PARAMETRIC REFERENCE AIRPORTS

Peter Biesslich[1]

Hamburg University of Technology, Hamburg, Germany

Bernd Liebhardt[2]

German Aerospace Center (DLR), Hamburg, Germany

Due to increasing passenger numbers, innovative aircraft designs and new passenger processing technologies will likely enter the aviation market in the mid-term. Therefore, airports have to adapt their infrastructure as well as their business models to the upcoming changes, whereof the latter is probably more challenging. Having a closer look at airports worldwide, their specific characteristics are quite different. Existing airport classifications are not adequate for detailed evaluations in research. Thus, this work proposes an approach for obtaining parametric reference airports for analysis and assessment of new technological impacts on airports.

The analysis starting point is a data collection of approximately 15 specific airport parameters, especially infrastructural and operational ones, from 146 airports out of the current biggest aviation markets Asia/Oceania, Europe, and North America. A hierarchical clustering method is used for determining airports having maximum similarity to each other and minimum similarity to airports of other clusters. This cluster analysis leads to a set of airports representing one cluster.

The following investigation shows that there is no single approach being applicable for all kinds of analysis. This work results in 7 functional airport groups. Each of them has distinctive characteristics and is applicable for further analyses on airports.

Keywords: Conceptual Airport Design, Airport Infrastructure, Airport Clustering

Classification: Airport Strategy, Airport and Airline Performance, Aviation Infrastructure

Corresponding author: Peter Biesslich

---

[1] Researcher and PhD student, Institute of Air Transportation Systems, Blohmstrasse 18, D-21079 Hamburg, Germany, Phone: +49 (0)40 42878 4591 E-Mail: peter.biesslich@tu-harburg.de

[2] Researcher and PhD student, Institute of Air Transportation Systems, Blohmstrasse 18, D-21079 Hamburg, Germany

1. Introduction

1.1. Research objective

Airports represent the knots of the air transportation system. On the one hand, there are large international airports, representing the rotation points of large passenger volumes, and on the other hand, there are hundreds of regional airports, running feeder traffic or point-to-point connections. Due to the high variety of airports, it makes sense to build functional groups for analysis and simulation purposes in research. Thus, the question of airport types and their characteristics arises.

Literature illustrates some approaches for grouping airports, yet usually distinguishing them by defined threshold values. Airport Council International (ACI), an association of airport operators, classifies four groups of airports simply using their yearly passenger volumes [1]. This distinction is very generic and non-adequate for detailed research evaluations.

This work pursues the following objectives: Firstly, airports shall be aggregated to functional groups based on statistical data and the application of clustering techniques, in contrast to rigid threshold vales. Parameters reflecting the importance of airports within the air transport network play a subordinate role compared to their infrastructural dimensions and operational performance. Furthermore, the final number of clusters should be five to eight, representing a definite point of time which here is 2010. Finally, out of the three performance areas infrastructure, operations, and business model, a set of parameters will be available for typical reference airports which can be applied easily for further analysis. The geographical location as a factor of influence is disregarded deliberately.

1.2. Literature review

In the field of airport research, the definition of reference airports and the usage of cluster analysis are not widespread. Azzam [2] introduces a new airport taxonomy based on flight plan data from 1979 to 2007 and network performance figures. By this, specific statements can be made about the evolution of airports and their function within the air transportation network at a defined point in time, putting them into a geographical context. Azzam uses a hierarchical cluster analysis, defining 12 different airport classes based on six network parameters.

Oettel et al. [3] use clustering techniques to develop an application-oriented airport classification for air traffic simulation purposes. In particular, the single linkage algorithm is applied to identify outliers and similarities among a set of airports which in this case are European secondary hubs. Oettel et al. indicate that it makes sense to limit the set of parameters according to the application. The more parameters are considered for classification, the bigger the application field becomes, but also the smaller the sample and the representativity of results.

In contrast to [2], the paper at hand does not focus on network parameters but on parameters reflecting the infrastructure and performance of an airport. The aspect of importance within the global network can be modelled at a later stage by a *Movement Share of Airline Type* parameter.

2. Principles of cluster analysis

Clustering is an algorithm for forming functional groups, whereby the objects of one group feature maximum similarity and minimal similarity to objects in other groups [4]. Countless methods of clustering can be found in literature. Hereafter, the principle and the proceeding of cluster analysis is outlined briefly, focussing on hierarchical, agglomerative clustering techniques which are applied in the following investigation.

2.1. Choice of parameters

Initially, a parameter selection has to be made which is implemented in the classification later on. It is important to recognise that the number of parameters should be chosen as low as possible, as on the one hand, computing time rises with an increasing number of parameters, and as on the other hand, there is a risk of finding a number of clusters that is above an interpretable extent or barely distinguishable [2]. Reducing the number of parameters on an experimental basis is one possibility for parameter choice. Alternatively a correlation analysis can be used to identify correlating parameters. Introducing well-established methods of correlation computation:

Correlation coefficient by Bravais-Pearson [5]

The Bravais-Pearson correlation coefficient is calculated by dividing the empirical covariance $\tilde{s}_{XY}$ by the product of the standard deviations $\tilde{s}_X$ or $\tilde{s}_Y$ of both features. This method can capture only linear coherences.

$$r = r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\tilde{s}_{XY}}{\tilde{s}_X \tilde{s}_Y} \tag{2.1}$$

Correlation coefficient by Spearman [5]

Determining the strength of a non-linear but monotonous coherence, the Spearman correlation analysis can be applied. The rank *rg*, sorted by size, is allocated to the original x and y values independently.

$$r_{SP} = \frac{\sum(rg(x_i) - \overline{rg}_X)(rg(y_i) - \overline{rg}_Y)}{\sqrt{(rg(x_i) - \overline{rg}_X)^2 (rg(y_i) - \overline{rg}_Y)^2}} \tag{2.2}$$

On both methods, the correlation coefficient r or r_{SP} can take a value between -1 and 1, whereby r = 1 represents the extremity of a consensual linear coherence and r = -1 represents the extremity of an opposing linear coherence. In the case of r ≈ 0, no linear coherence or no correlation exists. Nevertheless, a high correlation coefficient does not imperatively indicate a causal connection between two features. Therefore, it is mandatory to add practical reasoning.

2.2. The cluster algorithm

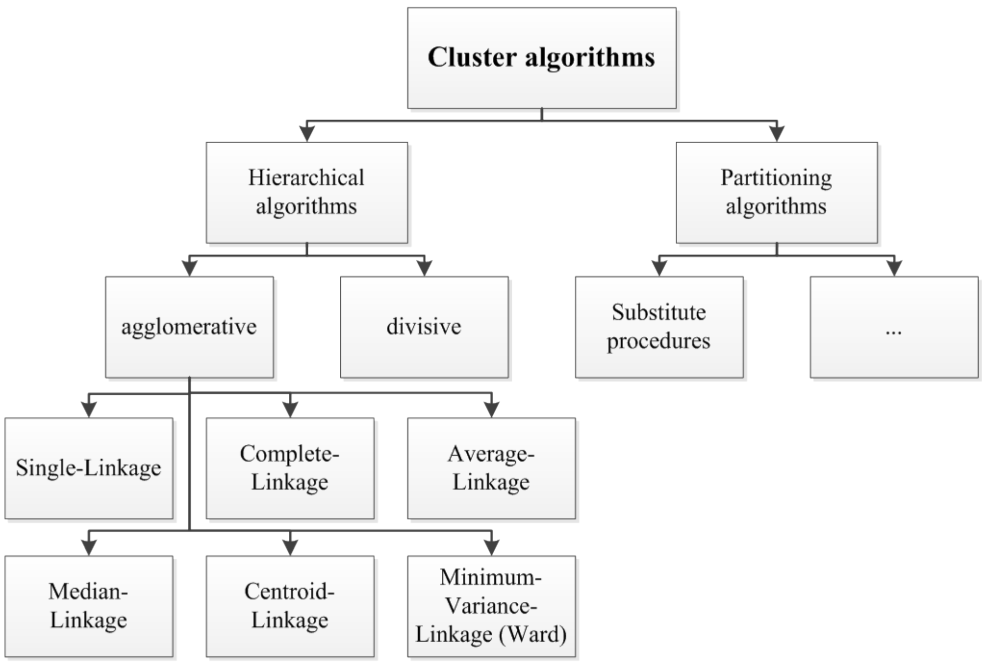

Figure 1: Cluster algorithms[3] [6]

[3] This figure is not intended to be exhaustive.

Figure 1 gives an overview about known cluster algorithms. The most common one probably is the hierarchical cluster algorithm, which can be subdivided into an agglomerative and a divisive approach. In the former, every object represents one cluster initially, incrementally merging into bigger clusters until all objects are concentrated in one cluster. The converse, divisive approach is seldom used. Advantages of the hierarchical approach are flexibility regarding the number of clusters and non-specification of initial conditions. At the same time, the absence of remapping after the initial cluster assignment represents a drawback, frequently producing non-optimal results.

Partitioning algorithms (whereof the best known representative is the so-called *k-means algorithm*) start with the determination of a fixed cluster classification, followed by an iterative optimizing process that consists of a relocation of objects to adjacent clusters until an optimum is found. This optimization reflects the greatest advantage of partitioning algorithms. Yet, there are two disadvantages: Firstly, the predetermined number of clusters is hardly estimable at the beginning of an investigation. Secondly, the random allocation of cluster centroids produces different results on every run. [6]

2.3. Measure of similarity / proximity d(X, Y)

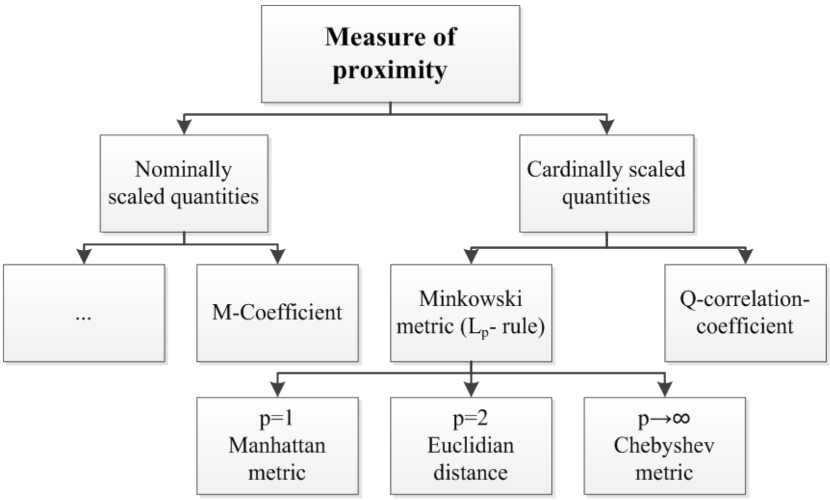

Figure 2: Measures of similarity / proximity

The determination of similarity or proximity between objects is essential for clustering. The choice of similarity measures depends on the form of scaling of the objects. For cardinally scaled quantities, the distance between the objects is usually used. This distance can be calculated using the so-called Minkowsi metric (derived from the L_p rule, see formula 2.3). Examples include the Manhattan metric, the Euclidean distance and the Chebyshev metric, whereby the Euclidean distance (L_2 rule) equals the human understanding of distances [4] and is therefore often used. [2]

$$d(X,Y) = \|X - Y\| = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{2.3}$$

To avoid the influence of features with large values that can overlay features with smaller values in the classification, the data are normalized to a uniform scale in advance of the clustering. This is done by a Z-score standardization following formula 2.4, whereby $\bar{X}$ represents the mean and $SD(X)$ the standard deviation [4].

$$X^* = \frac{X - \bar{X}}{SD(X)} \tag{2.4}$$

2.4. Fusion algorithm (Linkage rule)

The linkage rule determines the merging criterion for objects or clusters at each iteration step, based on the chosen measure of similarity. The calculation is performed for all possible pairs of objects d (X, Y). In the subsequent paragraph, only hierarchical agglomerative algorithms are discussed.

Using the *single-linkage* algorithm, the minimum distance between two objects decides on the merging of clusters. The algorithm is well suited for finding elongated and thread-formed clusters, which recognized either insufficiently or not at all by other algorithms. In contrast, the maximum distance of an object pair is used to determine the similarity in *complete-linkage*, resulting in compact clusters of approximately the same size. One disadvantage of both methods is a very high sensitivity to outliers. This can be avoided by using the *average-linkage* algorithm. As the name implies, the mean distance between all pairs of objects determines the similarity. Comparable to this method is the *centroid-linkage*. Here, the centres of the clusters are calculated first, significantly reducing the computational complexity. Another method is the *minimum-variance-linkage* or *Ward-linkage*. The algorithm merges the cluster pair that causes the smallest increase of the sum of squared errors, thus ensuring a maximum homogeneity per time step. Ward’s advantage is a low sensitivity to outliers whilst tending to form same-sized clusters. Having calculated the distances between all pairs of objects, the two objects merge which are most similar. This iterative process continues until the termination criterion is reached. [2]

The choice of the appropriate clustering algorithm largely depends on the investigated objects. There is no single approach that is applicable for all kinds of analysis. Therefore, it is important to have a basic understanding of the desired results and previous knowledge of the research topic.

2.5. Evaluation of cluster analysis

Special diagrams are applied for visualizing the results of cluster analysis. One is the dendrogram (see Figure 3, left), whereby the clustering is depicted by a hierarchical tree. The x-axis reflects the computed distance between two objects. In other words: The longer the "branch" of the hierarchical tree, the more dissimilar the clusters. The number of clusters is determined by an appropriate linkage distance. The structogram (see Figure 3, right) which is the representation of the number of clusters over the linkage distance can support a decision. If a distinct "knee" can be seen in the course of points, a reasonable number of clusters can be read off directly. Sometimes, the course is smooth (as shown), not allowing for a direct statement about the number of clusters.

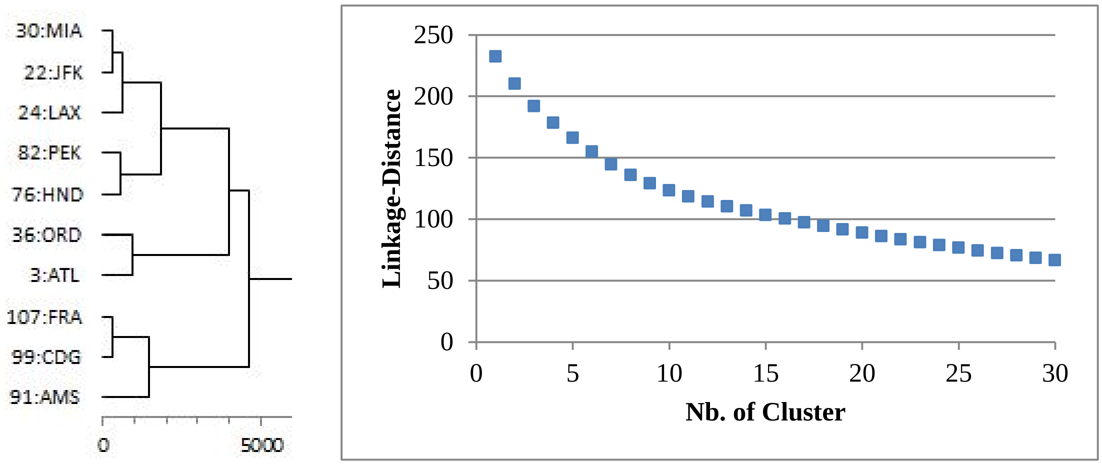

Figure 3: Dendrogram (left) and structogram (right)

As an additional aid, the so-called box-plot (five-point summary, Figure 4) is used. This diagram is well suited for a quick comparison of intra-cluster distributions. In addition to the median, the lower and the upper quartile are depicted by a "box". The smaller the median-to-edge distance, the more oblique is the distribution. Information on the spread of the cluster is given by the interquartile range (IQR) which is the difference between the upper and lower quartile, i.e. the length of the box. The "fence" outside the box represents a distance of 1.5

times IQR. Data points outside this fence show potential outliers that might require further consideration. [5]

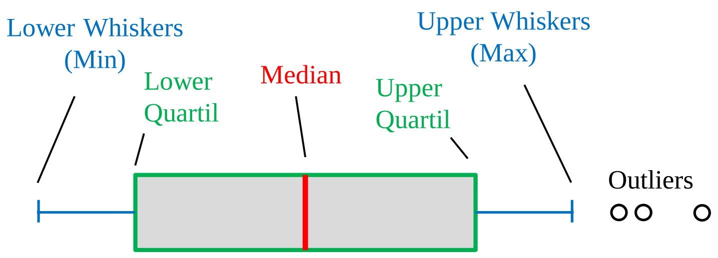

Figure 4: Box plot [5]

3. Determination of reference airports

3.1. Choice of parameters

For a description of reference airports, the following parameters have been selected out of the three key performance areas: infrastructure, operations, and business model. This preliminary selection of parameters might need expanding in case new data become available or in case new topics need to be assessed.

Total Passengers
Total Movements
Cargo
Share of Transfer Passengers
Share of International Passengers
Distance to City Centre
Terminal Size
(Movements Share of Airline Type)[4]

Number of Runways *(+ Length/Width)*
Runway Layout/Operation Mode
ILS Category
Number of Aircraft Stands
Traffic Mix (Super/Heavy/Medium/Light)
Share of Aeronautical Revenue
Share of Non-Aeronautical Revenue
Total Revenue
Revenue per Passenger (PAX)

Obtaining aggregated data for a representative amount of airports proved to be great a challenge. The current list of 146 airports is based on the ATRS Airport Benchmarking Report [7], supplemented by ICAO data [8] and featuring some additional German airports, all referring to the year 2010. Hereby, the absence of airport data from South America or Africa represents a certain shortcoming.

The parameters *Traffic Mix*, *Transfer Passengers* und *Movement Share of Airline Type* have been calculated using the commercial Airport Data Intelligence (ADI) web dialogue [9]. The

[4] This parameter set could not be included in the study due to lacking data ("Unknown Carrier") that could not be obtained in the short term.

present database is lacking capacitive performance of airports. However, parameters such as runway or terminal capacity and landing fees are planned to follow.

3.2. Procedure

As initially formulated, we had the objective of finding 5-8 functional airport clusters. Based on an independently performed trial-and-error process whilst following the cluster analysis of Assam [2], a hierarchical agglomerative clustering method is applied. Representing the most suitable linkage rule, the *minimum-variance-linkage/Ward-linkage* has been selected. Furthermore, the Euclidean distance is chosen as a measure of similarity, and prior to the cluster analysis, Z-score normalization is executed. In the following, two different cluster analyses are presented, whereby the cluster input parameters vary. The first cluster analysis is evaluated in detail, and its results are subsequently compared to those of the second analysis.

3.3. Cluster analysis 1

On the first approach, the initial cluster parameters were determined based on the authors' experience, not by means of a correlation analysis. For including all 146 airports in the cluster analysis, revenue parameters were neglected due to lacking availability. The Ward-linkage appeared to be robust and best suited for the analysis. Simultaneously, centroid-linkage produced almost exactly the same results. Table 1 shows an overview of the first analysis.

Table 1: Overview of cluster analysis 1

Input parameters	Cluster algorithm	Result
▪ Total Passengers	▪ Ward Linkage	+ 7 groups
▪ Total Movements	▪ Euclidian distance	+ 146 objects
▪ Cargo	▪ Z-score normalization	+ various large clusters
▪ Nb. of Runways		
▪ Transfer Passengers		

Figure 5 illustrates the dendrogram of the first cluster analysis. Since predictably, the structogram (not shown here) featured no distinct knee, the clustering was established at an arbitrary linkage distance (dashed red line), resulting in 7 airport clusters. The clusters are color-coded for clarity. The following explanations do not address all parameters of the clusters and only selected results are presented.

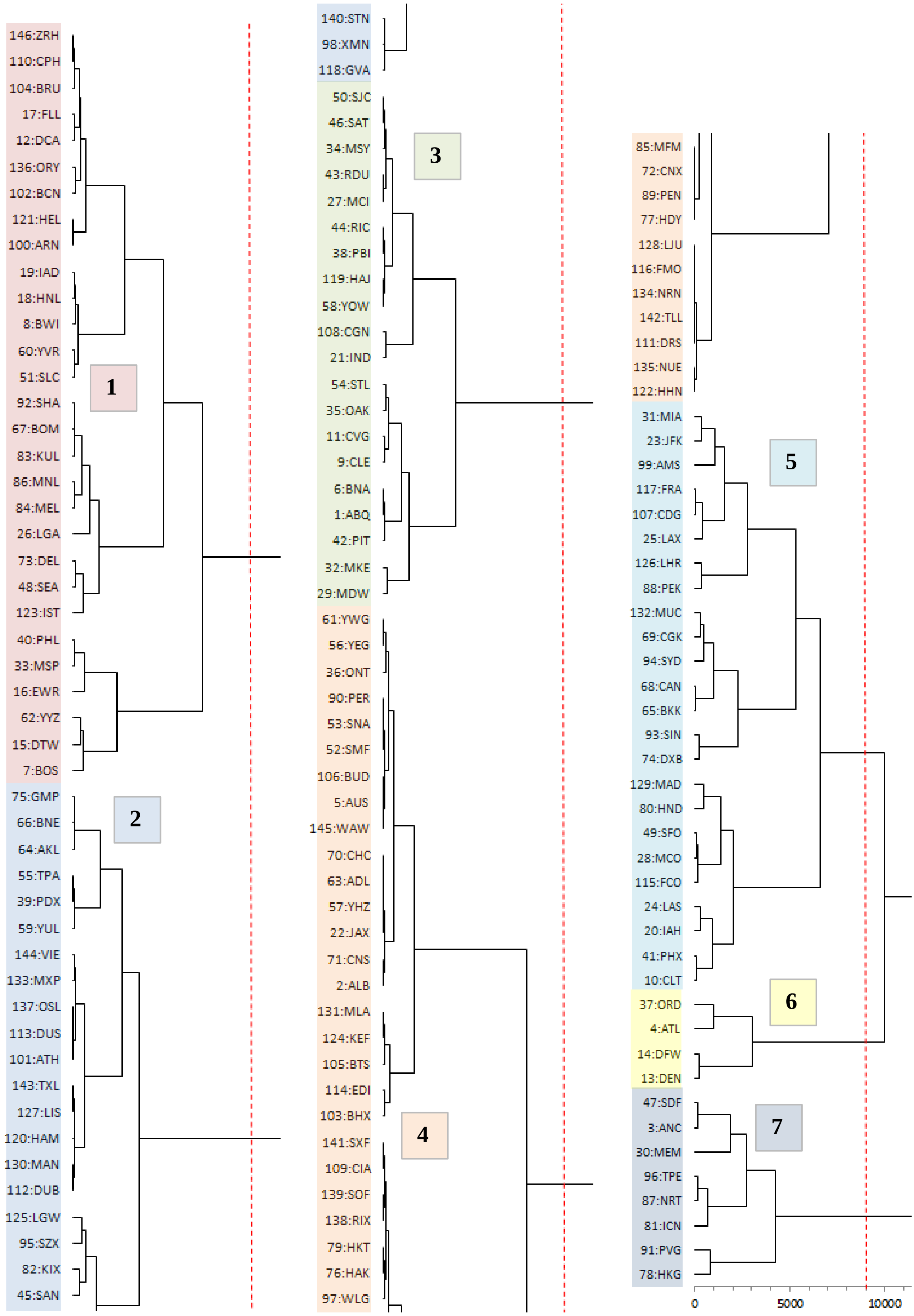

Figure 5: Dendrogram of cluster analysis 1[5]

[5] Due to size, the dendrogram is presented in three parts: part 1 left-hand, part 2 at centre and part 3 right-hand. The axis of part 3 is valid for all three parts.

The statistical distribution within the clusters was checked with the help of box plots. Figure 6 examines the parameters *Total Passengers*, *Total Movements*, and *Cargo* for all clusters, showing the original clustering on the left-hand side and the one after the manual reallocation of individual outliers (e.g. object 29 in the first figure) on the right-hand side.

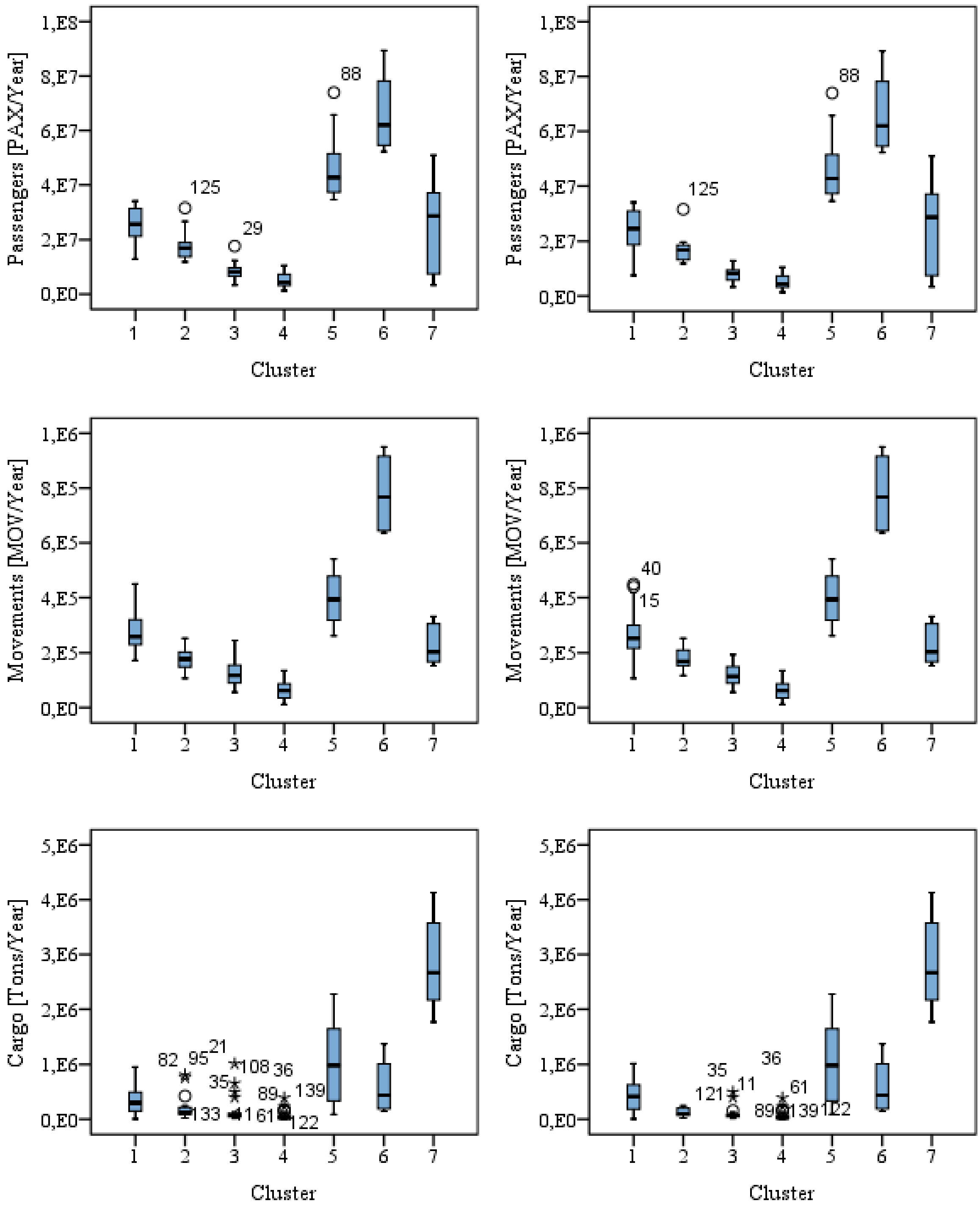

Figure 6: Box plots of cluster analysis 1 - left: original, right: outliers edited

The largest improvement was achieved for the *Cargo* parameter. The objects 21, 82, 95, 108, and 133 were manually assigned to cluster 1, whereby the spread in aircraft movements within the first cluster increases only modestly compared to the original distribution. The two new outliers in the *Total Movements* graph are close to the upper whisker value. However, the outliers in clusters 3 and 4 cannot be moved to cluster 1, as they significantly affect the distributions of *Total Passengers* and *Total Movements* of the first cluster. On the other hand, these are only identified as outliers of clusters 3 and 4 since the distributions themselves are very compact. Furthermore, the parameter *Revenue* (not shown here) recorded single outliers which have not been remapped due to a variety of aftermaths. Finally, Table 2 shows the average reference values for all clusters that will be named and described quantitatively and qualitatively in the following.

Table 2: Overview of cluster-specific average reference values

Cluster	Passengers [PAX/Year]	Movements [MOV/Year]	Cargo [Tons/Year]	Transfer PAX [%]	International PAX [%]	RWYs	RWY Layout/Operation
1	25.460.051	282.228	350.443	16,48	36,91	3	parallel/independent
2	17.236.402	175.610	199.145	15,05	54,32	2	single runway
3	8.477.082	126.338	175.760	12,79	9,36	4	parallel/dependent
4	5.001.488	65.298	59.295	8,20	38,03	2	single runway
5	45.843.935	398.439	1.038.753	38,84	54,03	3	parallel/independent
6	66.371.853	780.469	599.877	43,34	32,80	6	system/independent
7	25.098.445	229.578	2.838.209	20,46	46,88	3	parallel/independent

Cluster	Traffic Mix [%]				Aeronautical Revenue [%]	Non Aeronautical Revenue [%]	Revenue/PAX [US$]	Revenue [US$]	Distance to city center [km]	Terminal size [m²]
	S	H	M	L						
1	0,00	17,24	81,97	0,79	51,56	44,78	16,67	408.486.492	19,87	209.432
2	0,00	15,72	83,64	0,65	49,57	50,39	22,48	384.323.635	24,01	147.259
3	0,00	2,29	97,26	0,45	48,26	46,70	12,30	105.839.935	17,87	88.731
4	0,00	8,29	90,46	1,25	47,98	45,70	18,18	93.663.514	18,97	61.533
5	0,33	29,69	69,65	0,33	49,75	50,23	19,15	907.113.601	23,72	536.025
6	0,00	10,39	89,12	0,49	53,10	47,00	8,86	553.735.563	31,03	529.244
7	0,06	38,88	60,14	0,91	49,78	50,17	26,05	828.540.047	36,90	455.890

Small regional airports constitute cluster 4 which is also the largest one, featuring 38 objects. On average, about 5 million passengers use these airports annually on 65,000 aircraft movements. With around 60,000 tons, freight plays a negligible role. The airports mostly operate just one runway with a maximum of two, and 90 % of servicing aircraft are medium-sized. Their mean proximity to the respective city centre is 19 km. For example, this cluster includes Albany International Airport (ALB) and Edinburgh Airport (EDI).

Having similar distances to the city centre, airports of the 3rd cluster can be described as ***Medium regional airports***. Compared to the Small regional airports, they have four runways on average (usually parallel but dependent) which is mainly due to the above-average representation of American airports, e.g. Chicago Midway Airport (MDW) or Cleveland-Hopkins International Airport (CLE). Thus, utilization rises to around 126,000 aircraft movements, 8.5 million passengers and 176,000 freight tons annually. The share of medium sized aircraft is the highest of all classes with 97%.

With about 17 million passengers and 175,000 aircraft movements per year, cluster 2 consists of ***International airports***, providing both point-to-point and long-haul traffic, visible by the heavy aircraft share of about 15 %. The revenue per passenger is the second highest of all classes and amounts to US$22.48. Regarding infrastructure, these airports are equipped with fewer runways (usually two) compared to airports of the third cluster. Remarkably, the largest passenger traffic figures in this class are generated by single-runway airports like London Gatwick International Airport (LGA) and Shenzhen Baoan International Airport (SZX). With nearly 200,000 tons of cargo, International airports reside at the lower end of the scale.

Crucially important to the air transportation network, the airports of cluster 1 can be referred to as ***Secondary hub airports***, containing Newark Liberty International Airport (EWR) or Zurich International Airport (ZRH), for instance. Compared to cluster 2, aircraft movements increase to 282,000 departures/arrivals and cargo volumes rise to 350,000 tons. Although over 25 million passengers travel by these airports on average, only revenues of US$16.67/PAX are achieved, probably caused by higher operating costs for an average of three, mostly independent runways. The traffic mix of 17 % heavy aircraft and 82 % medium-sized aircraft is similar to the distribution of cluster 2.

Cluster 5 includes the major ***International hub airports***, e.g. Frankfurt Airport (FRA) or Los Angeles International Airport (LAX). This is obvious from the high proportion of transfer passengers (38.84 %) and from the share of heavy aircraft (ca. 30 %). In addition, about 60 % of the airports in this class are approached by the currently largest airliner, the Airbus A380. This results in an annual passenger volume of almost 46 million and in 400,000 aircraft movements. The second "revenue pillar" of these airports is cargo with a volume of about 1 million tons per year, mainly generated by freight carried in passenger aircraft (belly freight). The infrastructure consists of an average of 3 runways that are usually run independently. With a yield of US$19.15 per passenger, this class is located in the midfield; however, a mean total revenue of over US$900 million constitutes the first place.

A typical representation of major U.S. airports, referred to as ***High-frequency hubs***, is reflected in cluster 6. It includes, for example, Hartsfield-Jackson Atlanta International

Airport (ATL) or Chicago O´Hare International Airport (ORD). On the one hand, the number of flight movements is 780,000 per year which is by far the largest value of all classes. On the other hand, being only 10 %, a very low share of heavy aircraft is observed. This high number of flight movements is distributed to 6 runways on average, which allows for passenger volumes of around 66 million annually. The high frequency of flights is also due to a variety of transfer options for passengers, reflecting in the highest transfer passenger share of 43.34 %. The average cargo volume handled comes down to 600,000 tons. Despite these figures, only a yield of US$8.86/PAX is generated which is the lowest value of all classes.

Airports of the final cluster 7 are characterized by an annual cargo volume of 2.8 million tons and are therefore entitled as ***Cargo hubs***. 25 million passengers and 230,000 aircraft movements rank in the range of cluster 1. Due to the large volume of freight, a total income of US$828 million is generated, second only to the International hub airports. At the same time, this class achieves the best revenue per passenger value with US$26.05, caused by the relatively low numbers of passengers. As cargo aircraft are mostly large, this cluster holds the highest share of heavy aircraft which is around 39 %. This cluster features, for instance, Memphis International Airport (MEM) and Hong Kong Chek Lap Kok International Airport (HKG).

3.4. Cluster analysis 2

In contrast to the above-mentioned first cluster analysis, a regression analysis is performed to reduce the number of parameters, whereby correlating features are identified by calculation. The number of parameters is reduced incrementally, involving only cardinally scaled parameters (see Ch. 3.1, bold print). The parameters *Share of Aeronautical Revenue/Non Aeronautical Revenue* were not included due to obvious correlation. Since there are mostly busy airports in the sample, *Traffic Mix Light* is on average below 1%. The same applies to *Traffic Mix Super*, since currently, only the Airbus A380 and the Boeing B747-8 belong to this category. As a consequence, the two remaining parameters are strongly correlated (sum ≈ 1), and therefore, *Traffic Mix Medium* was not included. Table 3 shows an overview of parameters and initial results of the second analysis.

Table 3: Overview of cluster analysis 2

Input parameters	Cluster algorithm	Results
▪ Total Passengers	- Ward Linkage	+ (6) 8 groups
▪ Cargo	- Euclidian distance	- 131 objects only
▪ Nb. of Runways	- Z-score normalization	- one cluster with 3
▪ Traffic Mix Heavy		objects (8 classes)
▪ Revenue per PAX		

In contrast to the first analysis, the second one includes the parameters *Traffic Mix Heavy* and *Revenue per Passenger* instead of *Total Movements* and *Transfer Passengers*. This entails the negative affect that not all objects are covered by the clustering due to a lack of revenues.

In consideration of the dendrogram, there are two possible clusterings. The box plot distributions for 6 clusters show a number of outliers that even appear in the parameters *Total Passengers* and *Cargo* which were used to form the cluster. This disadvantage is only partially improved by the redesign of 8 clusters.

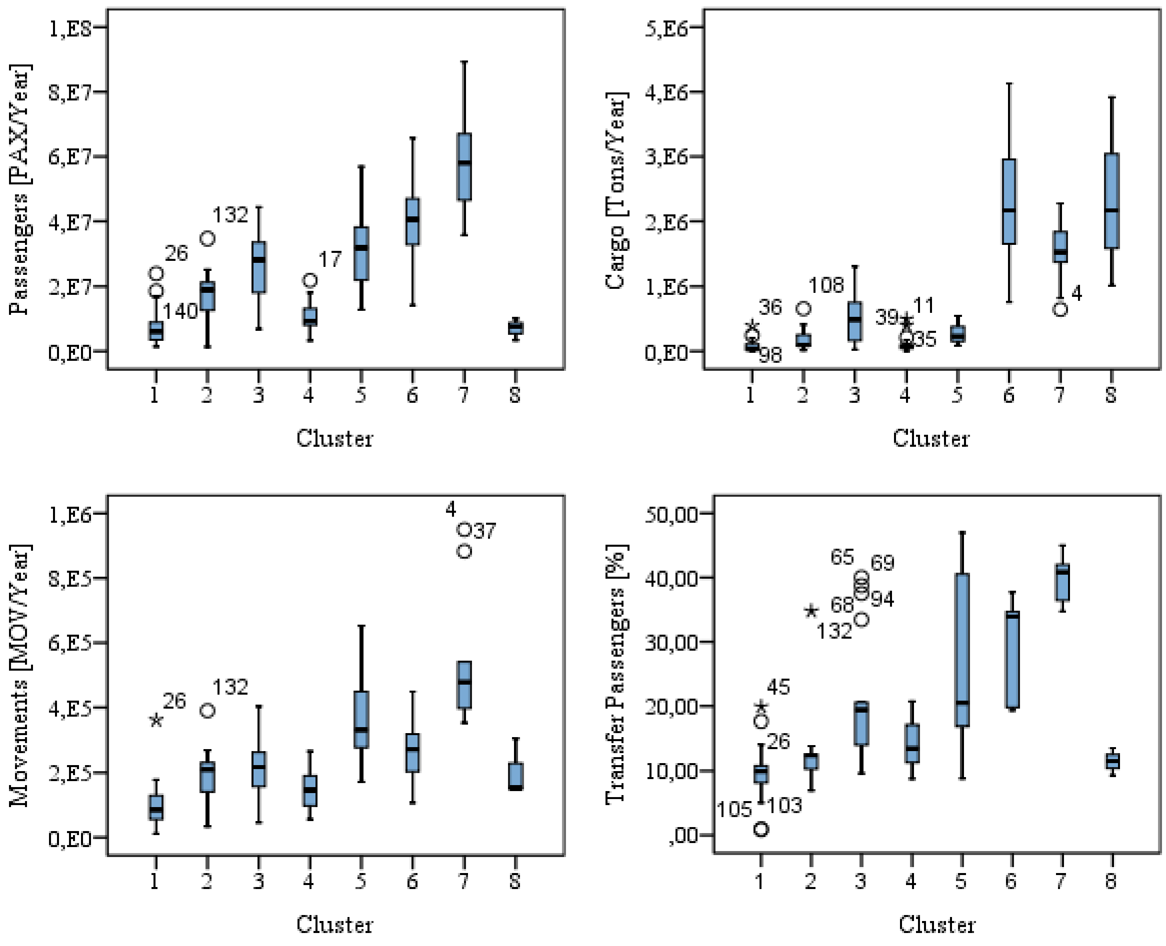

Figure 8: Box plots of cluster analysis 2 with 8 clusters

Figure 8 depicts, amongst others, the spread within the clusters 5-7 and 3 for the parameter *Total Passengers*, revealing an obvious deterioration compared to the first clustering approach (see Figure 6). The *Cargo* parameter within the clusters 6-8 shows similar spreads. Furthermore, a multitude of outliers are found regarding other parameters, for example in *Transfer Passengers* within the clusters 1 and 3, not being reallocateable to adjacent clusters.

Two more analyses were performed, showing weaker results compared to cluster analysis 1 as well. Therefore, they are not discussed further.

4. Conclusion and future work

The performed cluster analysis is a first approach to the desired objective of applying clustering techniques and of obtaining 5-8 functional airport groups only based on statistical data. The 7 clusters resulting from the first run represent a variety of airports within one manageable set. In particular, some clusters are already in harmony and fulfil the initial expectations. As mentioned in literature, the selection of input parameters used for the cluster analysis and the objective of the clustering are of paramount importance.

Cluster analysis 2 leads to the assumption that the addition of revenue parameters leads to an interpretation of too many dimensions and to an infeasibility of adequate airport classification. Another possible reason for the large spread in the cluster could be the variety of factors that affect the revenues of an airport, ranging from the general business model and country-specific charges up to non-aviation activities. Thus and in this approach, the focus was placed on infrastructure and the technical performance parameters for clustering. For investigations having a different focus, a fresh classification is advisable.

During the investigation, several obstacles for proper analysis were observed. One of these is the availability of data, especially in the area of revenues. Furthermore, different definitions of parameters cause susceptibility to errors. Aircraft movements vary by data source, depending on whether certain types (general aviation, non-commercial flights etc.) are included or not.

In order to expand the bandwidth of parameters for these reference airports, a similar procedure of airline clustering will be performed based on the work presented. The results will contribute to a *Movement Share of Airline Type* in the airport parameter set. Moreover, an extension to South American and African airports is desirable to allow for globally valid airport analyses.

The definition of reference airports in the wake of clustering is the first step towards further analyses in this research area. Depending on the object of study, reference airports deliver the input parameters for a parametric, system dynamic airport model. This model shall constitute the parameter networking and future changes, for example the construction of a new runway, over a pre-defined time period. Results of such studies could be the economic assessment of investments in airport infrastructure (keyword: net present value analysis), but also an evaluation of operational benefits.

References

[1] Schulz A. et al., “*Flughafen Management*“, 4. Edition, Oldenbourg Verlag, München, p. 8, 2010

[2] Azzam M., “*Eine Analyse der Evolution und Effizienz des Luftverkehrsnetzwerkes”*, Dr. Hut, Darmstadt, p. 57ff, May 2012

[3] Öttel G., Böck P., “*An approach to airport categorization - category definition and quantification for air traffic related simulation purposes”*, Journal of Aerospace Operations 1 (2012), p. 155-170

[4] Larose D., “*Discovering Knowledge in Data: An Introduction to Data Mining”*, John Wiley & Sons, Inc., 2005, p. 37, 99, 147

[5] Fahrmeier L. et al., “*Statistik - Der Weg zur Datenanalyse”*, Berlin, Springer, 2004, p. 66f, p. 134ff

[6] Bacher J. et al., *“Clusteranalyse - Anwendungsorientierte Einführung in Klassifikationsverfahren”*, Oldenbourg Verlag, München, 3. Edition, 2010

[7] Air Transport Research Society (ATRS), *“Airport Benchmarking Report - 2012”*, Vancouver, June 2012

[8] ICAO Database [online data base], URL: http://www.icaodata.com/default.aspx

[9] Sabre Airlines Solutions, *Airport Data Intelligence (ADI)* [online data base], URL: http://www.sabreairlinesolutions.com/home/software_solutions/airports