \geq 0$$
$$u_1, u_2, \dots, u_s \geq 0$$
Positive weights

Figure 5-2: Fractional Program introduced by Charnes, Cooper and Rhodes in 1978

This mathematical programming problem is run n times, once for each DMU being evaluated. The objective is to find the output weights (u_r) and the input weights (v_i) that maximise θ (the ratio) of the DMU for which the performance is assessed, in this case DMU_o. The weights have to be positive and once assigned to the other DMUs, have to give efficiency scores lower or equal to 1. However, to solve this mathematical problem, it is necessary to replace the fractional program with a linear program (LP) expressed as followed:

(*LP*)	$\max \boldsymbol{\theta} = \boldsymbol{\mu_1 y_{1o} + \mu_2 y_{2o} + \cdots + \mu_s y_{so}}$
Subject to	$\boldsymbol{v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj} = 1 \ (j = 1, \dots, n)}$
	$\boldsymbol{\mu_1 y_{1j} + \mu_2 y_{2j} + \cdots + \mu_s y_{sj} \le v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj} \ (j = 1, \dots, n)}$
	$\boldsymbol{v_1, v_2, \dots, v_m \ge 0}$
	$\boldsymbol{\mu_1, \mu_2, \dots, \mu_s \ge 0}$

Figure 5-3: Linear Program (Source: (Cooper, et al., 2007))

Considering the optimal solution obtained (θ^*, v^*, μ^*) where θ^* is the set of efficiency scores, v^* and μ^* are respectively the vectors of optimal input and output weights, the DMU_o is classified as CCR-efficient if $\theta_o^* = 1$ and there is at least one optimal (v^*, μ^*) with $v^* > 0$ and $\mu^* > 0$ (Cooper, et al., 2007). In other words, according to the Pareto-Koopmans definition, "full efficiency is attained by any DMU if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs" (Cooper, et al., 2011). Otherwise, the DMU is CCR-inefficient and the DEA provides a *reference set* which gathers all the CCR-efficient DMUs that force the DMU to be inefficient.

5.2 Step-by-step methodology

Even though the DEA basic principles seem easy to understand, it is essential to be as objective as possible and to specify all the choices made through the construction process of the indicator in order to build a trustful CI. Through the literature, there is currently no step-by-step methodology to wisely implement a DEA.

5.2.1 Framework proposed by OECD and JRC

Consequently, the methodology followed in this master thesis will mainly be based on the framework explained in the *Handbook on Constructing Composite Indicators: Methodology and User Guide* jointly prepared by OECD (the Statistics Directorate and the Directorate for Science, Technology and Industry) and the Econometrics and Applied Statistics Unit of the Joint Research Centre (JRC) of the EC (2008). The aim of this book is to provide guidelines to enhance the quality of the newly built CI. The framework proposed by OECD and JRC is broken down in 10 steps. For the sake of simplicity, we might group them into four stages:

Figure 5-4: Four stages for constructing a CI (Source: Personal)

		Steps	Descriptions	Procedure
FOUNDATION	1	Theoretical framework	It is the foundation for constructing a CI because it introduces all the relevant concepts and the variables which will drive it.	1. Define the concept to be measured. 2. Describe subgroups composing the CI.
FOUNDATION	2	Data Selection	After defining the main concept computed by the future CI, KPIs have to be selected according to selection criteria (Relevance, Simplicity, Validity, …). This step is quite subjective, but it is decisive for the quality of the CI.	1. Scrutinize the strengths and weaknesses of each possible indicator. 2. Summarize their characteristics (availability, source, type, frequency of update, …) in a table.
PREPARATION	3	Imputation of missing data	In practice, it is infrequent to have a complete data set without outliers. This issue might lead to a distortion of the final ranking. Consequently, it is important to choose the most adequate approach for managing missing values and to consider the outliers.	1. Clarify the reason of missing data: Is the absence of the value dependent on observed variables (Missing at Random), on the value itself (Not Missing At Random) or neither (Missing Completely At Random)? 2. Choose and implement suitable methods. 3. Discuss the presence of outliers
PREPARATION	4	Multivariate Analysis	After dealing with missing values, an explanatory analysis should investigate the interrelationships between the indicators. It is possible to group information on individual indicators or to detect similarities between DMUs.	1. Choose and implement methods to uncover the underlying structure of the dataset. 2. Identify groups of indicators or similar entities. 3. Compare to the structure to the theoretical framework and discuss differences.

	5	Normalisation	It is necessary to bring the indicators to the same standard to avoid mixing measurement units with different range.	1. Choose and implement a method to normalise the data. 2. Make scale adjustment, if necessary.
CONSTRUCTION	6	Weighting and Aggregating	This step aims to combine the indicators into a single index by choosing the most adequate weighting and aggregating methods which are consistent with the theoretical framework and the properties of the data. The correlation between indicators need to be taken into account in order to avoid double counting issues in the CI.	1. Test the correlation between indicators by using Pearson Correlation Coefficient. 2. Choose and implement a method to assign weights. 3. Specify if the CI allows compensability between its components. 4. Choose and implement a method to aggregate the indicators.
	7	Robustness and Sensitivity Analysis	The trust in the CI might be enhanced by assessing its robustness. Several choices have been made through the construction process. The uncertainty and the sensitivity analyses "can help gauge the robustness and improve the transparency of the composite indicator" (Joint Research Centre-European Commission, 2008). No scenario is a priori better than another. By consequence, alternative scenarios should be considered.	1. Determine all potential sources of uncertainty in the development of the composite indicator: selection of indicators, weighting methods, … 2. Assess how those sources of uncertainty propagate through the process and affect the final value of the outputs. 3. Assess how those sources of uncertainty contribute to the variance of the output.
INTERPRETA	8	Back to the data	After computing the CI, it is interesting to know the components that drive the results for each country.	1. Choose and implement a type of representation to show the contribution of each subcomponent
	9	Links to other indicators	The CI should be compared to other relevant measures in the same field.	1. Choose and implement a type of representation to show the link between the CI and other measurements.

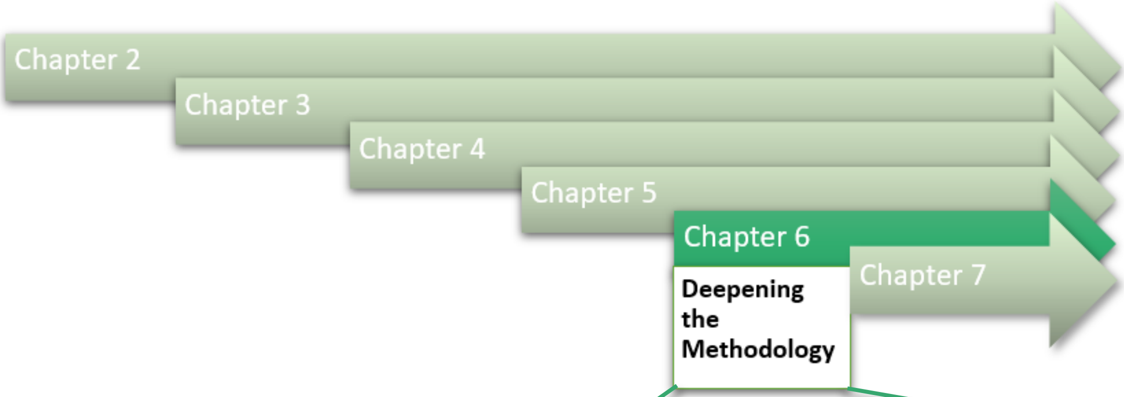
T I O N	10	Visualisation of the results	Finally, to ease the interpretation of the indicator by the target audience, it is important to clearly and accurately present the CI.	1. Choose a set of visualisation that communicate the most information 2. Choose the tool to represent the indicator 3. Show the results

Table 5-1: Overview of the methodology introduced by OECD and JRC (Extended from: (Joint Research Centre-European Commission, 2008))

5.2.2 Proposition

However, this framework gives only general recommendations which are independent of the model applied. Therefore, before running the DEA, we will develop a step-by-step methodology to implement trustful efficiency scores by completing this widely used framework with tools specifically used in the context of DEA.

Chapter 6: Deepening the methodology (Contribution 1)



As explained previously, the general methodology for constructing a CI is broken down into ten specific steps that we gathered in four stages. In this section, every part will be analysed and completed with technics applicable to DEA. The newly built framework is graphically represented in *Appendix 2*.

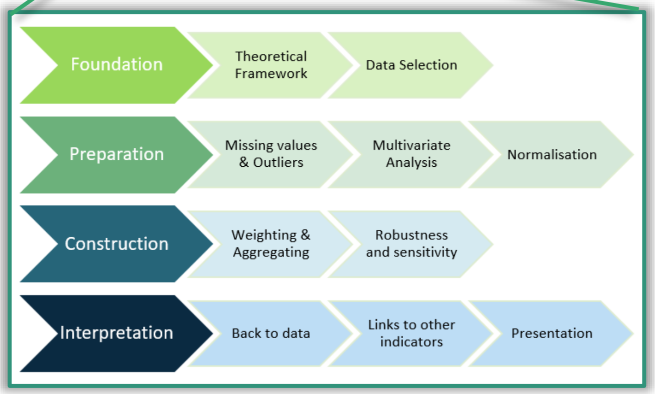


Figure 6-1: Four stages and ten steps for constructing a CI (Source: Personal)

6.1 Stage 1: Foundation

The first stage, we called "Foundation", implies the description of the theoretical framework and the selection of indicators.

6.1.1 Step 1: Theoretical Framework

The theoretical framework is the keystone for building a CI. It consists in specifying the concept to be measured and the subgroups composing the CI.

6.1.1.1 Definition

The notion estimated by the DEA is the relative efficiency of the DMUs. The efficiency can be defined as "performing activities as well as possible". In this context, this measure is obtained by constructing a "best-practice" frontier composed of efficient DMUs. The other entities are considered as inefficient if it is possible to improve either the inputs or outputs without worsening the others. Indeed, there exist many sources of inefficiencies. The three most common are "*Pure Technical Inefficiency*" (1), "*Mix Inefficiency*" (2) and "*Scale Inefficiency*".

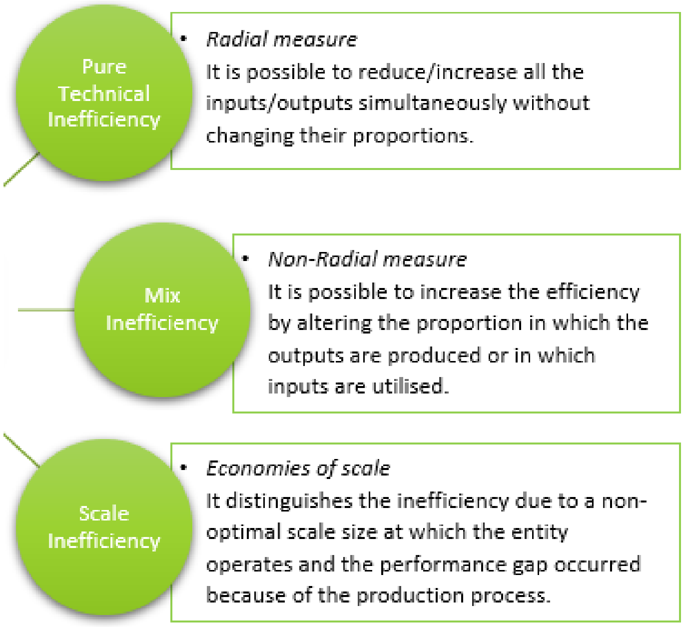


Figure 6-2: Sources of inefficiency (Source: Personal)

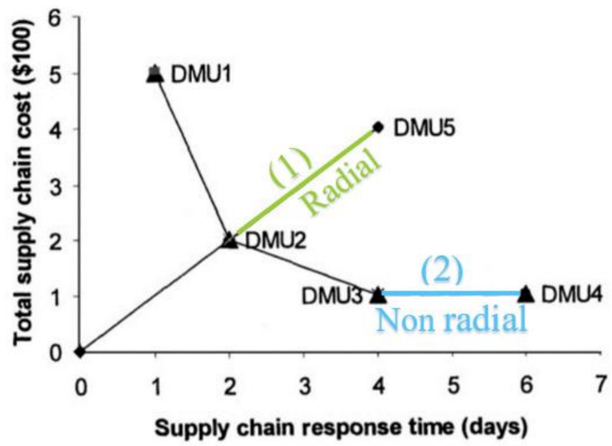


Figure 6-3: Radial and non-radial measures for two inputs and one output (Adapted from (Zhu, 2014))

In addition, in the case of the evaluation of cost-efficiency, the *Allocative Inefficiency* can be computed if the accurate prices of the inputs are available. It indicates whether the mix of inputs is effective regarding the prices.

In conclusion, the performance gap can be disaggregated into several components to better interpret the nature of the disparities between DMUs.

6.1.1.2 Subgroups

However, first and foremost, the efficiency reflects how well an organisation handles the **inputs** to generate a certain amount of **outputs** under **environmental** conditions.

To identify these three subgroups of components, it is necessary to firstly clarify the purpose of the benchmarking and so the concept to be assessed (Thanassoulis, 2001; Cook, et al., 2014). For instance, the aim might be to evaluate the operating efficiency or the ability to attract customers. Those two examples entail various inputs and outputs. Once this step is achieved, the variables to consider can be determined.

Variables	Description
Inputs	• Gather all the resources and are to be minimised (the less-the-better) (Cook, et al., 2014).
Outputs	• Represent all the outcomes generated by the organisation in accordance with the type of efficiency to be measured. • Are either: o Desirable: are to be maximised (the more-the-better).

	◦ Undesirable: are to be minimised (i.e. waste, pollutants).
Environmental	• Are the contextual factors impacting the performance of the entities (Thanassoulis, 2001). • Are defined as « an external variable which could nevertheless influence the availability or requirement of resources » (Wagner & Shimshak, 2007)

Table 6-1: Subgroups to consider in DEA (Source: Personal)

6.1.1.3 Recommendations

Those input/output/environmental variables are sometimes difficult to identify. In many studies carried out to assess efficiency, the researchers built a table summarising the models already applied in the same field and based on frontier construction (SFA and DEA). Besides this table, it could be useful to draw a schema modelling the process studied to measure the efficiency. Finally, the initial set of variables should ideally be determined in collaboration with experts.

6.1.2 Step 2: Data selection

Once the input/output/environmental variables are pinpointed, the most appropriate indicators to measure each of those components must be chosen. For each concept, many measurements or proxies are possible. As reported in the *Handbook on Constructing Composite Indicators* (Joint Research Centre-European Commission, 2008), "the strengths and weaknesses of composite indicators largely derive from the quality of the underlying variables". For that reason, the quality of the data should be evaluated before making the selection. Six criteria were suggested by the OECD in the "Quality Framework for Composite Indicators" (Joint Research Centre-European Commission, 2008) and are summarised below.

Criteria	Sub-criteria	Definition
Relevance		Refers to the extent to which the indicator meets the needs of the users. It has to be assessed in view of the overall purpose of the benchmarking.
Accuracy		Refers to "the closeness between the values provided and the true values". It is interesting to know whether the methodologies to compute the indicators evolved or whether the indicators were updated.

Timeless	Timeless	Represents the time lag between the availability of the data and the concept it describes.
	Punctuality	Represents length of time between the target delivery date and the actual date of release of the data.
Accessibility	Accessibility	Describes how easy it is to access those data.
	Clarity	Refers to the information provided on the statistics (explanation, documentation, …).
Interpretability		Describes how easy it is for the users to understand and accurately use as well as analyse the data.
Coherence		Reflects the degree to which the data are consistent over time and across countries.

Table 6-2: Criteria to judge the quality of indicators (Adapted from (Joint Research Centre-European Commission, 2008)

6.2 Stage 2: Preparation

The indicators selected should undergo some transformations before being combined into a single index.

6.2.1 Step 3: Dealing with missing data and outliers

One essential step to obtain a quality dataset is to deal with missing values and to also discuss the presence of outliers. Indeed, in practice, it is infrequent to have a complete data set without atypical observations. This issue might lead to a distortion of the final ranking. Therefore, it is important to choose the most suitable approach for managing missing values and to detect the outliers.

6.2.1.1 Missing data

The main assumption of the DEA is that all inputs, outputs and environmental variables are known and available. Consequently, the DEA is sensitive to missing variables and other issues related to the quality of data. There exist different methods to handle this issue.

Before searching for the potential technics to implement, the reasons of the inexistence of the variables should be clarified: Is the absence of the value dependent on observed variables (Missing at Random), on the value itself (Not Missing At Random) or neither (Missing Completely At Random)? The answer will help to pick the approach to take.

There are some general technics, which do not depend on the model applied, and others which are only relevant in the context of the DEA.

GENERAL TECHNICS	

Remove the observation	• Aim: Remove the DMUs with the blank entries from the dataset (Smirlis, et al., 2006; Kuosmanen, 2009; Azizi, 2013).
	• Good strategy when the sample is very large because only a few DMUs will be affected by the missing data. Nevertheless, the rejection of a DMU will lead to several issues :
	  ○ It worsens sampling errors.
	  ○ It will have an impact in an unpredictable way on the ranking of the other DMUs due to the loss of information.
	  ○ Efficiency score of the discarded DMUs will not be computed.

Discard the output or input from the analysis	• Aim: Discard the output or input with blank entries from the analysis. • It is “acceptable” when the variable is of poor quality or when the variable could be sufficiently approximated by some other indicators (Kuosmanen, 2009). • Problem: rejecting a relevant variable causes unpredictable changes in the efficiency scores and compromises the reliability of the analysis.

Estimate the exact approximation	• Aim: Apply imputation technics to estimate the exact value. This should be implemented with caution considering that it might lead to misleading efficiency results (Smirlis, et al., 2006).
	• Different methods explained in the *Handbook on construction composite indicators* (Joint Research Centre-European Commission, 2008), taken as reference book:
	  ○ **Single imputation**: replaces the missing values by results obtained from the mean/median/mode, a regression, hot- or cold- deck imputation, Expectation-Maximisation imputation or substitution.
	  ○ **Multiple imputation**: is based on a random process that reflects uncertainty (i.e. Markov Chain Monte Carlo algorithm).
	  ○ **Nearest Neighbour**: replaces the missing values by the value of the most similar case.

	• Other tools exist when dealing with time series (Wahab, 2017). For instance, the value (x_0, y_0) and (x_2, y_2) are available:
	  ○ **Interpolation**: is employed to find value between the known data (i.e. find y_1 for a given value x_1).
	  ○ **Extrapolation**: is applied when it is needed to estimate the value beyond the range (i.e. find y_3 for a given value x_3).

SPECIFIC TECHNICS FOR DEA	

Replace the missing output by zero and/or missing input by a large number	• Aim: the solution proposed by Kuosmanen (2009) is to replace
	  ○ Missing desirable output = 0 or value under which the variable will never fall.
	  ○ Missing input = large enough number that the missing value would never exceed.
	• Result: this method is at least as good approximation of the ideal frontier as the rejection of DMUs or output variables, even better.
	• Problem: it is unfair. So, this technic is suitable only if
	  ○ it is more important to include all the DMUs in the analysis rather than to obtain a fair comparison.
	  ○ it is needed to give incentives to the DMUs to share their data.

Intervals	• Aim: Replace the missing values by approximation through intervals in which the data is likely to belong (Smirlis, et al., 2006; Azizi, 2013). • Result: it is possible to estimate the lower and upper bounds of the efficiency scores of the problematic DMUs.
Fuzzy approach	• According to Kao & Lui (2000), replacing the missing data by an estimated crisp value results in misleading efficiency scores. To handle this problem, they proposed a new approach which applies ideas from the fuzzy theory. • Aim: the uncertain value is represented by membership functions built from the smallest (most pessimistic), the median (most likely) and the largest (most optimistic) of the observations. • Problem: This method produces good result but the efficiency scores will be expressed through fuzzy numbers, which, in contrast with real number, do not refer to one single value but to a set of possible values.

Table 6-3: Technics to deal with missing values (Source: Personal)

In conclusion, this table (Table 6-3) provides an overview of the possible technics to apply in case of missing data. However, there is no rule to select the "best" method. The recommendation given in the reference book is to use a complete sample of the dataset, to eliminate some values and to try different approaches. The evaluation of the performance of each technic is not addressed in this section but it is further explained in the reference book.

6.2.1.2 Outliers

The second problem related to the quality of the dataset is the presence of outliers. These are DMUs that disproportionately differ from the other observations. As mentioned by Bogetoft & Otto (2011), a firm might be viewed as an outlier for several reasons:

1. It may be due to **errors** in the data which should be corrected.
2. Those firms, that may be correct but **atypical**, should be discarded to avoid a distortion of the model which will fit these outliers.
3. Observations with **exceptionally low or high relative performance** are considered as outliers for precautionary reasons. Sometimes, they reflect for example the introduction of new technology or innovation in management practice.

They should be identified because the DEA analysis is sensitive to those atypical observations. A simple way to detect some of them is to generate a **scatter plot matrix**. This is used to visualise relationships between two variables.

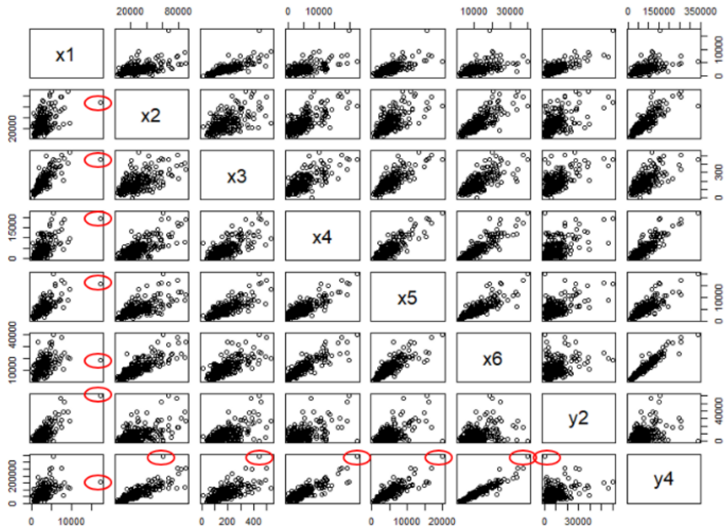


Figure 6-4: Scatter plot matrix on "pigdata" dataset (Source: Personal)

This solution allows to highlight extreme features through a linear combination of two variables. However, it is difficult to name the atypical observations. To solve this issue, the most common technic, used to precisely identify a group of outliers through the combination of more than two variables, is called the data cloud method introduced by Wilson (1993).

The **data cloud method** is based on the computation of the determinant of the combined input-output matrix. If removing a DMU from the dataset leads to a significant change in the determinant of the matrix, this observation is classified as outlier. The advantage of this technic is that it is possible to detect a group of outliers by removing several DMUs at a time.

Before implementing the model, it is indispensable to set the maximum number of DMUs to eliminate from the dataset at the same time. For instance, in the example (see Figure 6-4), it seems that there is only one atypical observation. So, to be sure, the maximum number of discarded firms could be 10 which corresponds to the total number of iterations of the algorithm.

During the first iteration, the algorithm removes one row at a time and computes the ratio between the determinant of the matrix before and after the operation. Then, the algorithm returns the observation that provoked the biggest change in the determinant of the matrix. The second iteration follows the same principle but with two observations and so for. The results of the algorithm are shown in the Figure 6-5. Of course, the greater the number of rows discarded, the smaller the ratio. This is why a second step is needed to identify the outliers by calculating all the logarithms of the ratios and the minimum ratio (log (R/R_{min})).

Iteration	**Observation**										$\mathbf{R_{min}}$
1	165										0.561
2	57	165									0.39
3	201	57	165								0.297
4	36	201	57	165							0.234
5	64	36	201	57	165						0.189
6	110	64	36	201	57	165					0.150
7	110	34	64	36	201	57	165				0.121
8	110	34	86	64	36	201	57	165			0.097
9	33	110	34	86	64	36	201	57	165		0.079
10	66	33	110	34	86	64	36	201	57	165	0.064

Figure 6-5: First step to detect outliers (Source: Personal)

The groups of outliers are detected by observing if there is a gap between the observations at 0 and above 0, in other words, when the curve connecting the second smallest value of the log-ratio reaches a peak. As predicted, it seems that there are two outliers, the observations 165 and 57.

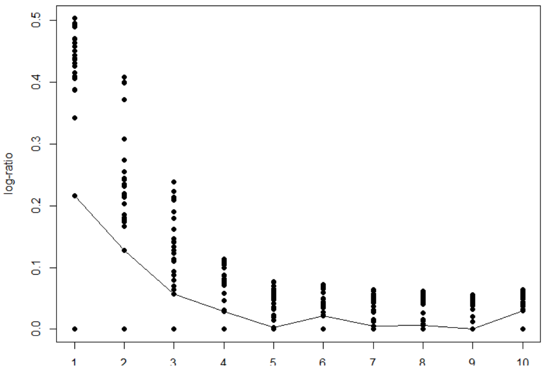


Figure 6-6: Second step to detect outliers (Source: Personal)

As soon as these atypical observations are spotted, the reasons why they are outliers should be further investigated.

6.2.2 Step 4: Multivariate Analysis

After dealing with missing values and outliers, let's take a look at the interrelationships between the indicators as well as the entities and reduce the dimensionality of the dataset if necessary.

6.2.2.1 At the indicators level

Firstly, to observe links between the indicators, the authors of the reference book suggested to perform a Principal Component Analysis (PCA).

PCA	• Aim: reduce "the dimensionality of the variable space by representing it with a few uncorrelated variables that capture most of its variability" (Tauler, et al., 2009) • It will provide an insight into the relationships between the variables when they are reduced to a few components.

6.2.2.2 At the observation level

Then, another interesting approach to tackle is to analyse the dataset at the observation level. The **clustering analysis** is a common technic to investigate similarities between DMUs. It allows to gather DMUs with similar properties in the same groups and to separate dissimilar DMUs in different clusters. It might be conducted through K-Means, DBSCAN, Expectation-Maximisation or a dendrogram algorithms.

6.2.2.3 Dimensionality reduction

Once those analyses are carried out, the selection of variables is a crucial step when the dataset contains a large number of characteristics comparatively with the number of available DMUs. Golany & Roll (1989) introduced a "rule of thumb" which implies that the number of DMUs should be at least twice the total number of inputs and outputs. However, Banker, et al. (1989) suggested that it should be three times instead of twice. Those rules have been established because the larger the number of variables considered, the higher the dimensionality and so the less discerning is the DEA. Adding a variable leads to a larger set of efficient DMUs and to higher efficiency scores. The challenge is thus to find a model with as many variables as necessary but as few as possible (Wagner & Shimshak, 2007).

There exist many methodologies to select the inputs and outputs to include in the model. In this table, a sample of the technics tested by Nataraja & Johnson (2011) and their remarks are summed up:

	Objectives	Remarks/Advice
Expert's opinion	Select variables based on their contribution to the objectives of the DEA.	• Quite subjective
Correlation	Remove variables which are highly correlated to avoid redundancies of the information.	Not sufficient because: • It is not because a variable is redundant in a regression model that it is in a DEA (Wagner & Shimshak, 2007). • The exclusion of a highly correlated indicator might significantly alter the DEA efficiency scores.
Efficiency Contribution Measure	Introduced by Paster et al. (2002), it evaluates the pertinence of a variable based on its contribution to the efficiency score. The principle is to assess the marginal impact of each variable. Then, a statistical test is applied to determine if the component is relevant for computing the DEA model.	• Works well with low correlations among inputs and may not work well with high correlation between variables (> 0.8). • Works well with a large sample size (n > 100). • Provides the input contribution. • Might be affected by the choices of the DEA models.
PCA-DEA	It was suggested by Ueda & Hoshiai (1997) and extended by Adler & Golany (2001). As explained, the principle is to combine PCA with DEA to reduce the dimensionality of the dataset and keep as much information as possible.	• Keeps information from all variables which improves the discriminatory power of DEA. • Works well with smaller sample sizes (n≈ 25). • Robust to high correlations between variables. • Vulnerable to the choice of technology. • May not work well with higher dimension datasets. • Not clear how many Principal Components are needed • Can never obtain true efficiency level.

Ruggiero's regression-based test	Introduced by Ruggiero (2005), the technical efficiency is firstly computed from a set of known inputs and outputs. Then, a regression model is implemented where the dependent variable is the technical efficiency score and the explanatory variables are the "candidates" to include in the model. If the coefficient of the candidate is significantly different from zero and has the proper sign, the variable is integrated in the DEA.	• Works well with low correlation (< 0.2) among inputs and a large sample size (n > 100). • Less vulnerable to the curse of dimensionality. • Robust to the choice of technology • May not work well with high correlation between variables (> 0.8). • Easy implementation.

Table 6-4: Technics to reduce the dimensionality of the data set (Adapted from Nataraja & Johnson (2011))

6.2.3 Step 5: Normalisation

The last transformation should be the normalisation before aggregating the indicators because these latter are often expressed in different units of measurement. This step consists in adjusting the values to a common scale. Nevertheless, the efficiency estimates obtained by means of DEA are independent of the units of measurement of the inputs and outputs provided that they are the same for every DMU (Cooper, et al., 2007). For that reason, this step is skipped.

6.3 Stage 3: Construction

Now that the indicators are analysed and transformed, they can be combined into a single index. The choice of the model is quite subjective. Therefore, all the assumptions made must be clearly stated and a robustness and sensitivity analysis must be undertaken.

6.3.1 Step 6: Weighting and Aggregating

This sixth step consists in selecting the model(s) to apply in order to compute the weights for the indicators. In our case, the general approach has already been chosen in light of the issue related to the ANSPs' cost performance. However, many extensions of the basic *CCR* model have emerged over the years and the most appropriate one should be executed depending on the assumptions stated. In this section, the two traditional models, *CCR* and *BCC*, will be explained and four questions must be asked to help to determine the technic to implement.

1. Variable Return to Scale or Constant Return to Scale?
2. Input or Output oriented model?
3. How to deal with undesirable outputs?
4. How to integrate the environmental variables?

6.3.1.1 Variable Return to Scale or Constant Return to Scale?

The difference between the *CCR* and *BCC* models lies in the assumption regarding the *economies of scale*. This phenomenon happens when an increase of the desirable output level provokes a decrease of the unit cost, in opposition to the diseconomies of scale. Those notions are related to the *Return to scale* (RTS) which describes the relationships between the input (x) and output (y) quantities. Indeed, if a firm is in a situation of *economies of scale,* then the raise of outputs will be proportionally larger than the raise of inputs which corresponds to an increasing return to scale.

The *CCR model* assumes a *constant return to scale* (CRS). In other words, if (x,y) is a feasible activity, then, for any positive t, (tx,ty) is also a feasible activity (Cooper, et al., 2007). Whereas, the *BCC model* had a convexity condition and this implies *variable return to scale* (VRS).

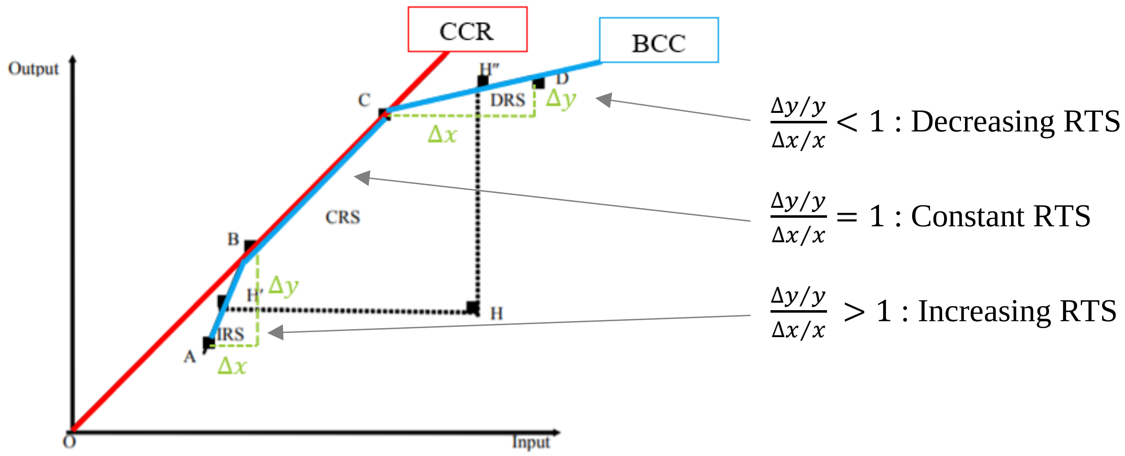


Figure 6-7: CCR and BCC (one input and one output) (Adapted from (Cooper, et al., 2007))

6.3.1.2 Input or Output orientation?

The *CCR* and *BCC* can be executed according to two angles: *output* or *input orientation*. The former is applicable when the outputs are controllable. The objective is thus to maximise the desirable output levels given the present quantities of inputs. On contrary, the *input orientation* is used when the inputs are manageable. The goal is to minimise the inputs while maintaining at least the present outputs quantities. Consequently, according to those definitions, the FP (introduced in *Chapter* 5) is solved through two linear programs:

INPUT-ORIENTED		OUTPUT-ORIENTED	
min	θ	max	η
$Subject\ to$	$\theta x_o - X\lambda \geq 0$	$subject\ to$	$x_o - X\mu \geq 0$
	$Y\lambda \geq y_o$		$\eta\, y_o - Y\mu \leq 0$
BCC only	$e\lambda = 1$	BCC only	$e\lambda = 1$
	$\lambda \geq 0$		$\mu \geq 0$

Table 6-5: Input and output oriented DEA (Source: (Cooper, et al., 2007))

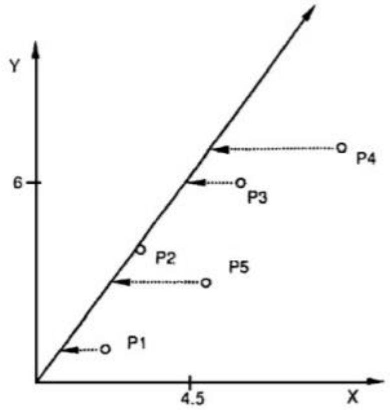


Figure 6-8 Projection to frontier for input-oriented CCR model (One input and one output) (Source: (Cooper, et al., 2004))

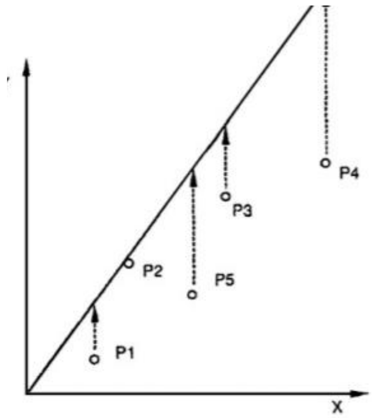


Figure 6-9: Projection to frontier for output-oriented CCR model (One input and one output) (Source: (Cooper, et

The θ and η have different ranges and meanings. The θ is situated between 0 and 1 because it indicates that the DMU_o can be more efficient by reducing all its input values by (1-θ) without changing the input mix. However, the η is higher or equal to 1 since it represents the output enlargement rate. To be efficient, the DMU_o has to increase all its outputs by this rate. Those two indicators only capture pure technical inefficiencies because they are radial measures. Therefore, to complete the analysis, the slacks, non-radial measures which represent the input excess or output shortfall, will have to be computed as well. For the sake of simplicity, the algorithm behind the slacks will not be explained. Further information on the subject is provided in the book written by Cooper, et al. (2007), Chapter 3.

6.3.1.3 How to deal with undesirable outputs?

Those traditional DEA can be performed only if all the outputs are to be maximised. However, in reality, the production process might generate undesirable outputs. The inclusion of those unwanted outputs can be done through a direct or an indirect approach (Zanella, et al., 2015). The latter consists in transforming the variables to apply a traditional DEA. The other possibility, referred to direct approach, is to treat it in a non-linear model or to use a directional distance function. In the aim of using *CCR* and *BCC models*, the focus is on indirect technics.

Options	Consequences	Applicable to
Consider the undesirable outputs as input	(+) If the undesirable outputs are treated as inputs, it will be minimised as wanted. (-) However, as mentioned by Seiford & Zhu (2002), it does not "reflect the true production process".	*CCR and BCC*
Data transformation	The data transformations allow to turn "the-smaller-the-better" into the "the-bigger-the-better" type of outputs. There are two possible modifications (u_i = undesirable output):	
	• **Non-linear**: for instance, the reciprocal multiplicative $f(u_i) = \frac{1}{u_i}$. The drawback is that the non-linear transformation might deform the efficient frontier (Cooper, et al., 2007).	*CCR and BCC*
	• **Linear**: for instance, Seiford & Zhu (2002) suggested to multiply the unwanted outputs by -1 and then add a number large enough to make all the values positive. This method is classification invariant. It means that this data transformation does not affect the classification between inefficient and efficient but it has an impact on the scores.	*BCC only because CCR is not translation invariant.*

Figure 6-10: Undirect approaches to handle undesirable outputs (Source: Personal)

6.3.1.4 How to integrate the environmental variables?

At this stage, all the inputs and outputs are eventually integrated into the model. However, the DEA is based on the hypothesis that the DMUs are comparable and face similar environmental conditions. This is not always the case. Therefore, the impact of the characteristics of the operational environment should be evaluated to dissociate the inefficiency due to exogenous and endogenous factors.

The contextual variables might be either categorical or numerical. The categorical factors spotlight the presence of groups. In this case, it would be interesting to compute group efficiency scores. Whereas the environmental variables are continuous, two main approaches have been investigated in the literature: the one-stage and two-stage (Daraio & Simar, 2005).

	One-stage approach	Two-stage approach
Description	The environmental variables are included as free disposal inputs (if they are advantageous to the efficiency) or outputs (if they are unfavourable to the efficiency). They are used to define the set of all the feasible activities, called the production set, but they are not active in the optimisation process applied to define the efficiency scores.	The first stage of this approach consists in computing the efficiency scores only with the input-output matrix. Then, a regression analysis is undertaken with the efficiency scores as dependent variable and the characteristics of the operational environment as explanatory variables. The most common method is the bootstrap algorithm proposed by Simar & Wilson (1998; 2007).
Pros	(+) The contextual variables are taken into account when computing the efficiency scores.	(+) The direction and the intensity of the impact of environmental variables on the efficiency scores are evaluated.
Cons	(-) The role of the environmental variables must be known a priori. (-) The free disposability is assumed. This property means that increasing inputs quantities is always possible without reducing the outputs (Liu, et al., 2010)	(-) It requires the separability condition. It means that the environmental variables have only an influence on the distribution of inefficiencies and not on the position and the shape of the frontier (Daraio, et al., 2015) (-) It requires parametric assumptions.

Figure 6-11: One-stage and Two-stages approach (Source: Personal)

6.3.2 Step 7: Robustness and sensitivity

All those choices concerning the model to use and the treatment of undesirable outputs as well as environmental variables might affect the credibility of the CI. Therefore, as explained in the reference book, carrying out robustness and sensitivity analyses might improve the transparency of the CI. The robustness analysis allows to determine ranges for the efficiency scores and the sensitivity analysis attempts to indicate how much uncertainty in the CI is reduced if a source of incertitude is removed. The objective is thus to identify all the causes of uncertainty (such as the selection of indicators, treatment of missing values, data quality, model, …) and to evaluate their impact on the final results.

6.4 Stage 4: Interpretation

To interpret those results, the CI might be disaggregated, related to other variables and finally represented through a suitable visualisation. There are no specific tools to DEA for those steps. Therefore, this section only provides a summary of the explanations given in the reference book.

6.4.1 Step 8: Back to the data

The authors recommended decomposing the CI in its various components with the aim of enhancing its transparency. This will highlight the differences between the DMUs regarding their inputs, outputs as well as environmental variables and reveal what drives the scores.

6.4.2 Step 9: Links to other indicators

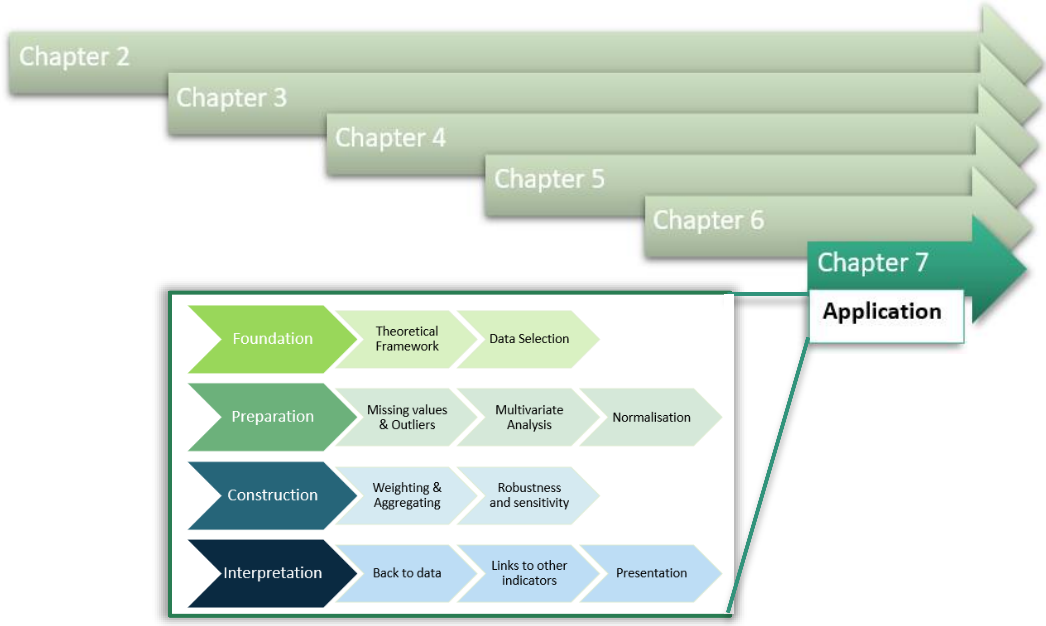
After analysing the internal structure, the CI should be linked to well-known indicators which are related to the concept measured. A simple cross-plot is sufficient to already observe a potential correlation between the indicators.

6.4.3 Step 10: Presentation and dissemination

Eventually, as quoted in the reference book, “a well-designed graph can speak louder than words”. Consequently, the choice of the visualisation tool is important to communicate the most information.

Chapter 7: Application (Contribution 2)

Now it is time to put the theory into practice by applying the methodology previously described to a concrete case related to the ANSPs' performance. The results and the visualisations are obtained from the R programming language and many of its packages including "Benchmarking" (Bogetoft & Otto, 2020) and "rDEA" (Simm & Besstremyannaya, 2020).



7.1 Stage 1: Foundation

7.1.1 Step 1: Theoretical Framework

The purpose of this master thesis is to create a cost-efficiency CI of ANSPs in Europe and to evaluate the impact of their operational environment on their performance.

7.1.1.1 Definition

The cost-efficiency indicator aims to reflect the level of optimal cost management of an ANSP to deliver services under environmental constraints. The final objective is to enable the managers to observe the performance gaps between the organisations and to open lines of thought for future improvement.

Firstly, the scope of the benchmarking analysis must be clarified to determine the outputs and the inputs to integrate. Amongst the services offered by the ANSPs, the to-be indicator will only focus on the provision of ATM/CNS services on continental territories. This activity generates different types of costs among which only the expenses controllable by the managers will be included in the construction of the CI.

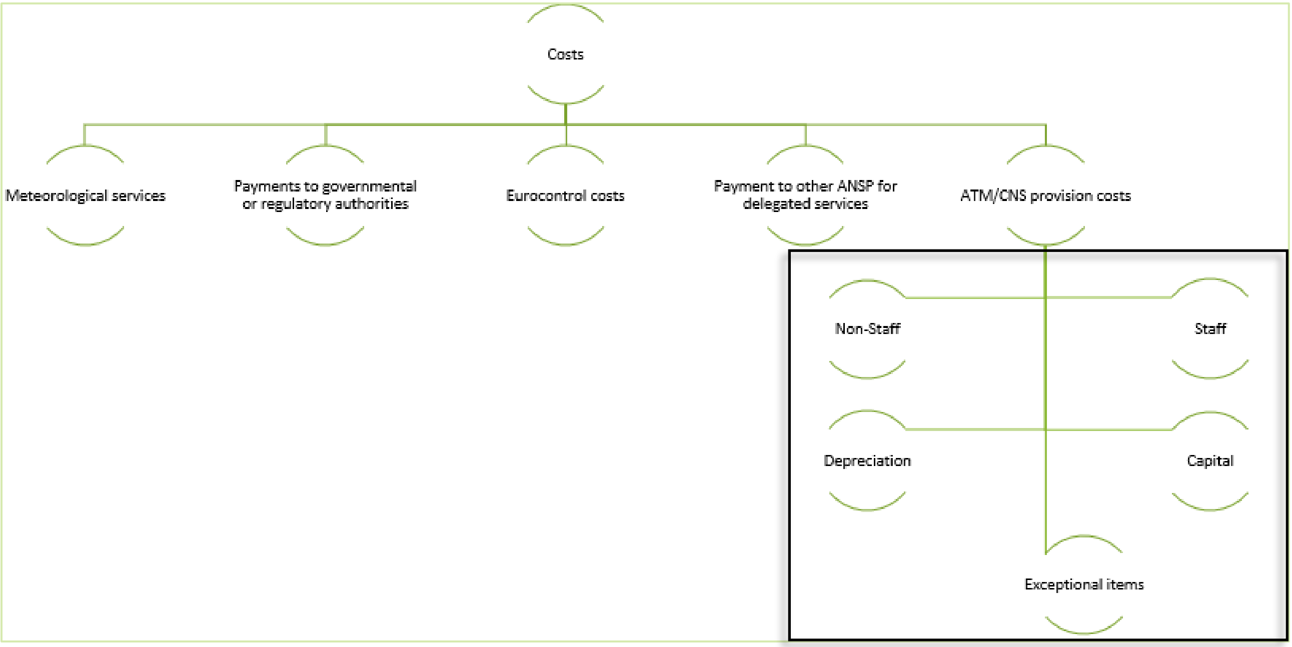


Figure 7-1: Cost structure of ANSPs (Adapted from PRU)

The expenses, others than ATM/CNS provision costs, are excluded for several reasons. On the one hand, for MET services, some ANSPs appeal to their national meteorological institutions, whereas others perform these activities in-house. On the other hand, "the payments to governmental or regulatory authorities", "EUROCONTROL costs" and "Payment to other ANSP for delegated services" are beyond ANSPs' control.

On the contrary, the managers have decision-making power regarding the ATM/CNS provision costs which encompass the expenses associated with the staff, their capital, non-staff operating and exceptional items.

7.1.1.2 Subgroups

Consequently, the ATM/CNS activities require inputs, such as human and material resources, to be performed. The output produced is the management of aircrafts in the sky at each stage of the flight. Finally, the ANSPs' performance might vary from one to another since the ATCOs have to adapt to a set of environmental factors pinpointed by the PRU.

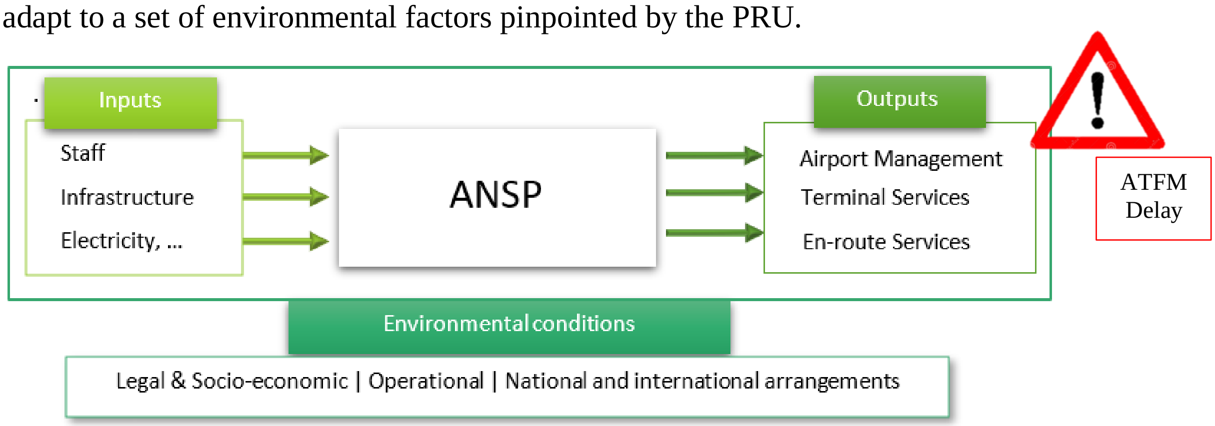


Figure 7-2: ATM/CNS activity process (Source: Personal)

From this schema (see Figure 7-2), it is easier to pick the variables to retain. The input prices were discarded due to the lack of accuracy.

Variables	Description	
Inputs	1. Staff 2. Capital	3. Others
Outputs	1. Outputs for En-route 2. Outputs for Terminal	3. Quality
Environmental	1. Legal & socio-economic conditions 2. Operational conditions 3. National and International arrangements	

Table 7-1: Components to consider in the analysis (Source: Personal)

For each component, there are many potential indicators. Therefore, a literature review is undertaken to build a comparative table summarizing the outputs, inputs and environmental variables integrated into the studies previously carried out in this field (see *Appendix 3*).

7.1.2 Step 2: Data selection

This table facilitates the selection process of indicators. A first sorting was made based on the most used variables in the models. Then, the potential indicators are analysed through the Quality Framework. The criteria *Timeless* and *Accessibility* are examined at the level of the data sources whereas the others (*Accuracy, Relevance, Interpretability and Coherence)* are studied at the indicator level. A table summing up the comments is attached to this document in *Appendix 4.*

7.1.2.1 Data sources

Most of the data are provided by the PRU that collects them directly from the ANSPs. Nevertheless, the Purchasing Power Parity (PPP) is needed to adjust the monetary values with the aim of taking the cost-of-living into account. This index is computed by both Eurostat and International Monetary Fund (IMF).

Regarding the *Timeless*, the information has to be submitted to the PRU by the participating ANSPs "by the 1st of July in the year following the year to which it relates" (Performance Review Unit, 2019). Therefore, the data are gathered six months after the concept occurs. Even though this time period seems long, it is necessary. Indeed, as seen at skeyes, the data are recorded, and they are adjusted later. Finally, the information, reported by the ANSPs, is "subject to an analysis and verification process" (Performance Review Unit, 2019). Concerning the PPP indicators of Eurostat, they are compiled on an annual basis and are published in four steps (see Figure 7-3) (Eurostat, 2020). For the IMF, the *World Economic Outlook* historical data are continuously updated and the estimations remain until complete information is available (International Monetary Fund, 2020).

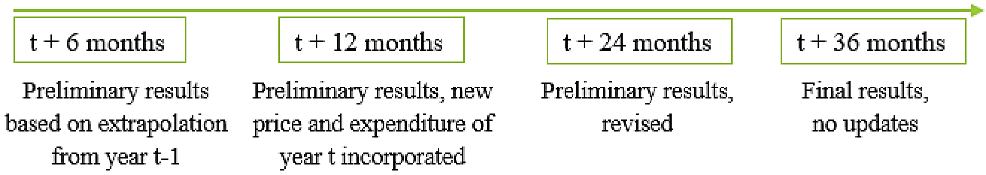


Figure 7-3: Four steps of the publication of PPPs (Adapted from EUROSTAT)

From an *Accessibility* perspective, the PRU sample is released in paper format in the yearly ACE report. To complete this project, a member of the PRU sent the data in electronic form by email. For the PPP indicator, it is accessible to the public on Eurostat and IMF websites.

7.1.2.2 Selected data

After assessing the different options, the interpretability proved to be essential to improve the transparency of the CI. By consequence, the better solution is to use all the costs instead of items (ATCO hours, Full Time Equivalent, …) as inputs and to see the information the CI will provide. For the environmental variables, the PRU introduced indicators to measure four exogenous factors: the size of ANSP, adjusted density, structural complexity and seasonal traffic variability. In addition, the ANSPs' ownership will be considered as well.

Components	Indicators for the period 2014-2018	
Inputs	1. Staff costs 2. Non-staff operating costs 3. Depreciation costs	4. Cost of capital 5. Exceptional costs
Outputs	1. IFR flight hours 2. IFR airport movements	3. ATFM delays (undesirable outputs)
Environmental variables	1. Traffic Variability 2. Network Size 3. Adjusted density	4. Structural complexity score 5. Ownership (Commercialised, Public or Private)

Table 7-2: Indicators used as foundation for the CI (Source: Personal)

7.2 Stage 2: Preparation

Those selected indicators are transformed before being combined into a single index. The treatment of blank entries, the detection of outliers and the pre-analyses are performed at this stage.

7.2.1 Step 3: Dealing with missing data and outliers

7.2.1.1 Missing values

In this study case, the datasets from the PRU and from Eurostat are incomplete. Those observations are Missing Completely At Random.

Indeed, in the PRU sample, the Georgian ANSP “Sakaeronavigatsia” has blank entries in 2014 since it joined only in 2015. This item is removed from the dataset because, given that the entire row is missing, it is impossible to approximate one single value for each variable. Furthermore, the aim of this master thesis is to create an indicator which is as fair as possible. Therefore, the technic suggested by Kuosmanen (2009) (see Table 6-3) was also discarded. Finally, the choice is motivated by the fact that this ANSP is present after 2015, so it is possible to get an efficiency score for this ANSP and the ranking of the other DMUs will not be impacted for the recent year.

Concerning the PPP, some countries included in the PRU sample are not part of the Eurostat database. This issue has already been investigated by the PRU that uses information contained in the IMF database to calculate the missing values. Nevertheless, an adjustment must be made to turn the PPP index from the IMF database into PPP from Eurostat. As a matter of fact, the latter is described as a currency conversion rate which allows to convert national currencies to an artificial one called the Purchasing Power Standard (PPS). In the IMF database, “the PPP index is expressed in local currency per international dollar rather than PPS” (Performance Review Unit, 2019). To solve this problem, the PRU based its reasoning on the assumption that “the difference in PPPs between two countries shall be the same in the Eurostat and IMF databases” (Performance Review Unit, 2019).

For instance, first, we divide PPP_IMF of Armenia (199.859), which is missing in Eurostat dataset, by PPP_IMF of Albania (42.996) and we obtain a factor (4.65). Then, we multiply 4.65 by the PPP_EUROSTAT of Albania (60.46) and the result is 281.03. This procedure is followed for all the other countries and the median of all the results (2018_Factors_Armenia) is computed to get the PPP_EUROSTAT of Armenia.

	Country	2018_IMF
Armenia	Armenia	199.859

	Country	2018_IMF	2018_Division_Armenia	2018_EUROSTAT	2018_Factors_Armenia
Albania	Albania	42.996	4.648316	60.458100	281.0284
Austria	Austria	0.836	239.065789	1.127840	269.6280
Belgium	Belgium	0.820	243.730488	1.131490	275.7786
Bosnia and Herzegovina	Bosnia and Herzegovina	0.697	286.741750	0.987565	283.1761
Bulgaria	Bulgaria	0.663	301.446456	1.012580	305.2387

Figure 7-4: Example of the computation of Armenia PPP (Source: Personal)

7.2.1.2 Adjustment

Then, the monetary values are adjusted thanks to the dataset of Eurostat. Before doing any transformations, it is important to know that the PPP is expressed in national currency per PPS. The costs collected by the PRU, though, are already converted into euros and in 2018 prices. The adjustment factor is, thus, estimated by dividing the exchange rates by the PPP index.

$$Adjustment\ factor = \frac{Exchange\ rate\_2018\ (National\ currency/€)}{PPP\_2018\ (National\ currency/PPS)}$$

$$Adusted\ cost = Costs * Adjustment\ factor$$

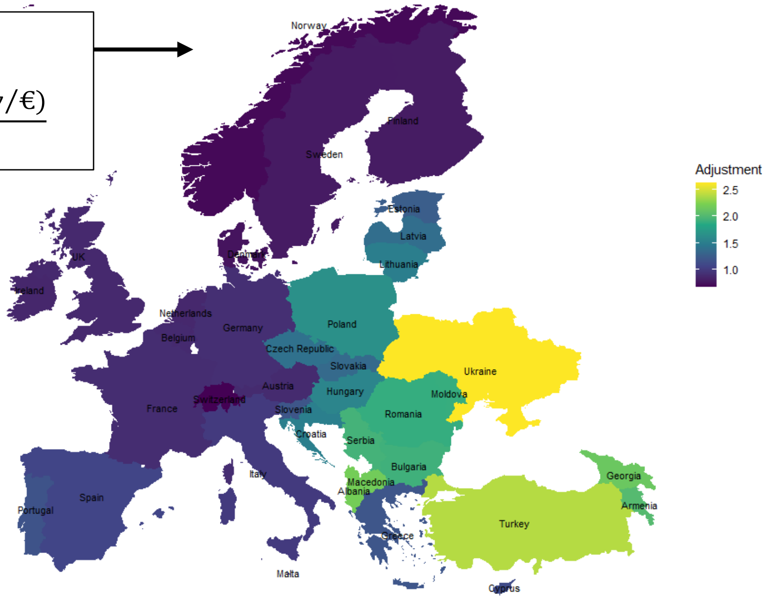


Figure 7-5: Adjustment factor for each country (Source: Personal)

7.2.1.3 Outliers

The dataset is now complete which enables us to undertake an analysis in order to detect atypical observations. The first step consists in generating a scatter plot matrix where outliers might be spotted by the combination of two variables. Then, the data cloud method is applied to highlight groups of outliers. It is executed, at first, on the whole dataset which contains values from 2014 to 2018 and, then, on the most recent data of 2018 (see Table 7-3).

The algorithm detected, on the dataset 2014-2018, the three ANSPs with the highest costs and the highest level of outputs generated: DSNA, DHMI and DFS. This assumption can be proved by visualising the data through PCA and clustering.

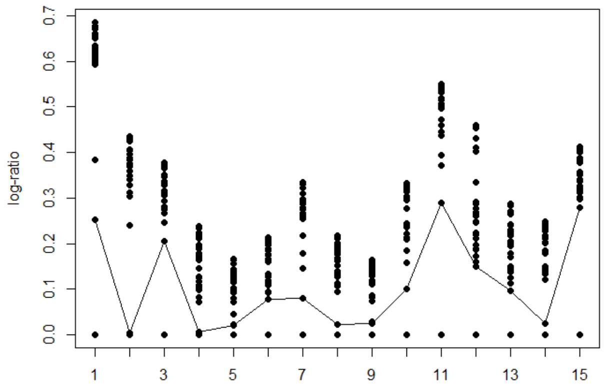
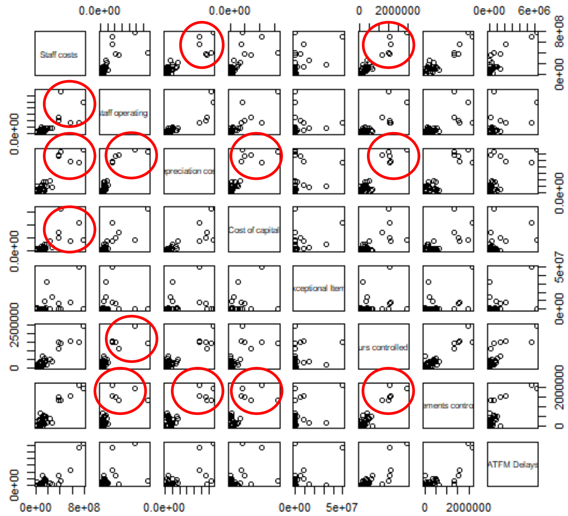
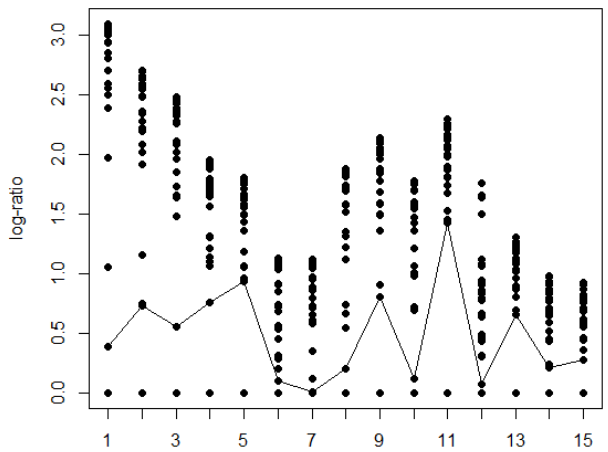
<table>
<tr><th></th><th>Scatter plot</th><th colspan="3">Outliers</th></tr>
<tr><td rowspan="4">2014
-
2018</td><td rowspan="4">
</td><td rowspan="4">
</td><td>1</td><td>DFS_2018</td></tr>
<tr><td>3</td><td>DSNA_2018, DHMI_2018, DFS_2018</td></tr>
<tr><td>11</td><td>DHMI_2014, DHMI_2016, DSNA_2017, DSNA_2015, DSNA_2016, DHMI_2017, DHMI_2015, DSNA_2014, DSNA_2018, DHMI_2018, DFS_2018</td></tr>
<tr><td>15</td><td>DHMI_2014, DHMI_2016, DFS_2014, DSNA_2017, DSNA_2015, DFS_2017, DFS_2016, DSNA_2016, DHMI_2017, DHMI_2015, DSNA_2014, DSNA_2018, DHMI_2018, DFS_2018, DFS_2015</td></tr>
<tr><td rowspan="4">2018</td><td rowspan="4">
</td><td rowspan="4">
</td><td>1</td><td>DFS</td></tr>
<tr><td>5</td><td>ROMATSA, UkSATSE, DSNA, DHMI, DFS</td></tr>
<tr><td>9</td><td>Austro Control, ROMATSA, ENAIRE, ENAV, NATS (Continental), UkSATSE, DSNA, DFS, DHMI</td></tr>
<tr><td>11</td><td>Skyguide, Sakaeronavigatsi Austro Control, ROMATSA, ENAIRE, ENAV, NATS (Continental), UkSATSE, DSNA, DFS, DHMI</td></tr>
</table>

Table 7-3: Outliers detection (Source: Personal)

7.2.2 Step 4: Multivariate Analysis

7.2.2.1 PCA

The PCA is conducted with the aim of reducing the dimensionality of the dataset to observe the interrelationships between the indicators through 2D and 3D representations. Those graphs are obtained by plotting the Principal Components (PC) which are linear combinations of the variables.

It is first carried out on the input-output matrix (the costs considered separately) of 2018. Those figures (Figure 7-7 and Figure 7-6) are reliable because the first and second PCs capture more than 80% of the information provided by the indicators. As predicted, DHMI, DSNA and DFS are outliers since they have a higher level of costs and outputs than the ANSPs agglutinated in the centre. Even though the monetary values are not adjusted with the “Adjustment Factor”, the outliers are the same. The gaps are just bigger between, for instance, DFS and UkSATSE.

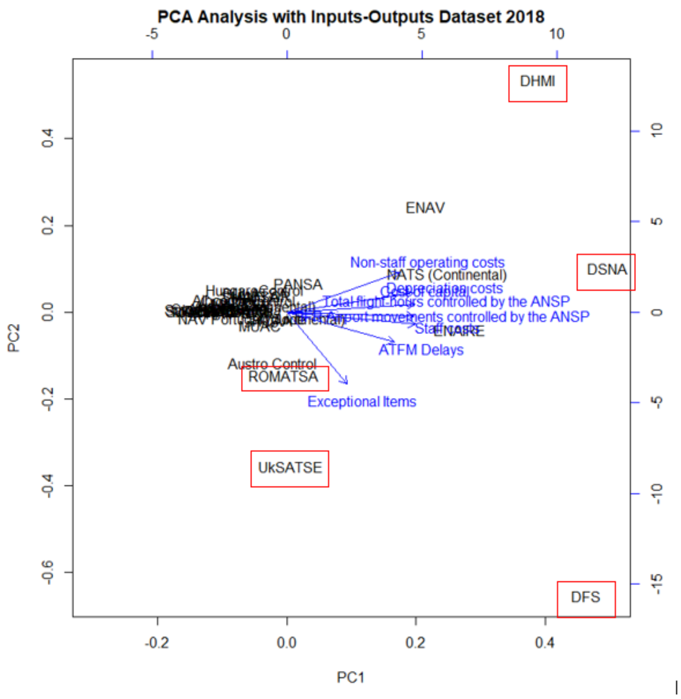


Figure 7-7: PCA with inputs and outputs data from 2018 with Adjustment factor (Source: Personal)

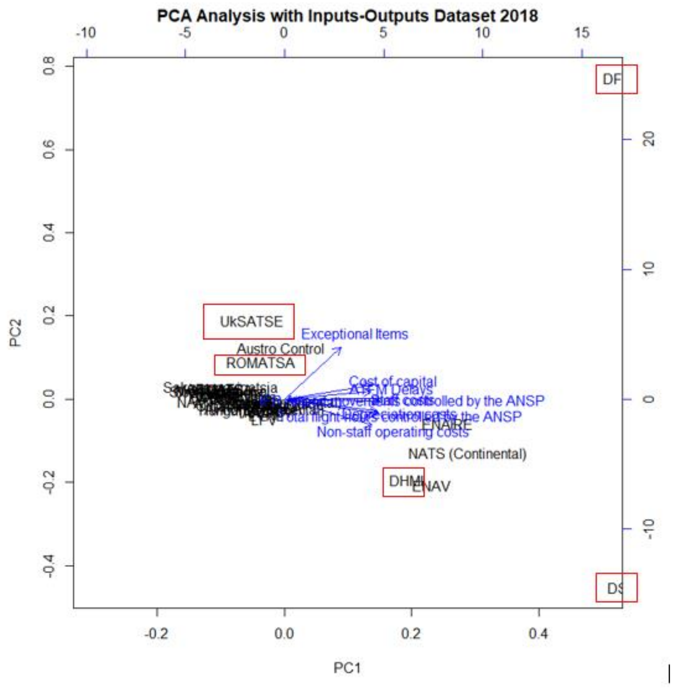


Figure 7-6: PCA with inputs and outputs data from 2018 without the Adjustment factor (Source: Personal)

The PCA is also performed on the four quantitative environmental factors to bring out the predominant characteristics of each ANSP. For instance, as expected, skeyes is featured by a complex environment. Indeed, the variable “Structural” pulls skeyes above on the right. Another example is ENAIRE that controls the biggest area. The 3D graph is more relevant as the three PCs explain together 87.75% of the variance against 65% for the first two PCs.

Therefore, it is more obvious on the 3D representations (Figure 7-8, Figure 7-10, Figure 7-11) that DHMI and DSNA, considered as outliers, are characterised by a larger airspace and a higher density, whereas DFS is operating in a complex airspace in terms of "Density" and "Structural".

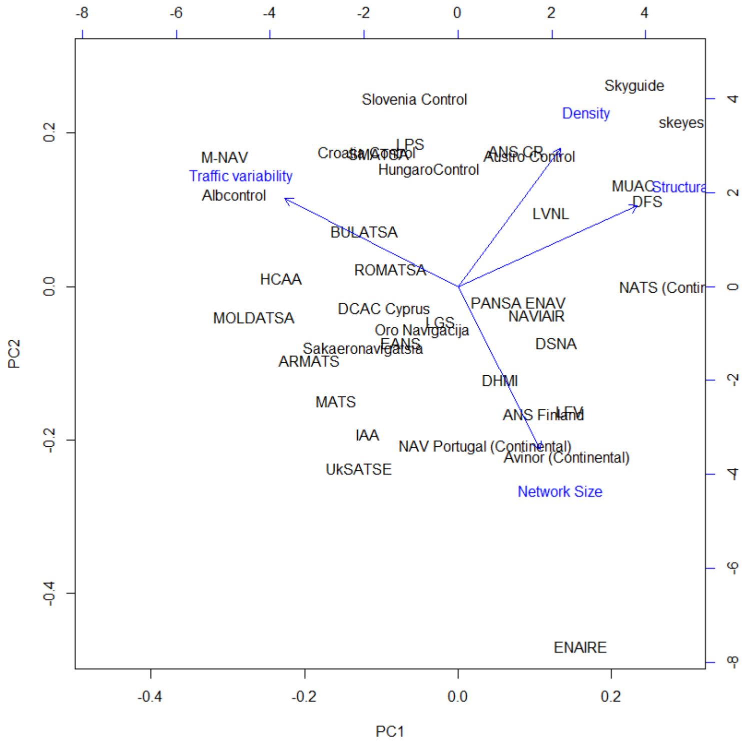


Figure 7-9: PCA analysis on four environmental variables in 2018 (Source: Personal)

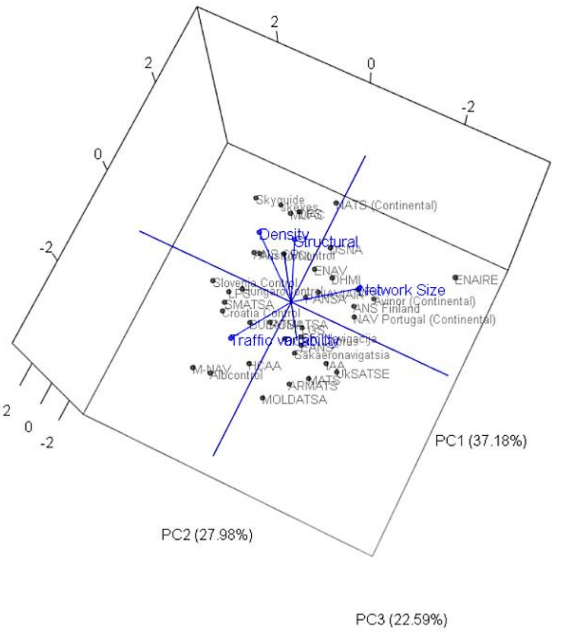


Figure 7-8: 3D PCA representation on four environmental variables in 2018 (Source: Personal)

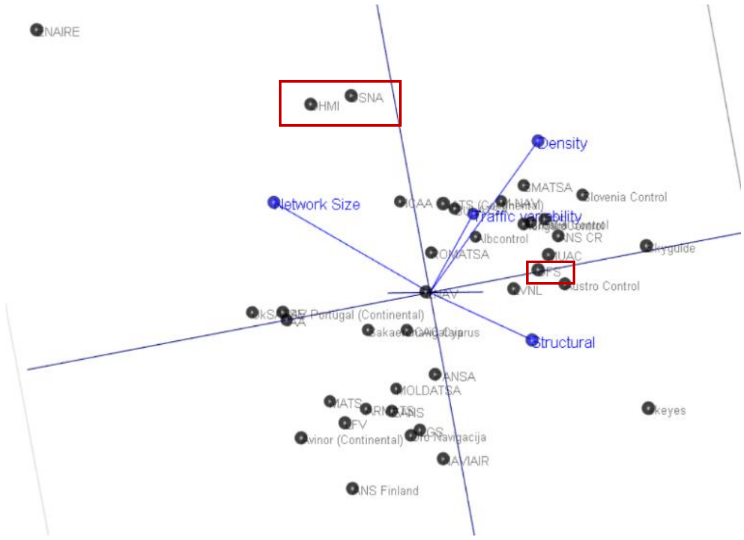


Figure 7-11: Zoom on the 3D PCA representation (Source: Personal)

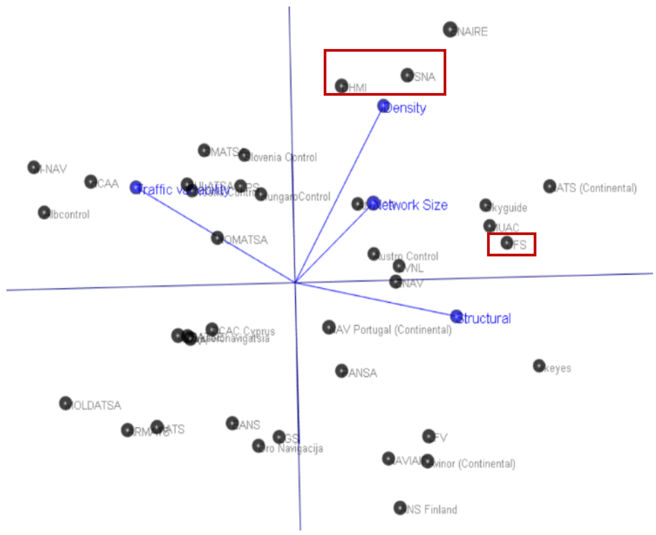


Figure 7-10: Zoom under another angle of 3D PCA representation (Source: Personal)

7.2.2.2 Clustering

After exploring the links between the indicators, it could be interesting to uncover if the ANSPs might be grouped according to their production process (inputs/outputs) or to their environment. The clustering analysis can be implemented through different algorithms such as hierarchical clustering (dendrogram), K-Means, DBSCAN, … In this section, the two most common technics, the dendrogram and the K-Means clustering analysis, will be run.

Dendrogram

To build the dendrogram, it is indispensable to compute the distance between the observations. The most famous one is the Euclidean distance. This formula works only with numerical values, therefore, the status of ANSPs were ignored.

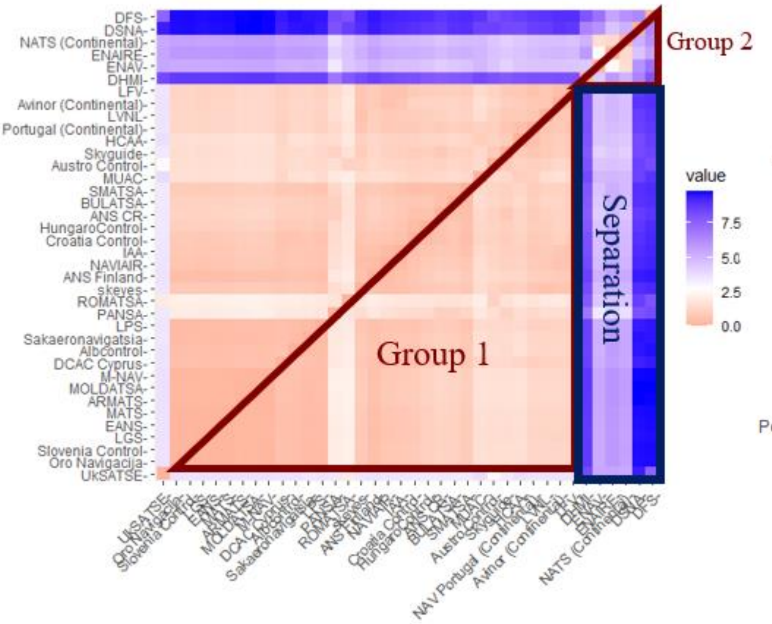


Figure 7-13: Euclidean distance on the input-output matrix of 2018 (Source: Personal)

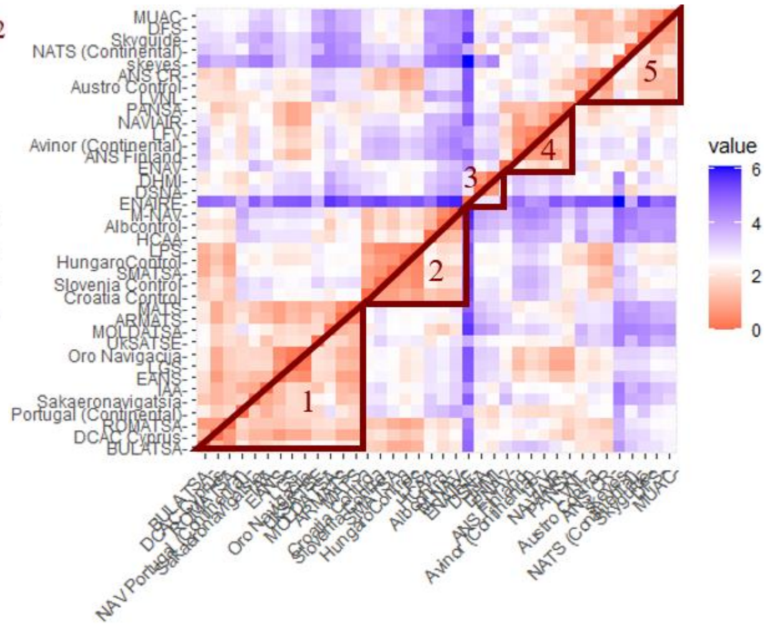


Figure 7-12: Euclidean distance on the environmental factors of 2018 (Source: Personal)

On the Figure 7-13, it seems that the trend on DSNA, DHIM and DFS is, once again, confirmed. According to the Euclidean distance, they seem to be apart from each other and also from the other ANSPs when only the input and output variables are taken into account. Furthermore, NATS, ENAIRE and ENAV form a smaller group far from the other ANSPs. Finally, UkSATSE, ROMATSA and PANSA are slightly less similar. Almost the same groups are observed as well when the costs are not transformed with the "Adjustment Factor".

The same analysis is conducted on the numerical environmental factors. The clusters are less well-defined. To ease the interpretation, the clustering of the ANSPs according to the characteristics of their operational environment are displayed by means of dendrogram based on the Euclidean distances.

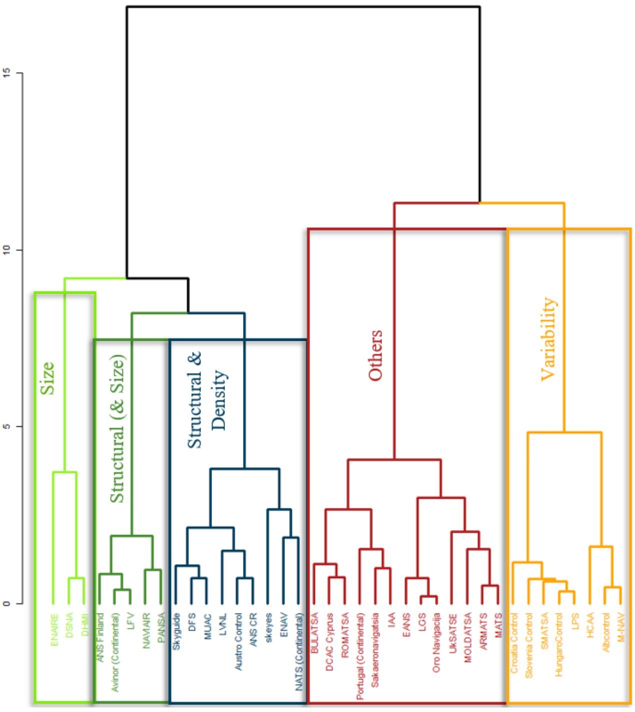


Figure 7-16: Dendrogram of Euclidean distance

(Source: Personal)

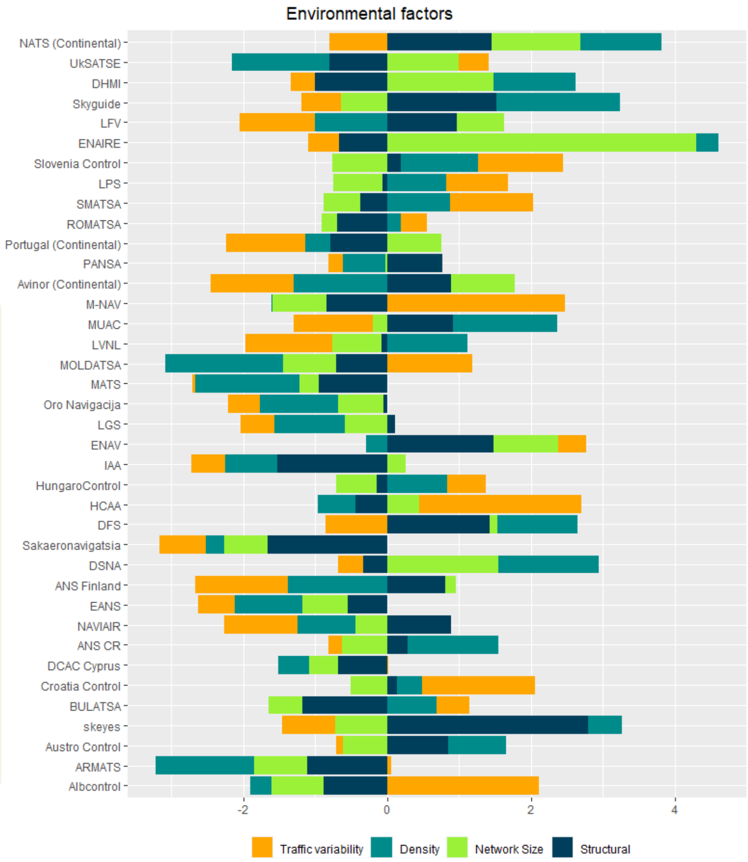


Figure 7-14: Scaled environmental factors used to compute the Euclidean distances (Source: Personal)

The dendrogram allows us to better visualise the groupings created from the computation of the Euclidean distances. The clusters are labelled thanks to the 3D PCA representations generated previously and to the Figure 7-14 plotting the scaled environmental data. More dendrograms are available in *Appendix 5* where the Daisy and Manhattan distances are calculated.

K-Means

The other clustering algorithm run in this case study is the K-Means analysis. This method is an iterative procedure which aims to find *k* clusters. It is also based on the Euclidean distance but the computation technic is different. Before implementing the K-Means algorithm, the elbow method is applied to determine the number of clusters to choose. Regarding the input-output matrix, 2 clusters seems to be the optimal number and for the numerical environmental factors, 5 clusters are considered. On the Figure 7-18, two groups are clearly separated. Finally, on the other figure, the groups are almost the same as discovered from the dendrogram analysis. Just some ANSPs moved to another group.

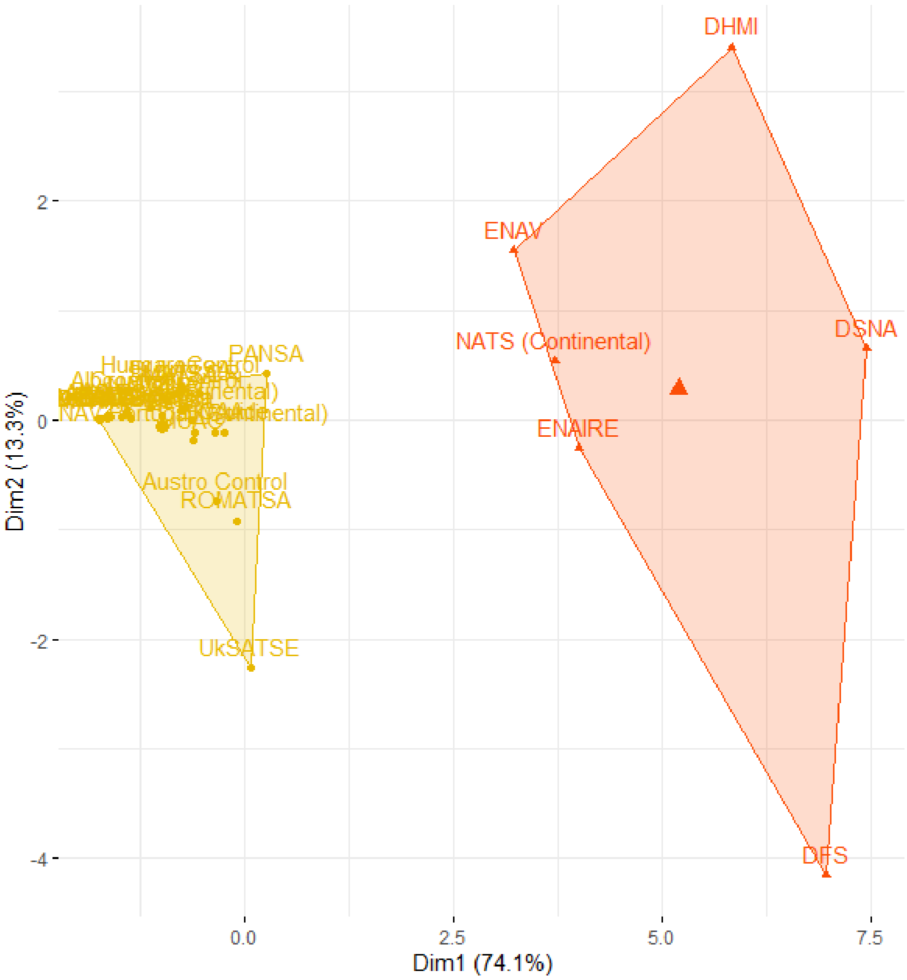


Figure 7-18: K-Means on the input-output matrix from dataset 2018

(Source: Personal)

Figure 7-17: K-Means on the numerical environmental variables from dataset 2018
(Source: Personal)

7.2.2.3 Dimensionality reduction

In this study case, the dimensionality reduction is unnecessary because the number of inputs-outputs variables is at maximum equal to 8 (Staff costs, Non-staff operating costs, Depreciation costs, Cost of Capital, Exceptional items, IFR hours controlled, IFR airport movements and ATFM delays) and the number of observations per year is 38. Thus, the rule established by Golany & Roll (1989) and Banker, et al. (1989) are respected.

7.2.3 (Step 5: Normalisation)

This step is skipped as well because the DEA estimates are independent of the units of measurement and take into account the differences between scale among ANSPs.

7.3 Stage 3: Construction

7.3.1 Step 6: Weighting and Aggregating

Now that the indicators are analysed and transformed, they can be combined into a single index. The table below (Table 7-4) shows the statistics of the data for 2018 which will be used to execute the DEA. Aware of the drawbacks of DEA, all the costs are gathered in one variable (Total Costs adjusted).

Variable	Min	Median	Mean	Max	Categories
Total Costs adjusted	17,049,366	146,448,116	248,468,972	1,204,728,754	Input
IFR flight hours controlled	12,452	277,771	450,679	2,458,363	Output
IFR airport movement	0	195,111	434,230	2,129,744	Output
ATFM Delays	0	148,736	652,926	6,300,231	Undesirable output

Table 7-4: Statistics of the dataset of 2018 (Source: Personal)

To determine the type of model to implement, two assumptions are made. The first one concerns the presence of economies of scale, suggested by the models applied by the CEG (2011) and Bilotkach, et al. (2015). The second is related to the outputs, considered as uncontrollable because it depends on the airspace users' needs. Those two hypotheses lead to the use of the *BCC input-oriented* model.

7.3.1.1 BCC input-oriented DEA model without ATFM delays

For the first DEA, the quality variable, ATFM delays, is ignored as the PRU did by developing the « Financial Cost-Effectiveness » indicator.

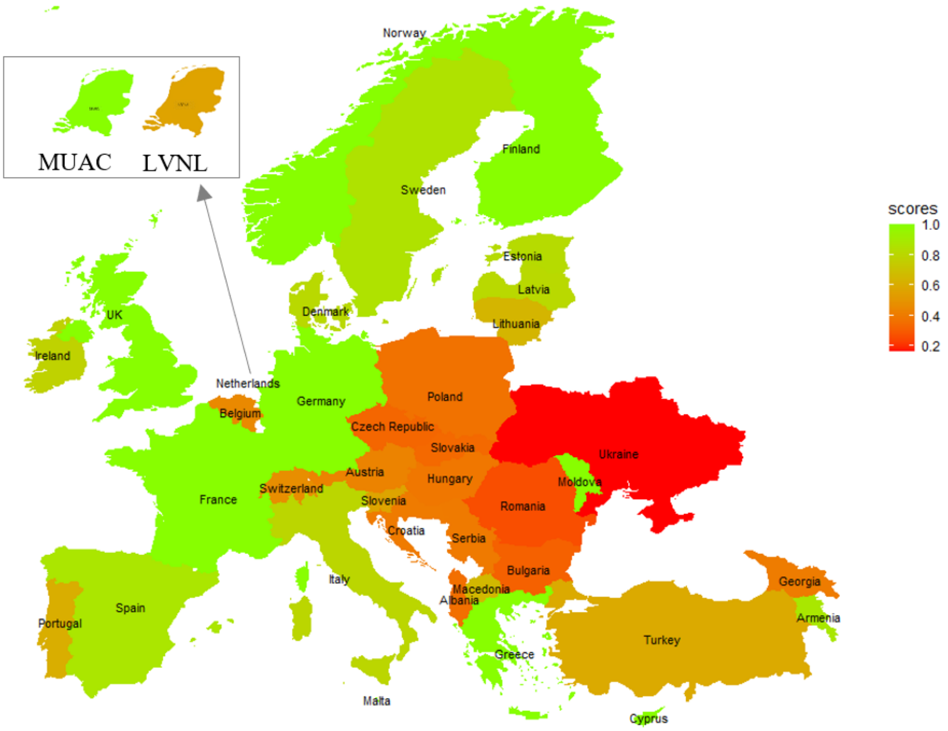


Figure 7-19: BCC input-oriented DEA efficiency scores (Source: Personal)

This graph (Figure 7-19) only shows the pure technical efficiency scores. It seems that there are spatial autocorrelation issues, as mentioned by Neiva (2014), where the performance of an ANSP is influenced by the level of efficiency of those around. But the grouping of inefficient ANSPs might be also due to the type of territory handled. Indeed, the countries on left side manage more maritime environment whereas the others on the right are located inland.

The pure technical efficiency is one form of performance measurement. With DEA, it is possible to have an idea of the levels of scale and mix efficiencies (see *Appendix 6*).

ANSP	Pure Technical	Scale		Mix	
ANSP	Pure Technical	Scale	RTS	Slack En route	Slack Terminal
skeyes	0.454	0.962	IRS	72669.67	0.00

Table 7-5: Decomposition of efficiency for skeyes (Source: Personal)

For instance, skeyes has a **pure technical efficiency** of 45.4% which expresses a reduction rate. According to the data, it is possible to decrease the costs by 54.6% without worsening the amount of outputs produced (costs improved = 146620358 – 0.546*146620358).

The **scale efficiency** is obtained by dividing the efficiency scores computed with *CCR model* by the ones from *BCC model*. It indicates that skeyes did not reach its optimal scale size and it could take advantages of economies of scale (IRS = Increasing RTS).

Lastly, the slacks, associated to output shortfalls, capture the **mix-inefficiencies**. skeyes could be more efficient by increasing the number of IFR hours controlled at the same costs. However, the pure technical efficiency is the most interesting one. Indeed, it is impossible to increase or decrease the scale of their production and to change the output mix.

In conclusion, even under the best set of weights, skeyes is classified as "inefficient" because of two ANSPs located on the "best practice" frontier: ANS Finland and Avinor (Continental) (for the other ANSPs see *Appendix 7*). It means that the radial projection of skeyes on the frontier is situated between those two ANSPs. In other words, the virtual ANSP to which skeyes is compared is a combination of ANS Finland and Avinor (Continental) (see Table 7-7).

skeyes' reference set	ANS Finland	Avinor (Continental)
Costs adjusted	48879345.8	112581418
IFR Airport movements	262327	650295
IFR hours controlled	124274.25	364765
Lambdas (see Table 6-5)	0.72344626	0.27655374

Table 7-6: Reference set of skeyes (Source: Personal)

<table>
<tr><th colspan="3">Virtual ANSP</th></tr>
<tr><td rowspan="2">Costs</td><td>=skeyes' efficiency scores * costs
= 0.454 * 146620358</td><td rowspan="2">6649391.9</td></tr>
<tr><td>= lambda ANS Finland * costs Finland + lambda Avinor * costs Avinor
= 0.723 * 48879345.8 + 0.277 * 1125811418</td></tr>
<tr><td rowspan="2">IFR Flight hours</td><td>= skeyes' Output + slack En route
= 118113.2 + 72669.67</td><td rowspan="2">190782.87</td></tr>
<tr><td>= lambda ANS Finland * Output Finland + lambda Avinor * Output Avinor
= 0.723 * 124274.25 + 0.277 * 364765</td></tr>
<tr><td rowspan="2">IFR airport movements</td><td>= skeyes' Output + slack Terminal
= 369621 + 0</td><td rowspan="2">369621</td></tr>
<tr><td>= lambda ANS Finland * Output Finland + lambda Avinor * Output Avinor
= 0.723 * 262327 + 0.277 * 650295</td></tr>
</table>

Table 7-7: Virtual ANSPs to which skeyes is compared (Source: Personal)

7.3.1.2 Undesirable output

The second indicator introduced by the PRU, called the "Economic Cost-Effectiveness", integrates the ATFM delays by multiplying it by 104€ in 2018. To avoid resorting to a fixed parameter (104 €), the efficiency scores will be obtained by transforming the undesirable outputs, ATFM delays expressed in minutes, to use the traditional DEA. Two methods are tested. Firstly, the ATFM delays are considered as inputs (see Figure 7-21). Secondly, the solution suggested by Seiford & Zhu (2002) is applied ($ATFM\ delays * -1 + M > 0$).

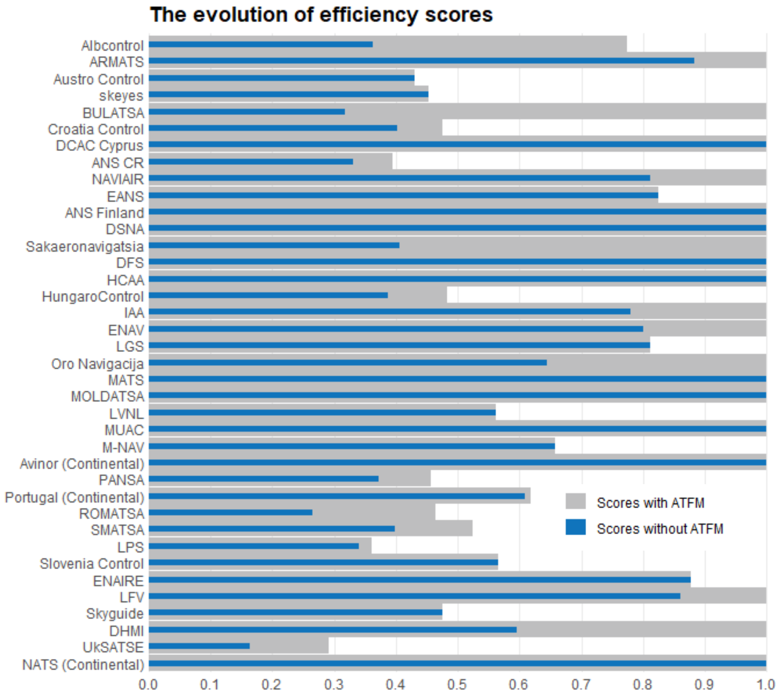


Figure 7-21: Evolution of the efficiency scores when the ATFM delays are considered as inputs (Source: Personal)

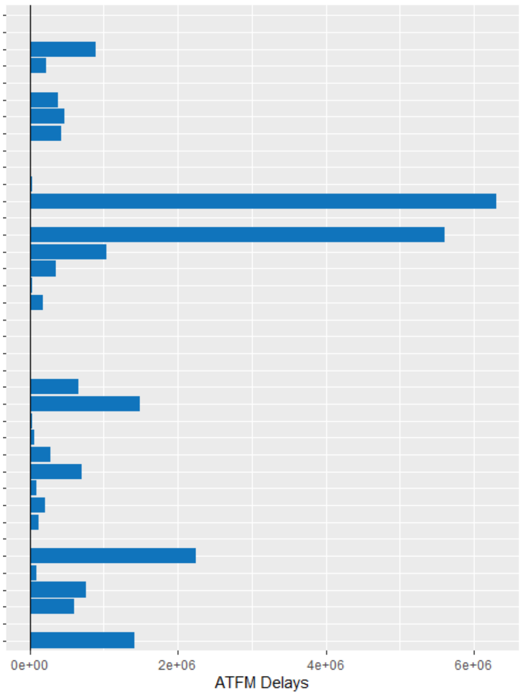


Figure 7-20: ATFM Delays in minutes without transformations (Source: Personal)

The inclusion of ATFM delays, as inputs or as desirable outputs, makes many ANSPs efficient (ARMATS, BULATSA, NAVIAR, Sakaeronavigatsia, IAA, ENAV, Oro Navigacija, LFV and DHMI) which, for the majority, records 0 ATFM delays. The classification between efficient and inefficient ANSPs is the same for both technics. However, the scores for the inefficient ANSPs are slightly higher when the ATFM delays are taken as inputs.

7.3.1.3 Environmental variables

The level of inefficiency might be partially due to exogenous factors. By grouping the ANSPs according to their predominant characteristics, it seems that the traffic variability and the network size have, respectively, a negative and positive impact on the cost-efficiency scores, without taking the ATFM delays into account. Concerning the structural aspect of the airspace, the direction and the intensity of the effects are less obvious.

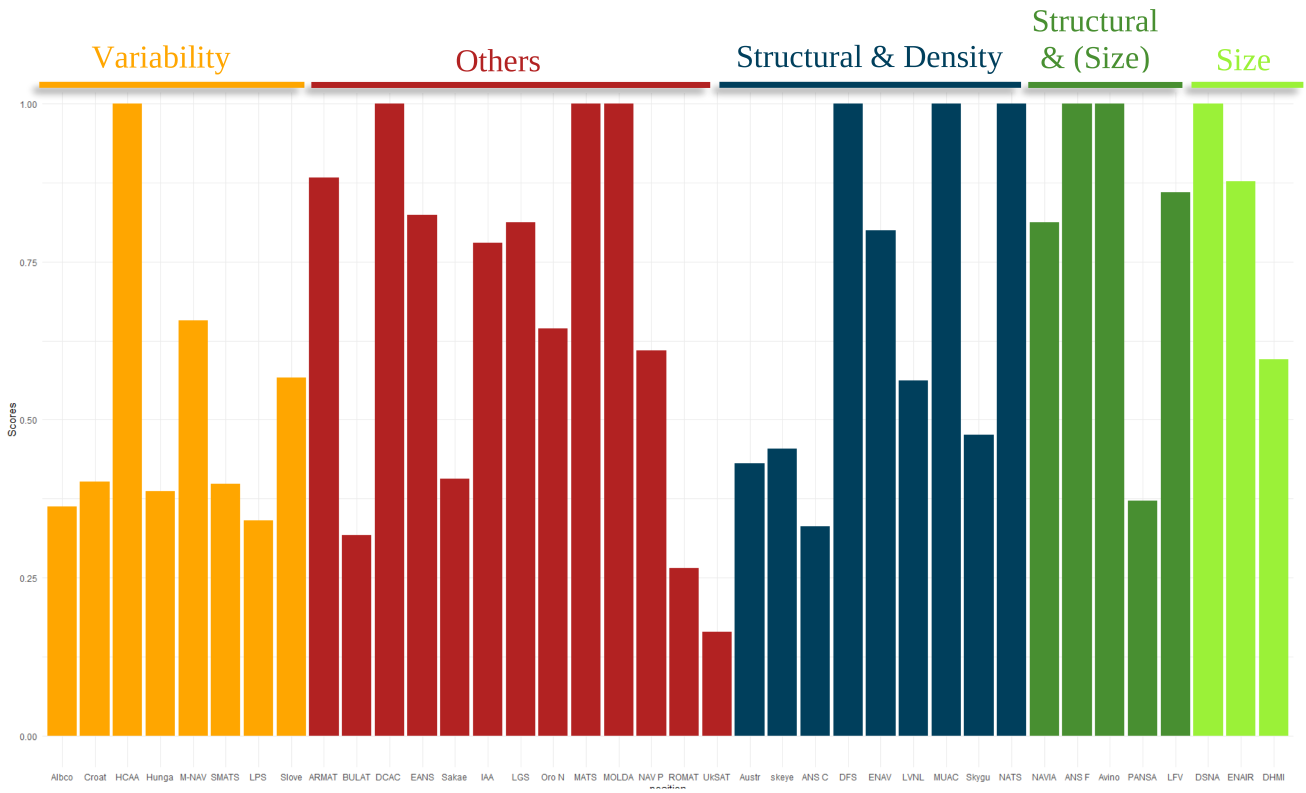


Figure 7-22: Efficiency scores grouped according to the clusters detected by the Dendrogram analysis (Source: Personal)

The ANSPs might be also grouped depending on their ownership and the type of territory they manage. Concerning ANSPs' status, the sample of private and public entities is small given that the dataset is unbalanced. It is, thus, impossible to compute separate best-practice frontier for each group. However, it seems that the private and public ANSPs performed better in 2018 than the commercialised. This conclusion is weak due to the lack of representativity of public and private ANSPs.

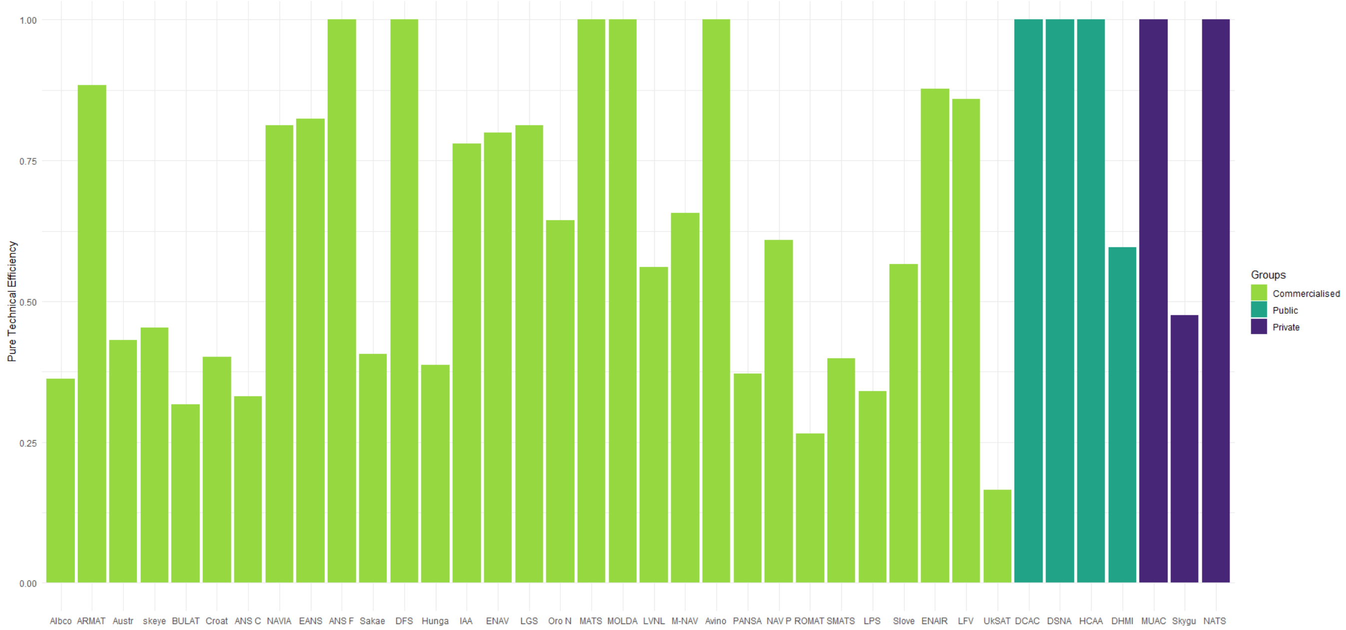


Figure 7-23: Efficiency scores grouped according to their ownership (Source: Personal)

About the type of territory under the supervision of the ANSPs, it seems that the ANSPs which manage more maritime territories (Sea) have in general higher pure technical efficiency estimates than the ANSPs located inland (Land). This trend is accentuated when the ATFM delays are integrated as inputs.

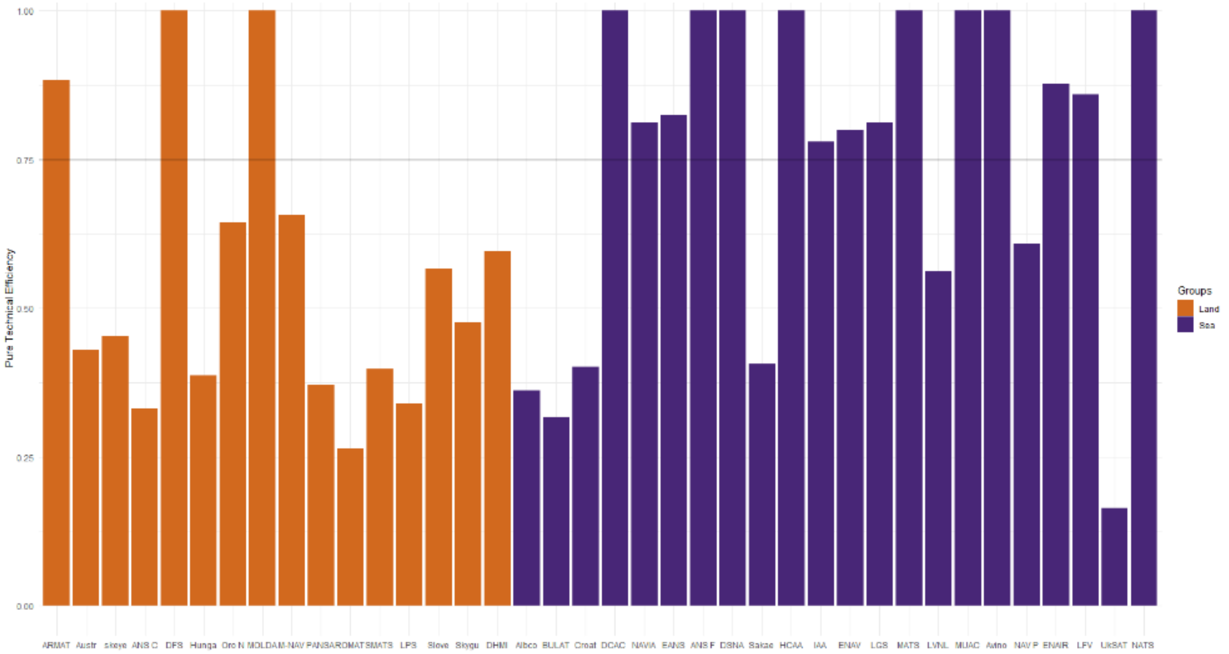


Figure 7-24: Efficiency scores grouped according to the type of territory handled (Source: Personal)

Considering the sample size of each cluster, two separate DEA production frontiers are constructed. As shown on the *Table 7-8*, the ANSPs, located near a sea, have almost the same scores when they are grouped together. However, the efficiency scores of inland ANSPs increase considerably when they are apart from the others. Therefore, the question remains: is the production frontier identical for both type of ANSPs or is the combination of inputs-outputs of the “Sea” ANSPs impossible to reach by the “Land” ANSPs?

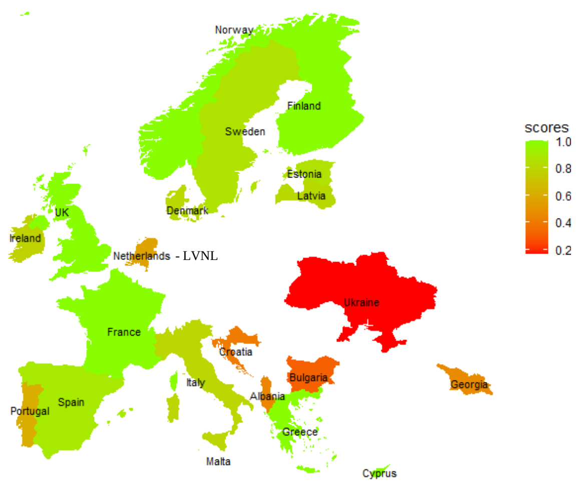
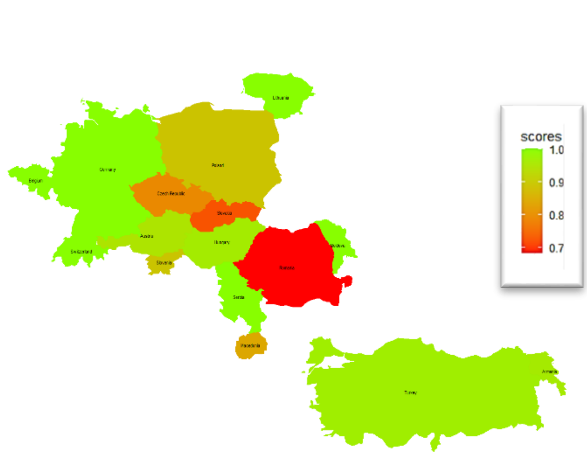
Sea	Land
	

Table 7-8: DEA production frontier computed separately depending on the type of territories (Source: Personal)

For the numerical variables, a two-stage approach is implemented because there is uncertainty on the effect of environmental variables on the performance. Due to the sensitivity of the scores to sampling variations, the bootstrap technic, suggested by Simar and Wilson (2007), is applied to obtain confidence level for each estimate and also bias-corrected scores (see Figure 7-25).

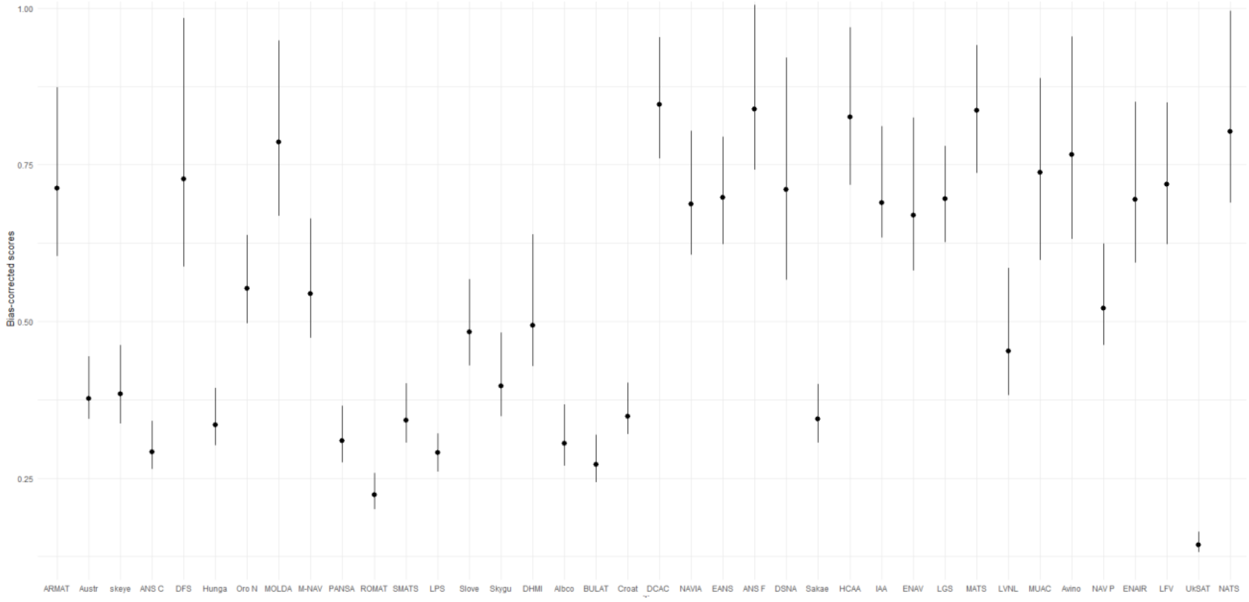


Figure 7-25: Confidence level (line) and bias-corrected efficiency scores (dot) (Source: Personal)

Then, a truncated regression analysis is performed on the inverse of those bias-corrected efficiency scores. If the coefficient is positive, it means that it is unfavourable for the ANSPs. Otherwise, if it is negative, the variable reduces the ratio (1/bias-corrected efficiency scores) and so increases the efficiency.

Variable	Coefficient	Confidence Interval	
		2.5%	97.5%
Traffic Variability	6.287933	-6.09025	21.86105
Network Size	-0.0000004448710	-0.0000064709	0.000008759729
Density	0.1624611	-0.452835	1.171327
Structural	-3.281201	-21.00111	12.06588

Table 7-9: Coefficients obtained from regression analysis (Source: Personal)

This study delivered results in the same vain as the research carried out by COMPAIR. Indeed, according to their interpretation, the traffic variability induced greater costs and the structural complexity was surprisingly associated with lower costs.

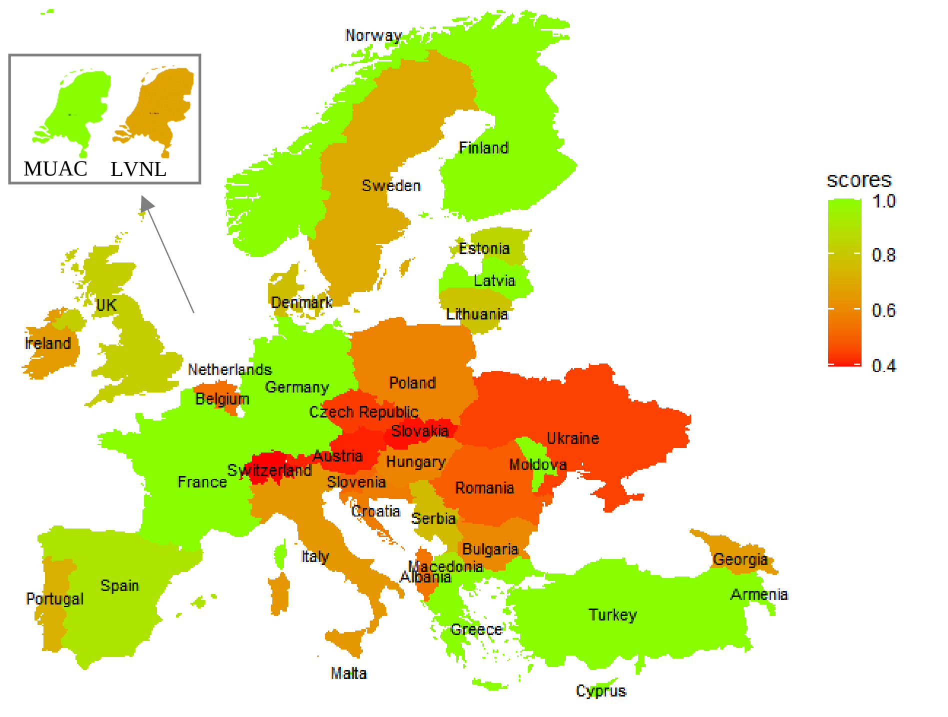
Nevertheless, those coefficients should be interpreted carefully because the two-stage bootstrap model is based on several assumptions. The most important one is the “separability condition” which supposes that the environmental variables have an influence only on the distribution of the efficiency scores and not on the shape or the position of the frontier. Therefore, it is possible that the coefficients are not consistent. Further analyses are needed to verify if this assumption is not violated. If it is the case, other technics should be used to separate the inefficiency due to exogenous and to endogenous factors to better uncover best practices among ANSPs.

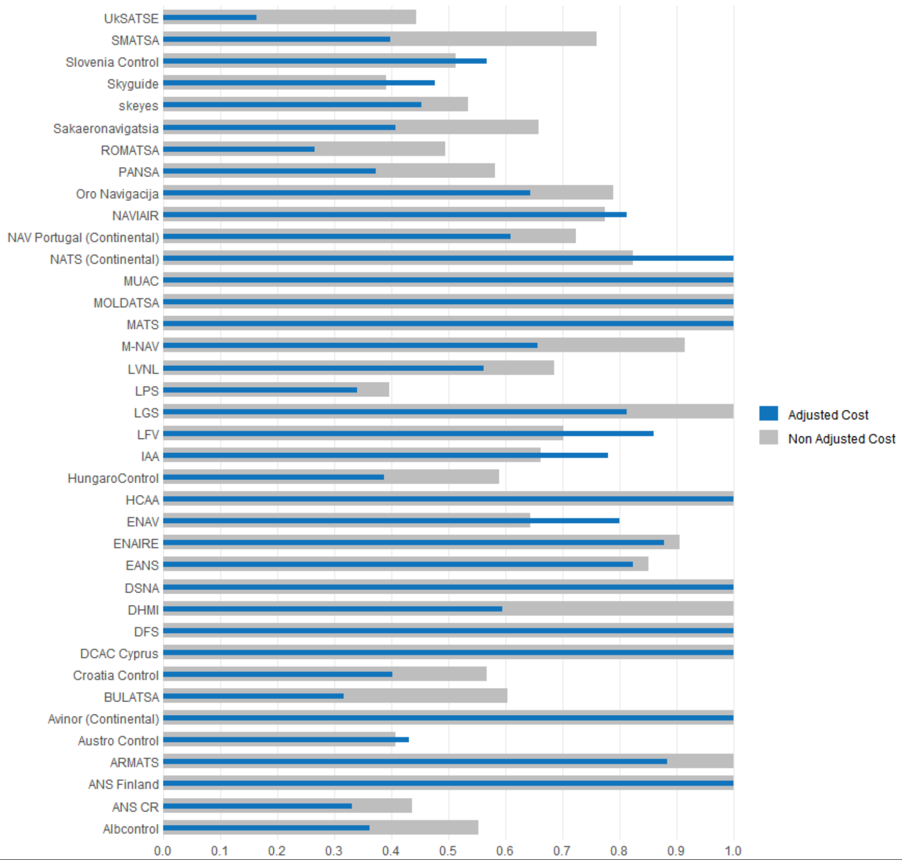
7.3.2 Step 7: Robustness and sensitivity

Ideally, all the sources of uncertainty should be identified. However, in this section, we will only focus on the influence of PPP on the efficiency scores.

The figures below show the evolution of efficiency scores when the costs are not transformed with the "Adjustment factor". NATS (Continental) is the only ANSP which was classified as efficient under cost adjusted assumption and inefficient otherwise. In contrast, LGS, DHMI and ARMATS perform better if the cost-of-living is not taken into account.

Cost Non Adjusted without ATFM Delays (Source: Personal)





Evolution of efficiency scores without ATFM Delays (Source: Personal)

7.4 Stage 4: Interpretation

The aim of this stage is to better understand what drives the CI and to observe the interrelationship with the "Financial Cost-Effectiveness" indicator conceived by the PRU. The following analyses consists in disaggregating the CI and in comparing the efficiency scores to the "Financial Cost-Effectiveness" indicator.

7.4.1 Step 8: Back to the data

The CI is composed of the total costs adjusted (staff costs, non-staff operating costs, depreciation costs, cost of capital and exceptional items) and the outputs (IFR flight hours and IFR airport movements) (see Figure 7-26 and zoom in *Appendix 8*).

For instance, as explained previously, skeyes is considered as inefficient because its radial projection is situated between two ANSPs located on the “best practice” frontier: ANS Finland and Avinor (Continental). It makes sense because Avinor produces more outputs with less inputs and ANS Finland generates less outputs but with a lot less inputs than skeyes.

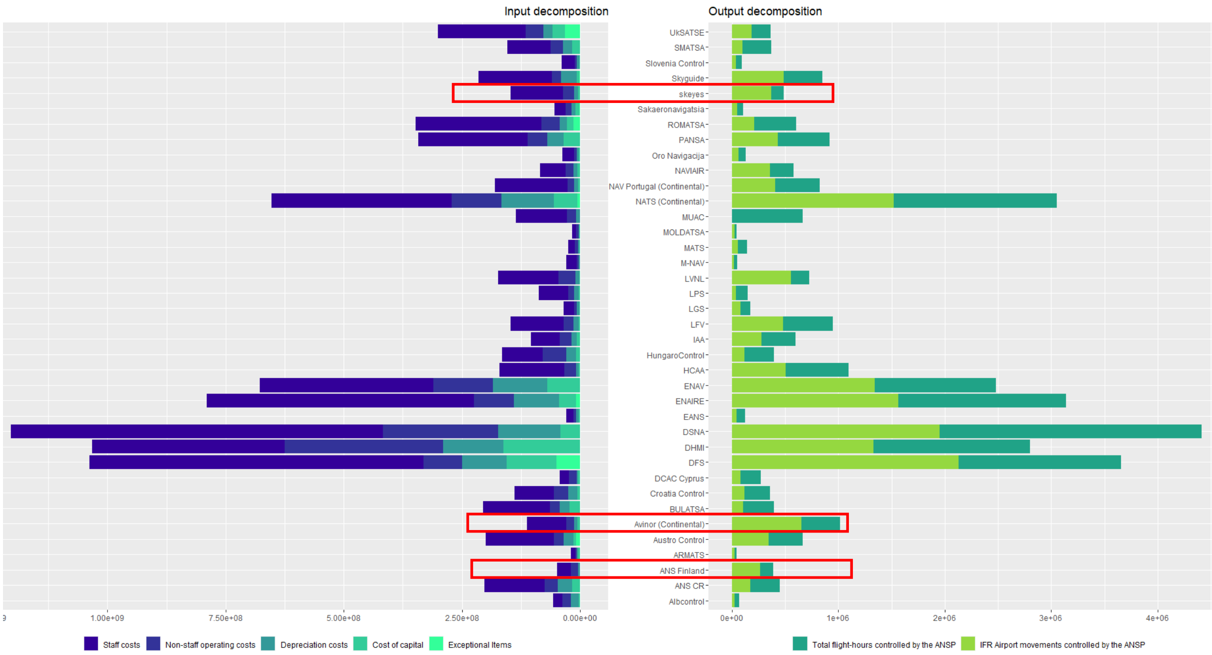


Figure 7-26: Decomposition of the indicators used to build our CI (Cost adjusted and outputs) (Source: Personal)

7.4.2 Step 9: Links to other indicators

After analysing the internal structure of the CI, let’s take a look at the relationship with the “Financial Cost-Effectiveness” indicator. The costs were adjusted to compute the efficiency scores but not to calculate the “Financial Cost-Effectiveness” indicator. The correlation between those measurements are -38%. The negative sign was predictable because the lower the Financial Cost-Effectiveness, the higher the efficiency scores should be. However, the correlation is not strong because the DEA is more flexible regarding the weights and take economies of scale into account.

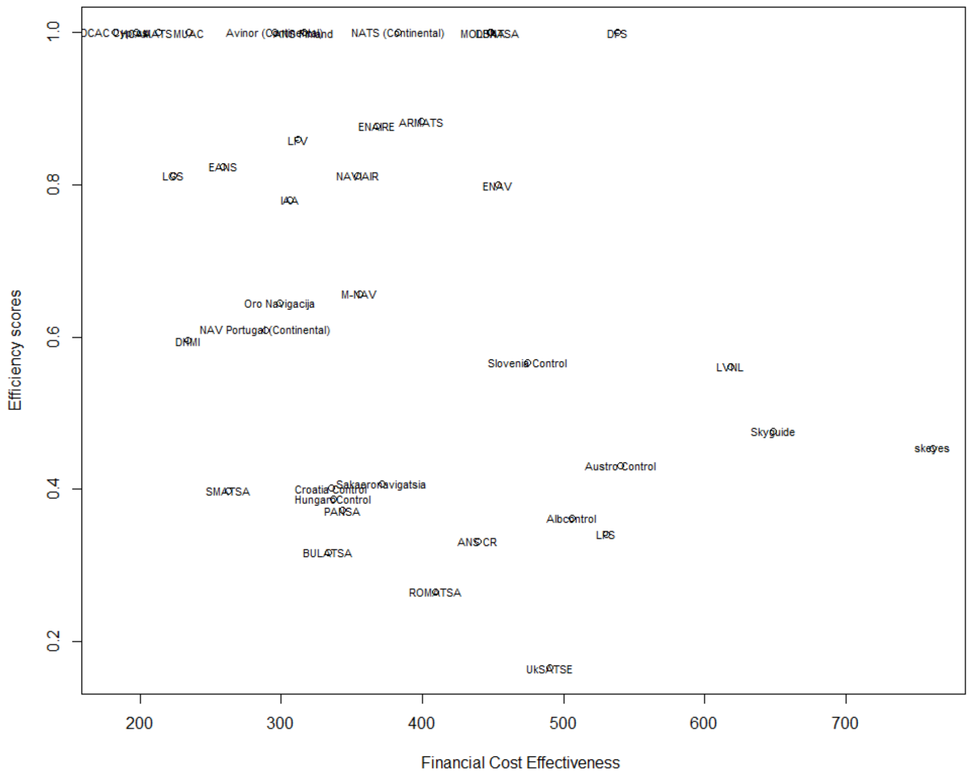


Figure 7-27: Scatter plot of Efficiency scores and Financial Cost Effectiveness (Source: Personal)

7.4.3 Step 10: Presentation and dissemination

Aware of the importance of a good visualisation, many graphs were already provided throughout the analysis. This representation summarises the constructed scores which capture the pure technical efficiency.

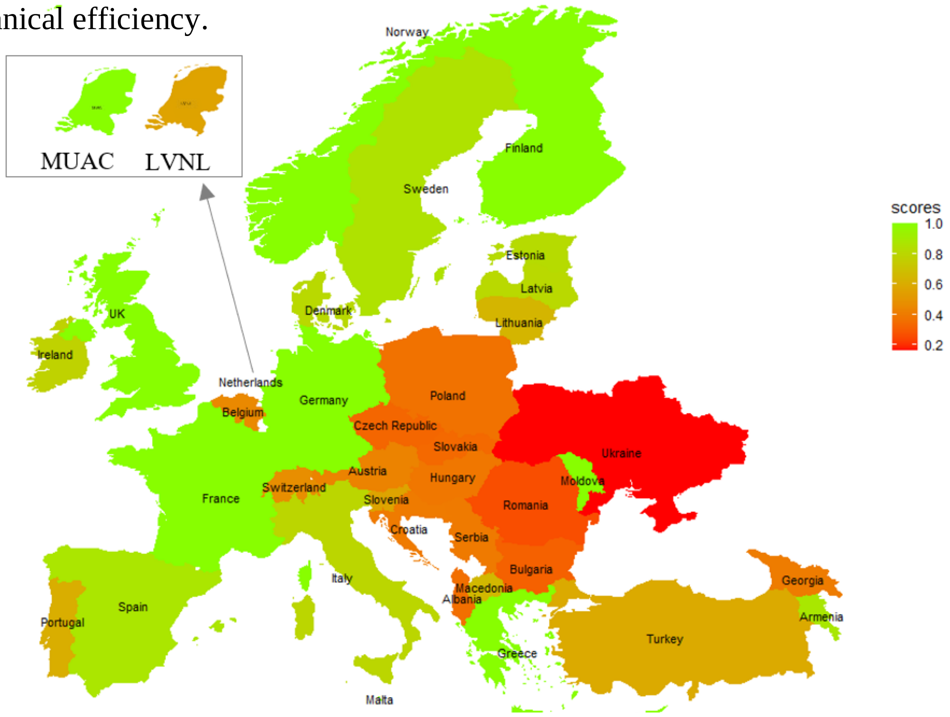


Figure 7-28: Pure technical efficiency scores (Source: Personal)

Chapter 8: Discussion

Before concluding this master thesis, a brief summary of the findings, both at managerial and theoretical level, is provided in this section as well as possible improvements of the methodology designed and of the efficiency scores computed.

8.1 Managerial implication

From a managerial perspective, cost-efficiency scores were obtained by means of the DEA and interpreted through different angles. Then, the impact of categorical and numerical contextual variables was evaluated.

Regarding the categorical factors, including the ownership and the type of territory handled, the efficiency estimates were analysed by groups. Due to the lack of representativity of public and private ANSPs, the conclusion, which leads to believe that those organisations performed better than the commercialised ANSPs, is weak. However, by plotting the results on a map, it seems that there are either spatial autocorrelation issues or disparities on account of the type of airspace managed. This inference derives from the fact that the majority of the ANSPs located in eastern Europe have lower efficiency scores. Therefore, the reason may be that the performance of an ANSP is dependent on those around or that they are situated inland.

For the characteristics measured quantitatively, a two-stage DEA was applied in order to estimate the direction and the intensity of the influence of environmental variables on the efficiency scores. The coefficients of the truncated regression analysis suggest that the traffic variability and the density have a negative impact on the performance whereas the network size and the structural complexity are favourable to the ANSPs. Concerning the structural complexity, the variability and the network size, those findings seem consistent in the light of the results obtained by Competition Economists Group in 2011. In both studies, the structural complexity is surprisingly positively related to the performance. An interesting track to dig would be to measure the variability (intensity and smoothness) and the cyclical pattern of the complexity scores over time. Those items could come forward with information about the predictability of the traffic. Nevertheless, the startling coefficients might also be due to the bootstrap model developed by Simar & Wilson (2007) which is based on several assumptions. The most important one is the separability condition which supposes that the environmental variables only affect the distribution of efficiency scores and not the shape or position of the production frontier.

8.2 Theoretical implication

All those analyses were realised by applying a new framework based on the methodology described in *Handbook on Constructing Composite Indicators: Methodology and User Guide* (Joint Research Centre-European Commission, 2008). This book gives broad-based recommendations to build a Composite Indicator. Therefore, completing it was indispensable to consider the particularities associated with DEA and to achieve the objective of creating a trustful measurement of ANSPs' performance. Indeed, the inception of DEA is represented by an article published by Farrell in 1957 and plenty of extensions have emerged over the years. Consequently, the methods used in the context of DEA supplement the user guide and help to provide a comprehensive view of the Composite Indicator designed.

8.3 Area for improvement

However, this master thesis is not an end in itself but it opens lines of thought for future improvement on a theoretical and a practical level.

For the sake of simplicity, only the traditional *CCR* and *BCC* DEA models were addressed while a variety of adaptations exist. As for the other steps, apart from the implementation, only some suggestions were reported. Those choices are reflected in the study case where only a few models were tested.

In conclusion, on the one hand, the newly built methodology should be enriched to extend its application. On the other hand, the DEA seems to be a better alternative than the SFA to evaluate ANSP's cost-efficiency. It is easier to understand, it relies on more accurate information and it better fits the multiple output production process. However, other extensions of DEA should be applied, the separability condition should be checked, and other environmental variables should be integrated such as, for instance, variability (intensity and smoothness) and the cyclical pattern of the complexity scores over time.

Chapter 9: Conclusion

In a nutshell, it all started with an interview at skeyes about their performance and the indicators already developed in the field of ANSP. As a matter of fact, the Performance Review Unit, which supports the implementation of the Single European Sky initiative, conducts yearly benchmarking analyses by computing, inter alia, the "Financial Cost-Effectiveness" and the "Economic Cost-Effectiveness" indicators. Those measurements are purely factual (EUROCONTROL, 2019) because they only consist in dividing the costs by an artificial output, Composite Flight Hours, which assumes that the intensities of en-route and terminal activities are uniform within Europe. Besides the fact that the particularities of the production process are not well represented in the indicators, the operational characteristics of the ANSPs' airspace, which affect their performance, are also put aside. Consequently, the objective of this master thesis was to construct a new measurement that takes those aspects into account.

Given the multiple outputs produced, a simple ratio, as calculated by the Performance Review Unit, is not suitable and the usage of a Composite Indicator is inevitable. The construction of an efficient frontier, through Data Envelopment Analysis (DEA) or Stochastic Frontier Analysis (SFA) models, is a common practice to implement a cost-efficiency indicator. The choice fell on the DEA for multiple reasons. On the one hand, the approach of SFA in the context of ANSPs' performance has been already largely and rigorously investigated by consulting groups, at the behest of the Performance Review Unit. On the other hand, the DEA does not require the specification of a functional form, which is unknown, and is compatible with production processes generating multiple outputs.

Aware of the reluctance of some analysts to resort to Composite Indicator, it was imperative to build the indicator according to a methodology allowing to improve its transparency. Nevertheless, no such framework exists in the literature to perform a DEA. Therefore, from a theoretical perspective, our first contribution is made by completing a widely-used methodology, jointly conceived by Organisation for Economic Co-operation and Development (OECD) and Joint Research Centre (JRC) (2008), with tools applicable to DEA. Secondly, this new step-by-step procedure is put into practice with the aim of undertaking a benchmarking analysis to assess ANSPs' performance. By means of the DEA, an efficiency score for each ANSP in 2018 was obtained and the impacts of some operational features were evaluated, such as the density, the traffic variability, the network size, the structural complexity, the ownership and the type of territory.

Bibliography

Adler, N. & Golany, B., 2001. Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western Europe.. *European Journal of Operational Research,* 132(2), pp. 260-273.

Andersen, B. & Pettersen, P.-G., 1996. *The benchmarking handbook.* London: Chapman & Hall.

Arblaster, M., 2018. *Air Traffic Management: Economics, Regulation and Governance.* 1 ed. Amsterdam: Elsevier.

Arnaldo, R. M., Comendador, V. F. G., Barragan, R. & Pérez, L., 2014. *European Air Navigation Service Providers' efficiency evaluation through Data Envelopment Analysis (DEA).* St Petersburg, 29th Congress of the International Council of the Aeronautical Sciences.

Asmild, M., Paradi, J. C., Reese, D. N. & Tam, F., 2007. Measuring overall efficiency and effectiveness using DEA. *European Journal of Operational Research,* 178(1), pp. 305-321.

Azizi, H., 2013. A note on data envelopment analysis with missing values. *The International Journal of Advanced Manufacturing Technology,* 66(9-12), pp. 1817-1823.

Balanced Scorecard Institute, 2020. *What is a Balanced Scorecard?.* [Online]
Available at: https://balancedscorecard.org/bsc-basics-overview/
[Accessed 10 May 2020].

Banker, R. D. et al., 1989. An introduction to data envelopment analysis with some of its models and their uses. *Research in governmental and nonprofit accounting,* 5(1), pp. 125-163.

Barone, D., Jiang , L. & Amyot, D., 2011. Composite indicators for business intelligence. In: *International Conference on Conceptual Modeling.* Berlin: Springer, pp. 448-458.

Bilotkach, V. et al., 2015. Cost-efficiency benchmarking of European air nvaigation service providers. *Transportation Research,* Volume 77, pp. 55-60.

Bogetoft, P. & Otto, L., 2011. *Benchmarking with DEA,SFA, and R.* 1 ed. New York: Springer Science & Business Media.

Bogetoft, P. & Otto, L., 2020. *Package "Benchmarking".* [Online]
Available at: https://cran.r-project.org/web/packages/Benchmarking/Benchmarking.pdf
[Accessed 24 July 2020].

Button, K. & Neiva, R., 2014. Economic Efficiency of European Air Traffic Control Systems. *Journal of Transport Economics and Policy,* 48(1), pp. 65-80.

CANSO, 2018. *Global Air Navigation Services Performance Report 2018,* JH Hoofddorp: CANSO (Civil Air Navigation Services Organisation).

Charnes, A., Cooper, W. W. & Rhodes, E., 1978. Measuring the efficiency of decision making units. *European journal of operational research,* 2(6), pp. 429-444.

COMPAIR, 2017. *Report on economic analysis,* Brussels: European Commission.

Competition Economist Group, 2011. *Econometric cost-efficiency benchmarking of Air Navigation Service Provider,* Brussels: EUROCONTROL.

Cook, W. D., Tone, K. & Zhu, J., 2014. Data envelopment analysis: Prior to choosing a model. *Omega,* 44(1), pp. 1-4.

Cooper, W. W., Seiford, L. M. & Tone, K., 2007. *Data envelopment analysis - A comprehensive Text with Models, Applications, References and DEA-Solver Software.* 2 ed. New York: Springer.

Cooper, W. W., Seiford, L. M. & Zhu, J., 2004. *Handbook on Data Envelopment Analysis.* 1 ed. Boston: Kluwer Academic Publishers.

Cooper, W. W., Seiford, L. M. & Zhu, J., 2011. Data envelopment analysis: History, models and interpretations.. In: MA, ed. *Handbook on data envelopment analysis.* Boston: Springer, pp. 1-39.

Daraio, C. & Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach.. *Journal of productivity analysis,* 24(1), pp. 93-121.

Daraio, C., Simar, L. & Wilson, P. W., 2015. *Testing the "separability" condition in two-stage nonparametric models of production.,* Pisa: LEM Working Paper Series, No. 2015/21, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM).

Dempsey-Brench, Z. & Volta, N., 2018. A cost-efficiency analysis of European air navigation service providers. *Transportation research,* Volume 111, pp. 11-23.

Dong, Y., Hamilton, R. & Tippett, M., 2014. Cost efficiency of the Chinese banking sector: A comparison of stochastic frontier analysis and data envelopment analysis. *Economic Modelling,* 36(1), pp. 398-308.

EFQM, 2018. *What is Excellence?.* [Online]
Available at: http://www.efqm.ie/what-is-excellence.html
[Accessed 12 May 2020].

EUROCONTROL, 2016. *ATFM Delay.* [Online]
Available at: https://ext.eurocontrol.int/lexicon/index.php/ATFM_delay
[Accessed 26 May 2020].

EUROCONTROL, 2016. *Performance Scheme.* [Online]
Available at: https://ext.eurocontrol.int/lexicon/index.php/Performance_scheme
[Accessed 7 June 2020].

EUROCONTROL, 2020. *Horizontal Flight Efficiency Dataset.* [Online]
Available at: https://ansperformance.eu/reference/dataset/horizontal-flight-efficiency/#:~:text=Horizontal%20en%2Droute%20flight%20efficiency,European%20airspace%20(reference%20area).
[Accessed 26 May 2020].

EUROCONTROL, 2020. *Pan-European ANS Performance Data Portal.* [Online]
Available at: https://ansperformance.eu/
[Accessed 17 December 2019].

European Commission, 2018. *Commission staff working document - on the ex-post evaluation of the Single European Sky Performance and Charging Schemes in Reference Period 1 and first year of Reference Period 2 {SWD(2018) 63 final},* Brussels: European Commission

European Commission, 2019. *10 Step Guide.* [Online]
Available at: https://composite-indicators.jrc.ec.europa.eu/?q=10-step-guide/step-6-weighting
[Accessed 28 May 2020].

European Commission, 2020. *Co-operation with EUROCONTROL.* [Online]
Available at: https://ec.europa.eu/transport/modes/air/single_european_sky/co-operation_eurocontrol_en
[Accessed 11 August 2020].

European Commission, 2020. *Performance Review Body.* [Online]
Available at: https://ec.europa.eu/transport/modes/air/single-european-sky/ses-performance-and-charging/performance-review-body_en
[Accessed 7 June 2020].

European Commission, 2020. *Single European Sky*. [Online]
Available at: https://ec.europa.eu/transport/modes/air/ses_en
[Accessed 26 May 2020].

Eurostat, 2018. *Air passenger transport in the EU: Record number of air passengers carried at more than 1 billion in 2017.* [Online]
Available at: https://ec.europa.eu/eurostat/documents/2995521/9428738/7-06122018-AP-EN.pdf/50a52d8d-3f61-4517-ace3-d3f56ed5cd91
[Accessed 17 December 2019].

Eurostat, 2020. *Revision of PPP time series.* [Online]
Available at: https://ec.europa.eu/eurostat/web/purchasing-power-parities/data
[Accessed 27 July 2020].

Gan, X. et al., 2017. When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators,* 81(1), pp. 491-502.

Gartner, 2019. *Corporate Performance Management (CPM).* [Online]
Available at: https://www.gartner.com/en/information-technology/glossary/cpm-corporate-performance-management
[Accessed 11 April 2020].

Gartner, 2019. *Gartner Glossary - Analytics and Business Intelligence.* [Online]
Available at: https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi
[Accessed 1 June 2020].

Golany, B. & Roll, Y., 1989. An application procedure for DEA. *Omega*, 17(3), pp. 237-250.

Incekara, A., Guz, T. & Sengun, G., 2017. Measing the Technology Achivement Index: Comparison and Ranking of countries. *Journal of Economics, Finance and Accounting,* 4(2), pp. 164-174.

Inmon, W. H., 2005. *Building the data warehouse.* 4 ed. Indianapolis: Wiley Publising Inc..

International Monetary Fund, 2020. *World Economic Outlook Database.* [Online]
Available at: https://www.imf.org/external/pubs/ft/weo/data/changes.htm
[Accessed 27 July 2020].

Isik, I. & Hassan, M., 2002. Cost and profit efficiency of the Turkish banking industry: An empirical investigation. *Financial Review*, 37(2), pp. 257-279.

Joint Research Centre-European Commission, 2008. *Handbook on constructing composite indicators: methodology and user guide.* 1 ed. Paris: OECD publishing.

Kao, C. & Liu, S.-T., 2000. Data envelopment analysis with missing data: an application to university libraries in Taiwan. *Journal of the Operational Research Society,* 51(8), pp. 897-905.

Kaplan, R. S. & Norton, D. P., 2005. The balanced scorecard: measures that drive performance. *Harvard business review,* 83(7), p. 172.

Kuosmanen, T., 2009. Data envelopment analysis with missing data. *the Operational Research Society,* 60(12), pp. 1767-1774.

Liu, W. B., Meng, W., Li, X. X. & Zhang, D. Q., 2010. DEA models with undesirable inputs and outputs.. *Annals of Operations Research,* 173(1), pp. 177-194.

Mendes De Leon, P. & Calleja Crespo, D., 2011. *Achieving the Single European Sky - Goals and Challenges.* 1 ed. AH Alphen aan den Rijn: Kluwer Law International BV..

Nataraja, N. R. & Johnson, A. L., 2011. Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research,* 215(1), pp. 662-669.

Negash, S. & Gray, P., 2008. Business Intelligence. In: *Handbook on decision support system.* Berlin: Springer, pp. 175 - 193.

Neiva, R., 2014. *Economic Efficiency of Air Navigation Service Providers: An assessment in Europe,* George Mason University: Doctoral dissertation.

Nera Economic Consulting, 2006. *Cost Benchmarking of Air Navigation Service Providers: A Stochastic Frontier Analysis,* London: Nera Economic Consulting.

Nero, G. & Portet, S., 2007. Five years experience in ATM cost benchmarking. *7th USA/Europe Air Traffic Management R&D Seminar.*

OECD, 2015. *Measuring Well-being in Mexican States.* Paris: OECD Publishing.

O'Rourke, L., 2012. *Handbook on Applying Environmental Benchmarking in Freight Transportation,* Washington: Transportation Research Board.

Parmenter, D., 2015. *Developing, Implementing and Using Winning KPIs.* 3e ed. Hoboken: John Wiley & Sons.

Pastor, J. T., Ruiz, J. L. & Sirvent, I., 2002. A statistical test for nested radial DEA models. *Operations Research,* 50(4), pp. 728-735.

Pavláková Docekalová, M. & Kocmanová, A., 2016. Composite indicator for measuring corporate sustainability. *Ecological Indicators,* 61(2), pp. 612-623.

Performance Review Unit, 2006. *ATM Cost-Effectiveness (ACE) 2004 Benchmarking Report,* Brussels: EUROCONTROL.

Performance Review Unit, 2019. *ATM Cost-Effectiveness (ACE) 2017 Benchmarking Report with 2018-2022 outlook,* Brussels: Report Commissioned by the PRC.

PRB, 2020. *PRB Work Program for 2020,* Brussels: Performance Review Body of the Single European Sky.

PRC, 2018. *Performance Review Report (PRR) 2017,* Brussels: EUROCONTROL.

Purvis, B., Mao, Y. & Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. *Sustainability science,* 14(3), pp. 681-695.

Rowena, J., 2001. Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. *Health Care Management Science,* 4(1), pp. 103-115.

Ruggiero, J., 2005. Impact assessment of input omission on DEA.. *International Journal of Information Technology & Decision Making,* 4(3), pp. 359-368.

Seiford, L. M. & Zhu, J., 2002. Modeling undesirable factors in efficiency evaluation.. *European journal of operational research,* 142(1), pp. 16-20.

Simar, L. & Wilson, P. W., 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science,* 44(1), pp. 49-61.

Simar, L. & Wilson, P. W., 2007. Estimation and inference in two-stage, semi parametric models of production processes.. *Journal of econometrics,* 136(1), pp. 31-64.

Simm, J. & Besstremyannaya, G., 2020. *Package "rDEA".* [Online]
Available at: https://cran.r-project.org/web/packages/rDEA/rDEA.pdf
[Accessed 26 July 2020].

skeyes, 2018. *Air Traffic Control Providers to manage Belgian airspace with one single system.* [Online]
Available at: https://www.eurocontrol.int/news/air-traffic-control-providers-manage-belgian-airspace-one-single-system
[Accessed 8 January 2020].

skeyes, 2018. *Nos yeux rivés sur votre sécurité - Rapport Annuel 2018,* Steenokkerzeel: skeyes.

SKYBRARY, 2019. *Single European Sky (SES).* [Online]
Available at: https://www.skybrary.aero/index.php/Single_European_Sky_(SES)
[Accessed 4 January 2020].

Smirlis, Y. G., Maragos, E. K. & Despotis, D. K., 2006. Data envelopment analysis with missing values: An interval DEA approach. *Applied mathematics and computation,* 177(1), pp. 1-10.

Taouab, O. & Issor, Z., 2019. Firm Performance: Definition and Measurement Models. *European Scientific Journal,* 15(1), pp. 93-106.

Tauler, R., Walczak, B. & Brown, S. D., 2009. *Comprehensive chemometrics: chemical and biochemical data analysis.* Amsterdam:Elsevier.

Thanassoulis, E., 2001. *Introduction to the theory and application of data envelopment analysis.* 2 ed. Dordrecht: Kluwer Academic Publishers.

Ueda, T. & Hoshial, Y., 1997. Application of principal component analysis for parsimonious summarization of DEA inputs and.or outputs.. *Journal of the Operations Research Society of Japan,* 40(4), pp. 466-478.

Wagner, J. M. & Shimshak, D. G., 2007. Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research,* 180(1), pp. 57-67.

Wahab, M., 2017. *Interpolation and Extrapolation.* [Online]
Available at:
https://www.researchgate.net/publication/313359516_Interpolation_and_Extrapolation
[Accessed 25 July 2020].

Wilson, P. W., 1993. Detecting outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs. *Journal of Business & Economic Statistics,* 11(3), pp. 319-323.

World Health Organization, 2014. *Child health and development.* [Online]
Available at: http://www.emro.who.int/child-health/research-and-evaluation/indicators/All-Pages.html
[Accessed 23 May 2020].

Wright, D., 2013. Ask Air Traffic Control: Instructions for takeoff. *USA TODAY,* 3 June.

Zanella, A., Camanho, A. S. & Dias, T. G., 2015. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis.. *European Journal of Operational Research,* 245(2), pp. 517-530.

Zhu, J., 2014. *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets.* 3 ed. Cham: Springer.

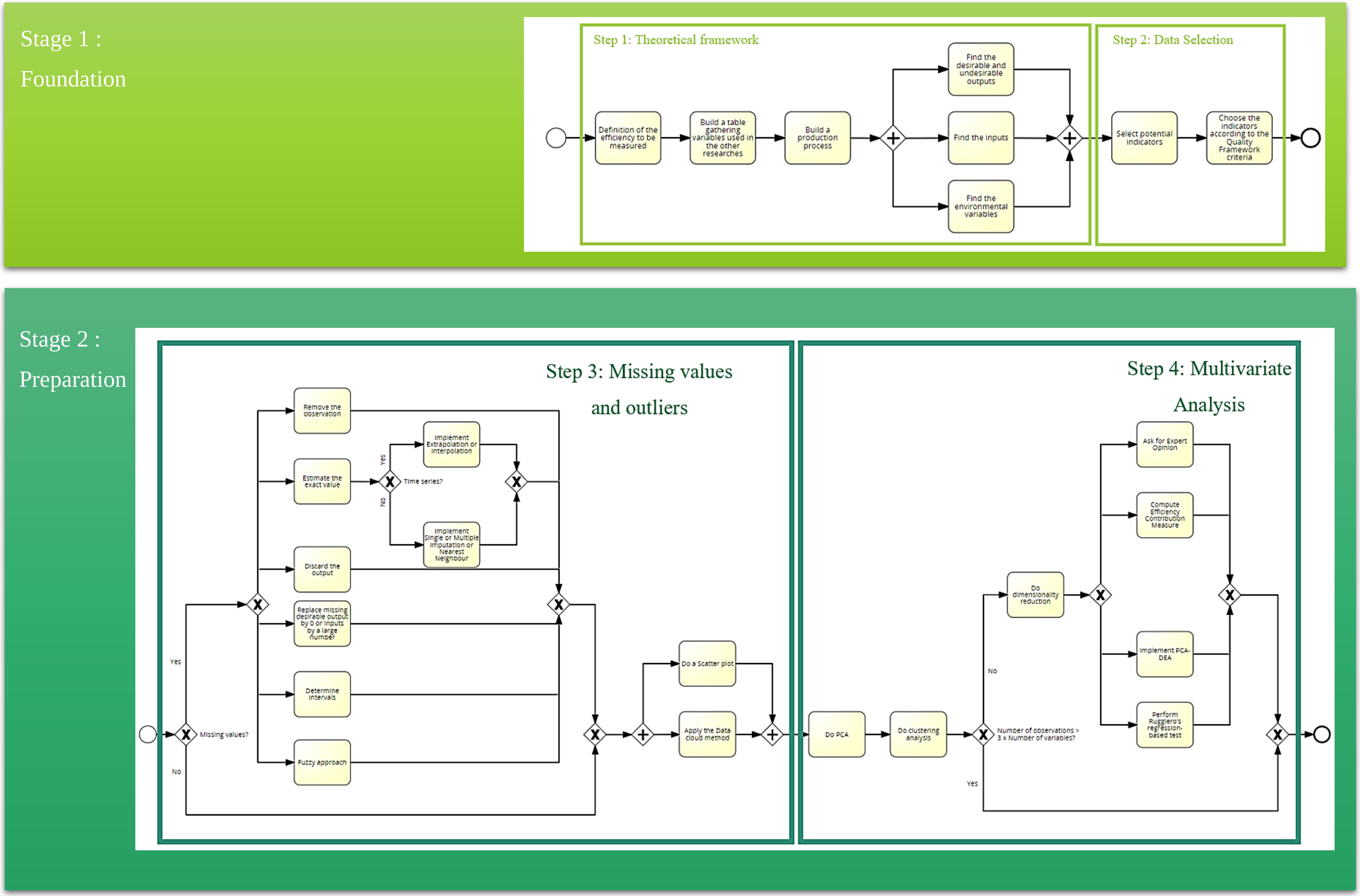
Appendices

Appendix 1: Acronyms

Acronyms	Full Name	First time introduced (Page)
ACC	Area Control Centre	8
ACE	ATM Cost-Effectiveness	25
AIS	Aeronautical Information Services	5
ANS	Air Navigation Service	9
ANSP	Air Navigation Service Providers	1
APP	Approach Control Unit	8
ASM	Air Space Management	6
ATC	Air Traffic Control	6
ATCO	Air Traffic Control Officer	7
ATFM	Air Traffic Flow Management	6
ATM	Air Traffic Management	5
BSC	Balanced Scorecard	12
CEG	Competition Economics Group	31
CFH	Composite Flight Hours	28
CI	Composite Indicator	17
CNS	Communication, navigation and surveillance systems	5
CPM	Corporate Performance Management	11
CRS	Constant Return to Scale	53
DEA	Data Envelopment Analysis	30
DMU	Decision Making Units	36
EC	European Commission	23
EFQM	European Foundation for Quality Management	12
FAB	Functional Airspace Bloc	2
FP	Fractional Program	37
IFR	Instrument Flight Rules	28
IMF	International Monetary Fund	61
KPA	Key Performance Areas	11
KPI	Key Performance Indicator	15

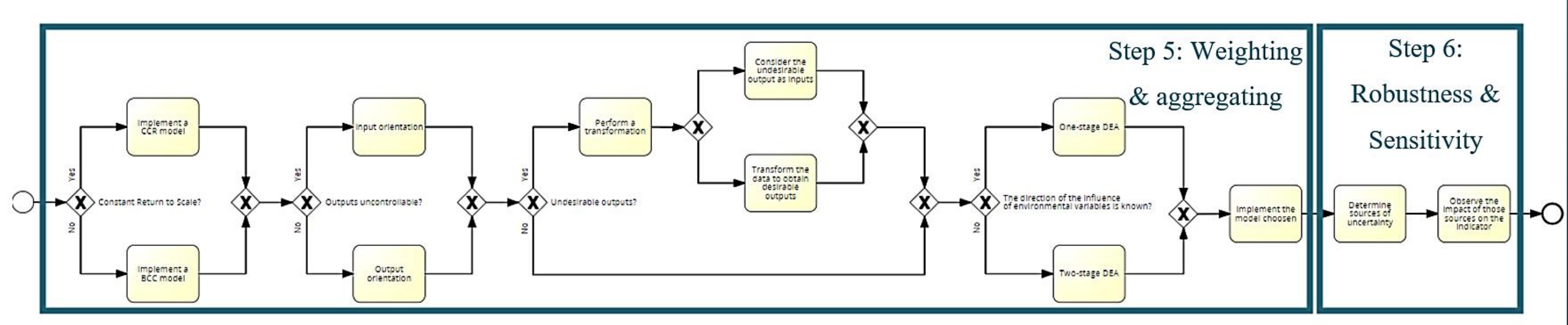
KRI	Key Result Indicator	15
LP	Linear Program	37
MET	Meteorological services for Air Navigation	5
PC	Principal Component	65
PCA	Principal Component Analysis	19
PI	Performance Indicator	15
PPP	Purchasing Power Parity	60
PPS	Purchasing Power Standard	63
PRB	Performance Review Body	25
PRC	Performance Review Commission	25
PRR	Performance Review Report	25
PRU	Performance Review Unit	25
RI	Result Indicator	15
RP	Reference Period	24
RTS	Return To Scale	53
SAR	Search and Rescue	5
SES	Single European Sky	23
SFA	Stochastic Frontier Analysis	30
VRS	Variable Return to Scale	53

Appendix 2: Methodology process

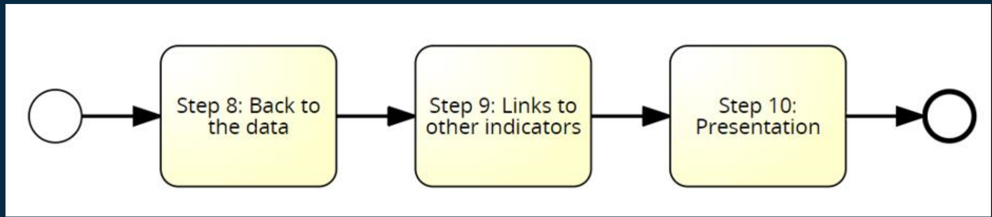


Stage 1 :
Foundation
Step 1: Theoretical framework
Definition of the efficiency to be measured
Build a table gathering variables used in the other researches
Build a production process
Find the desirable and undesirable outputs
Find the inputs
Find the environmental variables
Step 2: Data Selection
Select potential indicators
Choose the indicators according to the Quality Framework criteria
Stage 2 :
Preparation
Step 3: Missing values and outliers
Missing values?
Yes
No
Remove the observation
Estimate the exact value
Time series?
Yes
No
Implement Extrapolation or Interpolation
Implement Single or Multiple Imputation or Nearest Neighbour
Discard the output
Replace missing desirable output by 0 or inputs by a large number
Determine intervals
Fuzzy approach
Do a Scatter plot
Apply the Data cloud method
Step 4: Multivariate Analysis
Do PCA
Do clustering analysis
Number of observations > 3 x Number of variables?
No
Yes
Do dimensionality reduction
Ask for Expert Opinion
Compute Efficiency Contribution Measure
Implement PCA-DEA
Perform Ruggiero's regression-based test

Stage 3 : Construction



Stage 4 : Interpretation



Appendix 3: Comparative table

Name		Type	Output	Input	Input price	Environmental	Remark
Nera Economic Consulting (2006)	Project	SFA	1. Composite gate-to-gate flight hours	1. Capital units (to derive the capital input price)	1. ATCO in OPS labour price 2. Non-ATCO staff labour price 3. Capital input price 4. Price index for direct non staff operating input price	1. Network size 2. Traffic variability 3. Adjusted density 4. Structural complexity 5. Complexity score	
Competition Economics Group (2011)	Project	SFA	1. Composite flight hours	1. Capital physical inputs (to derive the capital input price)	1. ATCO in OPS labour price 2. Support staff labour price 3. Producer price index (PPI) for non staff operating input price 4. Capital input price	1. Traffic variability 2. Business environment quality 3. Network concentration 4. Network size 5. Structural traffic complexity 6. Time	All monetary variables are divided by the PPI
Rosa M. Arnaldo & all (2014)	Congress	DEA	1. ACC operational data 2. En-Route output data 3. Revenues	1. All costs (MET, EUROCONTROL, …) 2. ATM/CNS provision costs 3. Total staff	/	1. Variability 2. Complexity 3. Size	

				4. ATCOs in OPS 5. ATCO’s hours on duty 6. Staff cost for ATCOs in OPS			
Rui Neiva (2014)	Dissertation For Ph.D	DEA And SFA	For DEA 1. IFR flights hours 2. IFR airport movements 3. 1/[Minutes of air traffic flow management delays exceeding 15 min] And for SFA 1. Composite flight hours	For DEA 1. Gate-to-Gate ATM/CNS provision costs 2. Gate-to-gate of non-control services And for SFA 1. Staff costs 2. Other provision costs 3. Non provision costs	/	1. Year 2. ATCO 3. Spatial coverage of each controller 4. Ownership 5. Size 6. Sectors 7. Revenues/costs ratio 8. Staff/ATCO in OPS 9. Density 10. Structural complexity 11. Complexity	PPP used to adjust for the different purchasing powers across countries. (Eurostat) BCC variable return to scale Look at multicollinearity problem Slack analysis Spatial autocorrelation issues Bias-corrected bootstrapped efficiency results Malmquist indexes, Visualisation SFA: use of production function
K. Button & R. Neiva (2014)	Research Paper	DEA	1. IFR flight hours 2. IFR airport movements 3. 1/[Minutes of ATFM delays exceeding 15 minutes]	1. Gate-to-gate ATM/CNS provision costs 2. Other gate-to-gate costs of non-control services		1. Year 2. ATCO 3. ATCO km 4. Sectors 5. Size 6. Ownership 7. Revenues/costs 8. Staff/ATCO in OPS	Bias-corrected bootstrapped DEA estimated efficiencies

Bilotkach & all (2015)	Research Paper	DEA	1. Total flight hours controlled (en-route ouput) 2. IFR airport movement (terminal output)	1. Gate-to-gate ATM/CNS costs	1. ATCO input prices 2. Other staff input prices 3. Capital input prices 4. Non staff related input prices		For input prices, follow instruction in ACE report 2011. Not CFH to increase flexibility of the model → CFH source of biais Upper and lower bounds of the efficiency scores (biased-corrected)
Cujic & all. (2015)		DEA	1. Delay 2. Composite flight hours 3. Total revenues	1. ATCO costs 2. Total costs excluding ATCO costs			Slack based Measures
COMPAIR (2017)	Project	SFA	Production function 1. Total flight hours controlled 2. IFR airport movements Cost function 1. Total cost	Production function 1. ATCO hours ACC 2. ATCO hours APP+TWR 3. Capital input Cost function 1. Total flight hours 2. Labour input 3. Capital input	1. Labour input price (en-route) 2. Labour input price (terminal) 3. Capital input price (en-route) 4. Capital input price (terminal)	1. Size/ En-route sectors 2. Seasonality 3. Complexity 4. Ownership	• Cost & production functions • En-route & Terminal • All monetary variables/cost index • Non staff operating costs not integrated
Adler & all (2017)	Discussion Paper	SFA	1. IFR flight hours 2. IFR airport movements	1. Labour (Cost or ATCO hours) 2. Capital (Cost of En-route sectors)	1. Labour input price 2. Capital input price	1. Seasonality 2. Complexity 3. Ownership	• Intermediate goods and energy price index • PPP and PPI

Dempsey-Brench and Volta (2018)	Research paper	SFA	1. Composite flight hours	1. Capital input 2. Cost	1. ATCO labour price 2. Non-ATCO labour price 3. Capital input price 4. Material Input price (index)	1. Size 2. Variability 3. Complexity 4. Ownership	Monetary values are adjusted by Purchasing Power Parity (PPP)

Appendix 4: Quality Framework Analysis

Component	Indicators	Formula	Accuracy	Relevance	Interpretability	Coherence
			OUTPUT			
Services	Composite flight hours (CFH)	$en-route\ flight\ hours + 0.26 * IFR\ airport\ movements$	Taking the two indicators (IFR flight hours controlled & IFR airport movements) separately is closer to the reality than using Composite Flight Hours. Indeed, the intensity of the en-route and terminal activities are different for each ANSP. For instance, skeyes delegates a part of its en-route activities to MUAC.	YES because it is an indicator of the services (outputs) provided by the ANSP.	It is less easy for the users to understand how the weights factor is derived (0.26).	The CFH is a composite indicator of IFR hours controlled and IFR airport movements. By assigning a weight, it reduces the flexibility of the model. There are two other problems: • It measures the demand satisfied rather than the capacity provided. • It excludes the VFR movements controlled
	IFR flight hours controlled[5]				It is easy to interpret. Also to ensure comparability, those data are extracted from the same database.	
	IFR airport movements[6]					

[5] "difference between the exit time and entry time of any given flight in the controlled airspace of an operational unit" (Performance Review Unit, 2019).
[6] gather all the movements of take-off and landing. It is considered as an output measure for terminal ANS.

Quality	Minutes of ATFM delays exceeding 15 minutes		This metric reflects only a part of the concepts to define. Others indicators should be incorporated to capture other aspects of the quality of services provided.	YES Because CANSO report (2018) , “a lower cost per flight hours is not necessarily indicative of improved overall performance” and “there are costs associated with providing a safer and more punctual, predictable and efficient service”.	YES	This indicator is coherent over time and across countries because the methodology did not change over the period considered and it is based on the same concepts for all countries.
INPUT						
Capital	Capital units (1)	$ACC\ sector\ hours + \alpha * ILS\ localisers$	Those two indicators are quite far from the concept to be measured. It is preferable to distinguish the element (number of radars, …) than gathering it into the same indicator and using external index.	YES Because one of the inputs used by the ANSP is related to capital such as building, screens, …	The computation of α is not intuitive and this affects the interpretability of the indicators.	The data source for ILS localisers is not precise. Therefore, it is difficult to assess the coherence of the indicator across the countries.
	Capital units (2)	$\frac{Net\ Book\ Value\ of\ assets}{Annual\ producer\ capital\ index}$			This measure seems easier to interpret because the NBV is the value of assets reported by the ANSPs on their balance sheet.	A better measurement of the capital inputs is necessary for the improvement of the CI.

	Capital costs	$Depreciation\ costs + cost\ of\ capital$	It is closer to the true values.		It is much easier to interpret.	This is directly related to the capital used. However, this indicator is not coherent across countries because the depreciation method might be different.
Staff	ATCO in OPS	• ACC in OPS • APP + TWR in OPS	The hours should be taken into account instead of the physical people. Indeed, the number of people required will depend on the working shift allowed in their respective country.	YES Because one of the inputs used by the ANSP is related to the staff.	YES There are easy to interpret.	Normally, there is only one way to compute these indicators. Therefore, the coherence seems to be met.
	ATCO hours	• ACC ATCO hours • APP + TWRs ATCO hours				
	Full Time Equivalent (FTE)		It is coherent with the notion of staff.			
	Employment costs		This is directly related to the staff used.			The accounting standards are different across countries.
Direct non-staff operating input	Cost	Non staff operating costs + Exceptional items	This is directly related to the staff used.	YES This category gathers all the remaining costs which were not classified as capital or labour costs such as energy, insurance, …	YES	The accounting standards are different across countries.

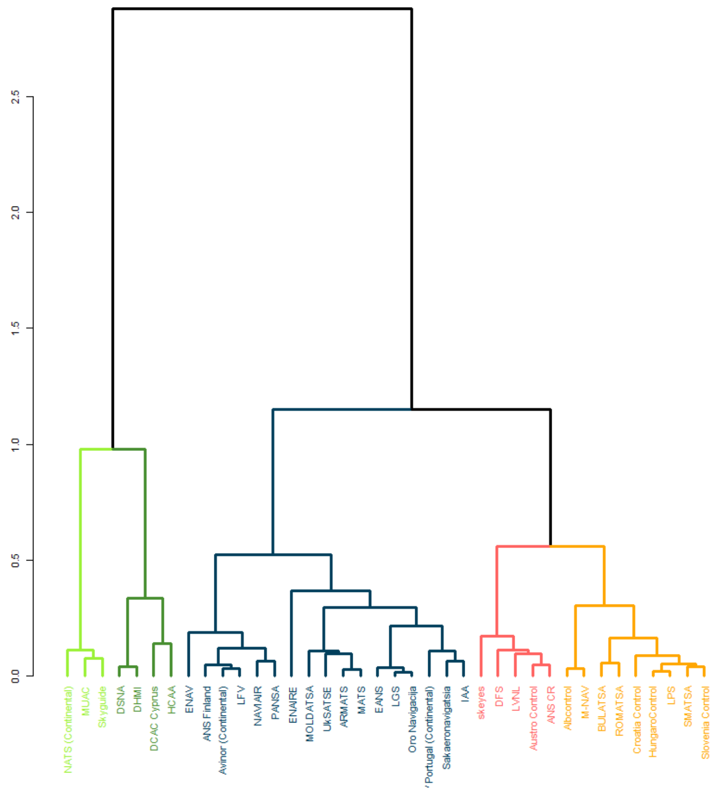
INPUT PRICE					
Capital	$\frac{Depreciation\ costs\ (x) + cost\ of\ capital\ (x)}{(NBV/Capital\ goods\ price\ index)}$	It is quite far from the true values.	YES	It is simply the costs divided by the capital input.	The accounting standards and the depreciation methods are different across countries.
Staff	$\frac{\boldsymbol{employment\ costs}}{\boldsymbol{\#ATCO\ in\ OPS}}$	These indicators seem close to the real values.	YES If we use the number of ATCO in OPS as input	YES	The accounting standards are different across countries.
	$\frac{\boldsymbol{employment\ costs}}{\boldsymbol{number\ of\ ATCO\ hours\ in\ OPS}}$		YES If we use the number of ATCO hours in OPS as input	YES	A distinction can be made if we separate ACC and APP/TWR • $\frac{Total\ employment\ costs}{ATCO\ hours\ in\ ACC}$ • $\frac{Total\ employment\ costs}{APP+TWR\ ATCOs}$
	$\frac{\boldsymbol{support\ staff\ employment\ costs}}{\boldsymbol{support\ staff\ Full\ Time\ Equivalent}}$		YES To separate the costs generated by ATCO and the costs generated by the support staff.	YES	The accounting standards are different across countries.
Direct non-staff	Price index	It is far from the real values.	YES	It is more difficult to interpret because it is an index.	This is a weak input prices because it is not based on the actual costs.
ENVIRONMENTAL					

Network size	Number of square kilometres of the airspace controlled		YES because bigger ANSP needs more capital (such as radar, buildings, …) to cover the area	YES	It is a coherent measure to estimate the network size.
	Volume of airspace controlled by ACC			YES	This measure has never been taken into account, it could be interesting to estimate the influence of this variable.
Variability	$\frac{Peak\ weak}{Average\ traffic\ week}$	This indicator seems to be accurate.	YES Because if unexpected changes in the traffic volumes occur, it is difficult for ANSPs to cut back on their capacity by adjusting their inputs (labour and capital). Those inputs are not so flexible to let ANSP to accommodate to those unanticipated evolutions.	YES	This measure is coherent across countries and the methodology did not change.
Complexity	Adjusted density	The question remains if it is better to take the complexity score as a whole or the	YES the complexity is important because it influences the human	The documentation about these metrics are abundant. The logic is clear but the	The methodology to compute those metrics change in 2014 and in 2017. Those two time series cannot be used jointly.
	Structural complexity				

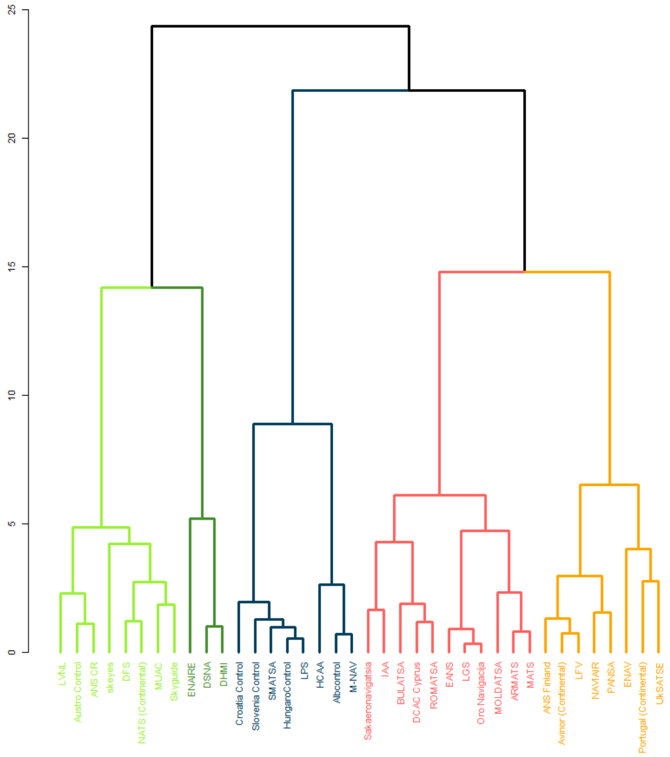
	Complexity score = Adjusted * Structure	separate the effects of density and complexity.	resources required to handle the traffic.	way to compute it seems more complex.	
Ownership	Ownership	The values are binaries, so the accuracy does not apply.	According to the PRU, the ownership could be a factor affecting the ANSPs' performance.	It is easy to interpret if the categories are well defined.	The categories should be well defined.
Business	Business environment quality	It is a proxy variable extracted from the Transparency Internal database.	This variable capture the "risk to invest in a given country taking into account the local business and institutional environments".	It is not easy to interpret and the relevance of this variable is questionable.	

Appendix 5: Dendrograms

Daisy Distance



Manhattan distance



Appendix 6: Efficiency scores

ANSP	Pure Technical	Scale		Mix	
ANSP	Pure Technical	Scale	RTS	Slack En route	Slack Terminal
Albcontrol	0.362	0.505	IRS	0.00	14854.00
ANS CR	0.332	0.995	DRS	0.00	0.00
ANS Finland	1.000	0.929	IRS	0.00	0.00
ARMATS	0.883	0.290	IRS	0.00	6428.27
Austro Control	0.431	0.997	DRS	0.00	0.00
Avinor (Continental)	1.000	1.000	CRS	0.00	0.00
BULATSA	0.317	0.996	DRS	0.00	0.00
Croatia Control	0.401	0.996	DRS	0.00	0.00
DCAC Cyprus	1.000	1.000	CRS	0.00	0.00
DFS	1.000	0.419	DRS	0.00	0.00
DHMI	0.595	0.612	DRS	0.00	74619.83
DSNA	1.000	0.508	DRS	0.00	0.00
EANS	0.824	0.761	IRS	0.00	7036.49
ENAIRE	0.877	0.594	DRS	0.00	0.00
ENAV	0.799	0.571	DRS	159823.42	0.00
HCAA	1.000	0.867	DRS	0.00	0.00
HungaroControl	0.387	0.995	DRS	0.00	0.00
IAA	0.780	0.996	DRS	0.00	0.00
LFV	0.860	0.972	DRS	0.00	0.00
LGS	0.812	0.818	IRS	0.00	0.00
LPS	0.341	0.837	IRS	0.00	28078.32
LVNL	0.561	0.991	IRS	135722.42	0.00
MATS	1.000	0.824	IRS	0.00	0.00
M-NAV	0.657	0.399	IRS	0.00	16631.26
MOLDATSA	1.000	0.285	IRS	0.00	0.00
MUAC	1.000	1.000	CRS	0.00	0.00
NATS (Continental)	1.000	0.614	DRS	0.00	0.00
NAV Portugal (Continental)	0.609	0.994	DRS	0.00	0.00
NAVIAIR	0.812	0.965	IRS	0.00	0.00
Oro Navigacija	0.644	0.707	IRS	0.00	0.00
PANSA	0.372	0.972	DRS	0.00	0.00
ROMATSA	0.265	0.991	DRS	0.00	0.00
Sakaeronavigatsia	0.407	0.633	IRS	0.00	0.00
skeyes	0.454	0.962	IRS	72669.67	0.00
Skyguide	0.476	0.998	DRS	0.00	0.00
Slovenia Control	0.566	0.636	IRS	0.00	10641.48
SMATSA	0.398	0.997	DRS	0.00	0.00
UkSATSE	0.165	0.968	IRS	0.00	0.00

Appendix 7: Reference set

ANSP	peer1	peer2	peer3
Albcontrol	MATS	MOLDATSA	
ANS CR	DCAC Cyprus	MUAC	Avinor (Continental)
ANS Finland	ANS Finland		
ARMATS	MATS	MOLDATSA	
Austro Control	DCAC Cyprus	MUAC	Avinor (Continental)
Avinor (Continental)	Avinor (Continental)		
BULATSA	DCAC Cyprus	MUAC	Avinor (Continental)
Croatia Control	DCAC Cyprus	MUAC	Avinor (Continental)
DCAC Cyprus	DCAC Cyprus		
DFS	DFS		
DHMI	MUAC	NATS (Continental)	
DSNA	DSNA		
EANS	MATS	MOLDATSA	
ENAIRE	DSNA	DFS	NATS (Continental)
ENAV	Avinor (Continental)	NATS (Continental)	
HCAA	HCAA		
HungaroControl	DCAC Cyprus	MUAC	Avinor (Continental)
IAA	DCAC Cyprus	MUAC	Avinor (Continental)
LFV	HCAA	MUAC	Avinor (Continental)
LGS	ANS Finland	MATS	MOLDATSA
LPS	DCAC Cyprus	MATS	
LVNL	ANS Finland	Avinor (Continental)	
MATS	MATS		
M-NAV	MATS	MOLDATSA	
MOLDATSA	MOLDATSA		
MUAC	MUAC		
NATS (Continental)	NATS (Continental)		
NAV Portugal (Continental)	DCAC Cyprus	MUAC	Avinor (Continental)
NAVIAIR	ANS Finland	MATS	Avinor (Continental)
Oro Navigacija	ANS Finland	MATS	MOLDATSA
PANSA	HCAA	MUAC	Avinor (Continental)
ROMATSA	DCAC Cyprus	MUAC	Avinor (Continental)
Sakaeronavigatsia	ANS Finland	MATS	MOLDATSA
skeyes	ANS Finland	Avinor (Continental)	
Skyguide	DCAC Cyprus	MUAC	Avinor (Continental)
Slovenia Control	MATS	MOLDATSA	
SMATSA	DCAC Cyprus	MUAC	Avinor (Continental)
UkSATSE	DCAC Cyprus	MATS	Avinor (Continental)

Appendix 8: Decomposition of the CI