Journal of the Air Transport Research Society 3 (2024) 100035

Contents lists available at ScienceDirect

Journal of the Air Transport Research Society

journal homepage: www.elsevier.com/locate/jatrs

A binary logistic regression model to identify key aspects that enhance global hub airports status

Gabriel Bordeaux[a,*], António Couto[a]

[a] Research Centre for Territory, Transports and Environment, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

ARTICLE INFO

Keywords:
Hub airports
Binary logistic regression
Key aspects

ABSTRACT

This paper provides a tool to identify key aspects for an airport to achieve global hub status for a given airline and determines whether these factors are related to the facility's infrastructure, its region, or both. Despite the frequent use of the term 'hub', there is little academic consensus on its exact definition in air transport. Many define a hub based on passenger numbers rather than the concentration of flights and passengers from the main carrier. This study addresses this gap by analyzing the factors influencing the definition of a hub and the commonalities among global hubs. Data from 300 major airports, including internal variables (runways, terminals, gates and area) and external variables (economy, population, attractiveness), were collected. A Binary Logistic Regression (BLR) model identified key aspects influencing hub status, with the assistance of an Exploratory Factor Analysis (EFA) that grouped the variables into factors. The binary 'hub' variable was defined by the main carrier's activity and the Global Airport Connectivity Index (GACI). The factor with the highest coefficient primarily involved internal variables and, to a lesser extent, global attractiveness and population. The factor with the lowest coefficient related to the region economy. The BLR correctly identified hub status in 93.3 % of cases, with 68.3 % accuracy for hub airports. Airports not correctly identified by the model mostly present a lack or underutilization of existing infrastructure.

1. Introduction

Hubs are special facilities that serve as switching points in distribution systems, concentrating flows to exploit economies of scale and reduce transport costs. They consolidate flows from the same origin with different destinations and combine them with flows from different origins to the same destination (Alumur and Kara, 2008). Various studies address algorithms for hub location problems. Wandelt et al. (2022) conducted an experimental benchmark comparing several of these algorithms and identified patterns in solution similarity and economies of scale. The hub-and-spoke network, based on central airports (hubs) connecting passengers to final destinations (spokes), emerged after regulatory changes and the advent of widebody aircraft. This model became prominent post-Airline Deregulation Act in the US (1978), transforming the industry and leading to new airlines and bankruptcies of some traditional ones. The growth of hubs in Europe after air transport liberalization indicates a global trend (O'Kelly, 1998; Button, 2002). The hub-and-spoke configuration, adopted by the majority of legacy carriers, optimizes fleet use and aircraft load factor to cover a greater number of cities served by each airline. This model may require a variety of aircraft sizes from the same airline, which increases the complexity of fleet capacity management and pilot training, so full-service carriers can hire regional airlines to serve locations with less demand. Despite passenger inconvenience, hub-and-spoke systems expands the range of destinations by an airline and promote territorial integration by allowing smaller communities to join airline networks. Another issue is that passengers in hub cities with dominant airlines often pay a "hub premium" fare for direct connections. Since connecting flights through a hub are usually cheaper than direct flights, passengers may use a controversial practice called skiplagging. In skiplagging, a passenger booking a multi-leg itinerary (e.g., A–B–C) only takes the first leg (A–B), benefiting from lower fares, especially on intercontinental routes (Sun et al., 2023).

Key advantages include: (a) economies of scope – spreading costs across services, (b) economies of density – high infrastructure utilization, (c) economies of scale – lower costs per passenger, (d) increased load factor, and (e) network enlargement (Uher, 2005). Environmentally, hubs have mixed impacts; while they contribute to global aviation growth and local pollution, they result in fewer flights and lower fuel use per passenger, presenting lower emissions compared to point-to-point networks (Button, 2002; Tian et al., 2020). Hubs also have weaknesses, such as high operating costs due to (a) additional takeoff and landing fees, (b) aircraft and crew utilization limited by network geography,

* Corresponding author.
E-mail address: gabrielfbordeaux@gmail.com (G. Bordeaux).

https://doi.org/10.1016/j.jatrs.2024.100035
Received 28 June 2024; Received in revised form 26 July 2024; Accepted 26 July 2024

connection timing and hub congestion. They are susceptible to delays, increasing missed connections and mishandled baggage risks (Cook and Goodwin, 2008). Other disadvantages include: (a) flight delays, (b) fluctuating capacity demand, (c) congestion during peak hours, (d) increasing fees, and (e) higher ticket prices (Uher, 2005). Longer ground times reduce aircraft and crew utilization, adding costs (De Neufville et al., 2013). Additionally, hubs often rely on the strategic decisions of their dominant carrier (Button and Lall, 1999). If at any time a de-hubbing occurs — meaning the hub carrier ceases operations at the airport due to bankruptcy, mergers, or cost-cutting measures (Redondi et al., 2010) —airport operators must act to fill the void left by the former airline with another hub carrier or low-cost carriers. Hubs depend more on airline network considerations than on regional economic characteristics (Martín and Román, 2003). While low-cost carriers focus almost exclusively on point-to-point traffic with short-term planning, full-service airlines build long-term relationships with airports, evaluating attractiveness based on traffic potential beyond the hub. Airports typically offer incentives to attract new airlines, such as reduced charges for new destinations and quantity discounts based on passenger volume (Bilotkach and Bush, 2020).

This paper aims to identify key aspects for an airport to achieve global hub status. The literature lacks a clear definition of the criteria for classifying an airport as a hub, which may relate to infrastructure capacity or be influenced by the region's socio-economic factors. By analyzing global airport data with statistical models, commonalities among hubs can be identified, explaining their differences from non-hub airports. The objective is to understand essential characteristics and global location factors attracting airlines to establish hubs, and whether internal or external factors play a more significant role.

1.1. Background

1.1.1. Hub classification

Studies have classified airports as hubs or non-hubs based on connectivity and traffic generation (Burghouwt and Hakfoort, 2001; Rodríguez-Déniz et al., 2013). Airport categories include: (a) primary international hubs with high share of transit traffic and large catchment areas, (b) secondary hubs with lower transfer traffic but sizable catchment areas, (c) international gateways with large catchment areas but are used as a hub only for long-haul flights, and (d) regional/small airports with minimal transfer traffic and smaller catchment areas, typically served by regional airlines and low-cost carriers. Rodrigue (2020) noted that the share of connecting passengers determines an airport's hub level, with higher percentages indicating more significant hubs.

Similarly to how Paulwels et al. (2024) depicted various meanings for the concept of a regional airport in aviation studies and proposed a uniform definition, there is also an absence of a standard classification for what can be considered a global hub airport. The term "hub" is normally ambiguously used, sometimes referring to the busiest airports without considering their role as a carrier's operations center. The Federal Aviation Administration (FAA) classifies US hubs based on passenger enplanements, but other studies also consider connectivity, transfer traffic, and accessibility (Ivy, 1993; Ryerson and Kim, 2013; Rodríguez-Déniz et al., 2013). Market concentration measures classify hubs where the main carrier handles over 50 % of local traffic (Button, 2002). Thus, a hub is not just a large airport but one that generates significant transit passengers (Redondi et al., 2010; Costa et al., 2010). However, detailed data on connecting passenger traffic is often unavailable (Maertens et al., 2020).

This present study defines an airport as a global hub facility when it meets two key criteria: the airport must have strong global connectivity; and the airport's primary airline should have a significant market share, offering a diverse range of destinations, including intercontinental long-haul flights, and maintaining numerous code-share agreements with partner carriers.

1.1.2. Internal aspects

In order to discern which attributes to adopt for delineating the internal characteristics of airports in the model described in this paper, it is necessary to consult studies on airport efficiency and productivity available in the literature, given the absence of similar research addressing hub-related aspects. Utilizing techniques such as data envelopment analysis and similar methods, these studies describe the gains (outputs) derived from specific resource allocations (inputs). Consequently, each airport's resources exemplify inherent characteristics of the facility, typically chosen based on airport infrastructure, designated zones within the airport site, operator assets, number of works, and labor costs. Outputs, in most cases, encompass the number of passengers and cargo movements, as well as aircraft operations — characteristics of the facility's production. Thus, the inputs outlined in these studies closely align with internal aspects of airports that influence their production and status, while the outputs contribute to the puzzle of defining what qualifies an airport as a hub. With this in mind, the internal aspects implemented in the model were as follows: (a) total runway length (Merkert and Assaf, 2015; Yoshida and Fujimoto, 2004), (b) passenger terminals — including terminal area, apron area and number of gates (Chaouk et al., 2020; Fragoudaki et al, 2016; Merkert and Mangia, 2014), and (c) airport area (D'alfonso et al., 2015; Merkert and Mangia, 2014).

(a) Total runway length: Global hubs are designed to accommodate large aircraft, requiring sufficient runway length, which is influenced by factors like altitude, temperature, and surface wind. There are four basic runway configurations: single, parallel, intersecting, and open-V (Budd and Ison, 2017). Parallel runways can vary in spacing and number, including close or medium-spaced, independent, and multiple pairs (De Neufville et al., 2013). Commercial runways can handle 30 to 60 movements per hour, but intersecting runways reduce capacity. Therefore, large airports favor parallel runways for simultaneous takeoffs and landings (Rodrigue, 2020).

(b) Passenger terminals: Most airports have one of the following passenger terminal configurations: standard (linear or semicircular), piers, concourses, satellites, and shuttles (Rodrigue, 2020). The suitability of a configuration depends on traffic levels, seasonality, and the percentage of transfer traffic, as hub connections need to be fast, reliable, and easy to navigate. Transfer passengers prefer short walking distances and dislike changing terminals (De Neufville et al., 2013). Hub terminals differ from origin-destination terminals by focusing on optimizing passenger flow during transfers, accommodating large volumes of passengers moving between gates. Designed to handle waves of passengers from arriving and departing aircraft, these terminals often use mechanized aids like automated people movers for efficient circulation (Ashford et al., 2013). Since transfer passengers spend more time in these terminals waiting for their connections, non-aeronautical services, such as retailing and advertising, emerge as the main source of revenue for many hubs (Kidokoro and Zhang, 2023).

(c) Airport area: Airports need expansive areas for runways, terminals, maintenance hangars, and other facilities. This requirement often leads to airports being located further from city centers to cover their physical footprint (Rodrigue, 2020).

1.1.3. External aspects

Geographic and political factors significantly impact an airport's hub status. European airports serve as global connectors, though they face competition from Middle Eastern hubs due to strategic geographical positions and supportive aviation policies (O'Connell and Bueno, 2018). Geographic location also plays a crucial role in shaping the market served by hubs and the strategies adopted by their operators. For instance, while Southeast Asian hubs focus on competing on Asia-Europe/Middle East/Oceania routes, Far East Asia hubs concentrate

Table 1
Descriptive statistics and data sources

Variable	Min	Max	Mean	Standard Deviation	Sources
RWYLENGTH	1738	24507	6890,59	3846,06	http://www.gcmap.com/
TERMINALAREA	2	180	33,21	32,53	Own Calculations via Google Earth
GATES	0	195	40,99	39,43	Own Calculations via Google Earth
AIRPORTAREA	75	4230	878,52	741,47	Own Calculations via Google Earth
GDPPERCAPITA	2,27	204,61	44,53	27,18	World Bank (2023), OECD (2023)
POPULATION	0,054	37,73	6,04	6,99	Demographia (2023)
CITYGLOB	0	12	5,39	3,71	GaWC (2020)
CARRIERDEST	1	5	2,64	1,31	
CARRIERFLIGHTS	1	5	2,63	1,21	http://flightconnections.com/ based on
INTERCONTDEST	1	5	1,77	1,11	market figures from November 2023
CODESHARE	1	5	2,00	1,16	
GACI	1	3,18	1,56	0,49	Cheung et al. (2020)

their efforts on Asia-North America routes (Chang et al., 2020). Socioeconomic factors like regional population and GDP per capita positively affect air transport demand (Albayrak et al., 2020; Boonekamp et al., 2018), although the correlation with airport efficiency is debated (Örkcü et al., 2016; Tsui et al., 2014). Hub airports significantly boost economic development not only in their own cities but also in connected cities within the same airport network. Increases in air passengers, cargo, or flights at a hub have substantial positive spillover effects on the regional economy (Chen et al., 2021). Additionally, city globalization indices influence air freight and logistics networks (Akhavan et al., 2020; Matsumoto and Domae, 2018).

2. Methods

2.1. Data collection

Data from 300 main international airports worldwide were collected. The airports included in the study, listed in Appendix A, handled over 5 billion passengers in 2022 combined, representing slightly more than 75 % of global air traffic during this period, according to Airports Council International – ACI (2023). The database was categorized into internal, external, and airport status variables, the latter distinguishing between hub and non-hub airports. The significant challenge of the research was obtaining standardized data from a single source that was available for all 300 airports in the dataset, aiming to contribute to a consistent and discrepancy-free database. Frequently, data published by airport operators varies in measurement methods. For example, one airport may disclose the total number of gates, considering all types, while another might only report gates equipped with airbridges. Similarly, an airport may publicize its area, encompassing the entire airport site, while other mentions only the airside area. In order to prioritize data standardization, certain variables were derived through own calculations based on aerial view information collected via Google Earth. This approach particularly applied to terminal area, number of gates and airport area. Similarly, for the airport status variables regarding the main carrier's activity, since these may vary according to the month consulted and in order to facilitate subsequent definition of cutoff scores, they were categorized on a scale of 1 to 5, shown in Appendix B, with the interval ranges established by the authors based on exploratory analysis. The following Table 1 present the descriptive statistics of the variables included in the model, along with the source of their data:

2.1.1. Internal variables

- RWYLENGTH: total length of all runways in meters;
- TERMINALAREA: total area in hectares of passenger terminal buildings plus the apron area in hectares for parking spaces with airbridges;
- GATES: number of gates equipped with airbridges;
- AIRPORTAREA: total area of the airport site in hectares.

2.1.2. External variables

- GDPPERCAPITA: GDP per capita values expressed in thousands of current international dollars (USD) converted by Purchasing Power Parity (PPP) of the city or region where the airport is located based on values from the year 2019;
- POPULATION: estimated population for the year 2023 of the metropolitan area served by the airport, in millions of inhabitants;
- CITYGLOB: classification based on the GaWC 2020 (Globalization and World Cities Research Network); a think tank established by the Department of Geography at Loughborough University in the United Kingdom. This index categorizes global cities into tiers known as alpha, beta, gamma, and sufficient, along with their respective subcategories. GaWC assesses cities worldwide based on inter- and intrarelations of offices located in their territories, focusing on accounting, advertising, banking/finance, and law. Consequently, this variable is represented by an ordinal value ranging from 0 to 12, where 0 pertains to cities not listed by GaWC, and 12 represents cities in the Alpha group, highly integrated into the world economy, in the alpha++ subcategory, the highest within this group. In cases where airports are part of multi-airport systems (Bonnefoy et al., 2010), the primary city of the system was chosen as the reference. For example, the city index of Miami was used for Fort Lauderdale Airport, the city index of Milan was used for Bergamo Airport, and so on.

2.1.3. Airport status variables

- CARRIERDEST: number of destinations offered by the main carrier operating at the airport;
- CARRIERFLIGHTS: sum of the longest flights by the dominant airline departing from the airport in the northeast, northwest, southeast, and southwest directions to detect the centrality of this facility within the carrier's network and potential market orientations;
- INTERCONTDEST: number of intercontinental destinations (i.e., those surpassing the International Air Transport Association (IATA) Traffic Conference Zone in which the airport is situated. IATA divides the world into three zones, namely: Zone 1 – Americas, Zone 2 – Europe, Africa, and the Middle East, and Zone 3 – Asia-Pacific), offered by the primary airline;
- CODESHARE: number of codeshare agreements held by the main carrier operating at the airport with other carriers also operating at this airport, based on global alliances and information available on the website of the primary airline;
- GACI: Global Airport Connectivity Index (GACI) developed by Cheung et al. (2020), with scores corresponding to the year 2022.

2.2. Exploratory factor analysis

There are several methods to uncovering data structure, among which factor analysis is a statistical approach to examine the underlying structure in multivariate data and reduce the number of P variables to a smaller set of parsimonious K < P variables (Washington et al., 2011).

In order to condense all previously collected internal and external variables into a smaller group and thus facilitate the insertion of these independent and uncorrelated variables in the next stage of the tool, an exploratory factor analysis (EFA) was performed using the statistical software SPSS version 28, employing the Principal Component Analysis (PCA) method for factor extraction. EFA examines all the pairwise relationships between individual variables (e.g., items on a scale) and seeks to extract latent factors from the measured variables (Osborne, 2014).

Two tests were employed to measure the adequacy level of the data. The first refers to the Kaiser-Meyer-Olkin (KMO) which measures the suitability of the data for factor analysis. A KMO value greater than 0.6 indicates that the sample is acceptable. Therefore, the value found in this EFA, which was 0.684, reflects a reasonable adaptability of this dataset. The second is the Bartlett´s Test of Sphericity which provides a measure of the statistical probability that the correlation matrix has significant correlations among some of its components. The results were significant ($p < 0.001$), which indicates a good level of adequacy for factor analysis. All variables exhibited communalities greater than 0.5. A solution with three factors with eigenvalues greater than 1 representing 86.08 % of total variance was obtained.

2.3. Binary logistic regression

The goal of Binary Logistic Regression (BLR) is to identify a well-fitting model that describes the relationship between a binary dependent variable and a set of independent or explanatory variables (Washington et al., 2011). This approach models the probability of presence and absence of the binary dependent variable (Ozdemir, 2011). Some studies have already used BLR on topics related to air transport, such as: wave-system structures of network airlines at hub airports (Jiang et al., 2020), multiple airport systems status (Hansen and Weidner, 1995) and passengers' satisfaction with safety (Ceccato and Masci, 2017).

The BLR constitutes the main stage of the model in this study, wherein the airport status is represented by the binary dependent variable, where 1 signifies the airport being a hub and 0 non-hub. The three factors generated in the EFA serve as the independent variables. Through this tool, the aim is firstly to ascertain whether the factors are related to the airport status and in what manner, and subsequently, to determine which initial variables contained in these factors most influence this process. The general form of BLR is described in Eq.s (1-5) and more details can be found in Hosmer and Lemeshow (2000).

$$\mathrm{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \tag{1}$$

$$\pi = \mathrm{P}\left(\mathrm{Y} = 1 | x_1, x_2, \ldots, x_m\right) = \frac{e^Y}{1 + e^Y} \tag{2}$$

$$1 - \pi = \mathrm{P}\left(\mathrm{Y} = 0 | x_1, x_2, \ldots, x_m\right) \tag{3}$$

$$\text{odds} = \frac{\mathrm{P}(Y = 1)}{\mathrm{P}(Y = 0)} = \frac{\mathrm{P}(Y = 1)}{1 - \mathrm{P}(Y = 1)} = \frac{\pi}{1 - \pi} \tag{4}$$

$$\text{Logit P} = \ln\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \tag{5}$$

Where Y is the dependent variable (when Y=1 the airport status is defined as hub, otherwise Y=0 means non-hub status); β_0 is the constant; $\beta_1 + \beta_2 + \ldots + \beta_m$ are the regression coefficients to be estimated; and $x_1 + x_2 + \ldots + x_m$ are independent variables. Regression coefficients $\beta_1 + \beta_2 + \ldots + \beta_m$ show the contribution of each independent variable on probability value P. A positive sign means that the explanatory variable helps to increase the probability of change, and a negative sign implies the opposite effect (Ozdemir, 2011). Thus, Y is a linear combination function of the independent variables that configure a linear relationship. The probability of Y=1 is represented by π and consequently the probability of Y=0 is 1-π. The ratio of the first probability to the last are the odds. Logit P is the logarithmic conversion of the odds. This model is a multivariate estimation approach that evaluates the relative strength and significance of the factors.

2.3.1. Determination of the dependent variable

Since the classification of what constitutes a global hub lacks a precise definition in the literature, several approaches have been employed using airport status variables to get as close as possible to a binary variable that accurately separates the sample into hub and non-hub airports. The study opted to establish two types of criteria as predominant in determining an airport's status, the first being the airport connectivity and the second being the activity of the main airline operating at that facility. The first criterion is represented by the variable GACI (Cheung et al., 2020) that combines existing centrality indicators, including degree, closeness, and eigenvector, along with two new volume-based indicators, flow betweenness, and regional importance, to study the evolution of aviation networks and understand an airport's competitive position and its influence over others. The second criterion involves an association of the ordinal variables calculated in this study, namely: CARRIERDEST, CARRIERFLIGHTS, INTERCONTDEST, and CODESHARE. Hence this model defines the status of an airport based on its GACI score while also highlighting the behaviour of the main carrier operating at its facilities. Thus, airports that handle a large number of passengers and have a significant range of flights and destinations, but where the main airline is a regional or low-cost carrier with limited connection traffic volume, are not considered hubs.

Once the dependent variable is binary, it is necessary to decide cut-off or thresholds values to classify airports in the sample as hubs and non-hubs. Regarding the GACI, the authors of this index determined that a hub airport is one with a GACI score greater than or equal to 2 (Cheung et al., 2020). For the ordinal variables, as there are no pre-established theoretical values, since they were calculated in this paper, four separate linear regressions were conducted, using SPSS version 28, to find cut-off values correlated with the determined score of 2 by GACI. Therefore, the analyses established GACI as the independent variable and aspects of the main carrier´s activity, in their form of continuous variables, as the dependent variables in each respective linear regression. GACI significantly predict CARRIERDEST, CARRIERFLIGHTS, INTERCONTDEST, and CODESHARE, with ANOVA test results of $F\,(1{,}299) = 455.658$, $p < .001$, $F\,(1{,}299) = 340.501$, $p < .001$, $F\,(1{,}299) = 262.964$, p < .001 and $F\,(1{,}299) = 221.941$, p < .001, correspondingly. Moreover, the $R^2 = 0.605$ in the CARRIERDEST regression, the $R^2 = 0.533$ in the CARRIERFLIGHTS regression, the $R^2 = 0.469$ in the INTERCONTDEST regression and the $R^2 = 0.427$ in the CODESHARE regression depicts that the first model explains 60.5 % of variance, the second 53.3 %, the third 46.9 % and the fourth 42.7 %. The regression Eq.s for predicting CARRIERDEST, CARRIERFLIGHTS, INTERCONTDEST, and CODESHARE from GACI score is described in Eqs 6-9, respectively.

$$\mathrm{y} = -46.57 + 67.13x \tag{6}$$

$$\mathrm{y} = -7355.82 + 12532.87x \tag{7}$$

$$\mathrm{y} = -16.31 + 13.46x \tag{8}$$

$$\mathrm{y} = -8.58 + 8.11x \tag{9}$$

Thus, it is observed that when x = 2 in Eqs 6-9, y takes the value of 88 destinations for CARRIERDEST, 17710 nautical miles for CARRIERFLIGHTS, 11 intercontinental destinations for INTERCONTDEST, and 8 agreements for CODESHARE. Since there are four different aspects of the main airline in a hub, and this carrier may not meet all these requirements, the cut-off scores for a more lenient specification were set at the category immediately below the category where Y is when x = 2

Table 2
EFA results

	Factor 1	Factor 2	Factor 3
RWYLENGTH	**0.848**	-0.371	-0.082
TERMINALAREA	**0.895**	0.010	-0.228
GATES	**0.903**	-0.147	-0.059
AIRPORTAREA	**0.802**	-0.345	-0.159
GDPPERCAPITA	0.382	-0.146	**0.889**
POPULATION	**0.552**	**0.740**	-0.174
CITYGLOB	**0.675**	**0.536**	0.312

(Appendix B). That is, for CARRIERDEST, the lower threshold was category 3 (51 to 75 destinations), for CARRIERFLIGHTS it was category 3 (more than 10 to 15 thousand nautical miles), for INTERCONTDEST it was category 2 (1 to 5 intercontinental destinations), and for CODESHARE it was category 2 (1 to 5 agreements). Therefore, it can be stated that a hub airport in this study is one that has a GACI score greater than or equal to 2 but also has the main airline operating at its facilities, offering more than 50 destinations with at least 1 being intercontinental, with the sum of the longest flights in the northeast, northwest, southeast, and southwest directions being greater than 10000 nautical miles and having at least 1 code-share agreement with another airline that also operates at the airport. The classification established in this study was based on airport connectivity through its GACI value and the comparison of this score with the characteristics of the dominant airline. Some of the requirements might raise doubts, given that there are hub airports in strong domestic markets focused exclusively on national and nearby international flights, and there may be local full-service carriers without code-share agreements or with a limited range of destinations. However, this study focuses on identifying key factors of global hubs rather than regional hubs. Out of the 300 airports in the study database, 41 met these requirements and had their status recognized as hubs, while the remaining 259 were classified as non-hubs.

3. Results and discussion

3.1. Exploratory factor analysis outputs

Table 2 shows the factors extracted in the EFA with loading factor values greater than 0.50 in bold and underlined. Upon observing the EFA results, it is noted that internal aspects are inserted in Factor 1, while GDPPERCAPITA is contained in Factor 3. POPULATION and CITYGLOB appear simultaneously in Factor 1 and Factor 2. The hypothesis considered by the authors was that Factor 1 relates to the intrinsic characteristics of airports as well as characteristics included in POPULATION and CITYGLOB that contribute to greater airport infrastructure and capacity. In Factor 2, there are characteristics of POPULATION and CITYGLOB that are not associated with internal aspects or even infer lesser value in these aspects. This can be explained by densely populated regions with low economic indicators where a large population does not translate into a large number of passengers or in cases where secondary or small airports are located in major centers. Finally, Factor 3 pertains to GDPPERCAPITA, which is not necessarily related to greater airport infrastructure and capacity; a city with good economic indicators may have a population that does not justify a large airport. When analyzing Factor 1 in-depth, it is evident that aspects with higher loading factor are, in descending order of importance, the number of gates equipped with airbridges, the passenger terminal area + apron area, the total length of all runways and the total airport area; all belonging to internal variables. A pattern can be identified where characteristics related to the passenger terminal have slightly greater importance than characteristics of airport airside. Among external variables, the globalization index of the airport city exhibits greater influence than population. Conversely, in Factor 2, population displays a higher loading factor.

Table 3
Classification table

Observed	Predicted		
	Airport Status		Percentage Correct
	0 (non-hub)	1 (hub)	
0 (non-hub)	252	7	97.3
1 (hub)	13	28	68.3
Overall Percentage			93.3

3.2. Binary logistic regression outputs

A BLR was conducted using SPSS version 28, with the independent variables established as the three factors resulting from the EFA, and the binary dependent variable being calculated by the requirements defined in section 3.3.1. Succeeding a preliminary analysis, Factor 2 did not prove to be significant, which was somewhat expected since the hypothesis presented by the authors is that this factor encompasses demographic and attractiveness characteristics that extend beyond air transport. Therefore, Factor 2 was removed from the analysis, and it proceeded with Factor 1 and 3 as the independent variables. The Omnibus Test of Model Coefficients was significant, indicating that there is a considerable improvement in fit compared to the null model, hence, the model is showing a good fit. Hosmer and Lemeshow Test statistic indicates a poor fit if the significance value is less than 0.05. However, the model presented a significance of .629, adequately fitting the data. Therefore, the values between the observed and predicted model are approximately equal. The model exhibited a Nagelkerke R^2 value of .653. In this case, it can be said that 65.3 % change in the dependent variable can be accounted to the independent variables in the model.

According to Table 3, using a cut-off probability of 0.5, the predicted accuracy is 97.3 % for non-hubs and 68.3 % for hubs. The overall predicted accuracy is 93.3 % which reveals that this model has a good fitness for uncovering the key aspects that enhance the hub status of an airport. Among the 300 considered airports, only 20 had predicted model outcomes different from the observed ones. Of these, 7 airports were previously classified as non-hubs according to the hub definition in this study, but the BLR categorized them as hubs, while the other 13 had been labeled as hubs in the observed model and were designated as non-hub airports by the BLR.

Table 4 lists the variables in the logistic regression Eq. and their coefficients (B). The statistical test used the Wald chi-square value [$(B/\text{Standard Error})^2$] and a 95 % confidence interval for the corresponding degrees of freedom (df). The values of the constant, Factor 1 and Factor 3 are shown in Table 4. Factor 1 presented significance < .001 and Factor 3 exhibited significance value below 0.05. Hence, both factors were accepted in the regression as influential predictor variables with coefficients of 2.577 and 0.483, respectively. The Exp(B) value greater than 1 for Factor 1 and Factor 3 reveals that these variables have positive effects on determining an airport's status. The negative value of the constant (-3.537) indicates that airports with minimal infrastructure and low socioeconomic values in their region tend to have near-zero probabilities of being designated as hubs.

Based on the coefficient values provided in Table 4 and the loading factors from the EFA present in Table 2, Eqs (10-13) of logistic regression were derived as follows to represent the probability of an airport achieving hub status.

$$Y = -3.537 + 2.577(\text{FACTOR 1}) + 0.483(\text{FACTOR 3}) \tag{10}$$

$$\begin{aligned} Y = {} & -3.537 + 2.298\,x\text{GATES} + 2.196\,x\text{TERMINALAREA} \\ & + 2.145\,x\text{RWYLENGHT} + 1.990\,x\text{AIRPORTAREA} \\ & + 1.890\,x\text{CITYGLOB} + 1.413\,x\text{GDPPERCAPITA} \\ & + 1.338\,x\text{POPULATION} \end{aligned} \tag{11}$$

Table 4
Variables in the Eq.

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0 % C.I. for Exp(B)	
							Lower	Upper
Factor 1	2.577	0.359	51.426	1	<.001	13.161	6.507	26.620
Factor 3	0.483	0.224	4.644	1	.031	1.621	1.045	2.516
Constant	-3.537	0.435	66.049	1	<.001	0.029		
	Coefficient	Standard Error	Wald chi-square	df	Significance	Exp. Coefficient	95 % confidence interval for Exp(B)	

Table 5
List of Cases with SResid greater than 1

Airport	Observed	Pre-dicted	Probability	SResid	Carrier Dest. ≥ cat 3	Carrier Flights ≥ cat 3	Inter-Cont. Dest. ≥ cat 2	Code Share ≥ cat 2	GACI ≥ 2
PHOENIX (PHX)	1**	0	0,115	2,090	√	√	√	√	√
MIAMI (MIA)	1**	0	0,488	1,211	√	√	√	√	√
MINNEAPOLIS (MSP)	1**	0	0,333	1,494	√	√	√	√	√
VANCOUVER (YVR)	1**	0	0,150	1,957	√	√	√	√	√
MONTREAL (YUL)	1**	0	0,064	2,351	√	√	√	√	√
MUNICH (MUC)	1**	0	0,288	1,590	√	√	√	√	√
ZURICH (ZRH)	1**	0	0,126	2,051	√	√	√	√	√
VIENNA (VIE)	1**	0	0,056	2,411	√	√	√	√	√
BRUSSELS (BRU)	1**	0	0,109	2,117	√	√	√	√	√
SAO PAULO (GRU)	1**	0	0,084	2,239	√	√	√	√	√
DOHA (DOH)	1**	0	0,430	1,321	√	√	√	√	√
SYDNEY (SYD)	1**	0	0,127	2,041	√	√	√	√	√
MANILA (MNL)	1**	0	0,052	2,441	√	√	√	√	√
DETROIT (DTW)	0**	1	0,625	-1,422	√	√	√	√	
WASHINGTON (DCA)	0**	1	0,613	-1,693	√				
BEIJING (PKX)	0**	1	0,742	-1,681	√	√	√	√	
TOKYO (NRT)	0**	1	0,799	-1,817		√	√	√	√
HONG KONG (HKG)	0**	1	0,590	-1,352	√	√	√	√	
KUALA LUMPUR (KUL)	0**	1	0,854	-1,989	√	√	√	√	
JAKARTA (CGK)	0**	1	0,835	-1,932		√	√	√	√
SAN FRANCISCO (SFO)	1	1	0,518	1,163	√	√	√	√	√
BOSTON (BOS)	1	1	0,512	1,178	√	√	√	√	√
TORONTO (YYZ)	1	1	0,583	1,052	√	√	√	√	√
FRANKFURT (FRA)	1	1	0,558	1,093	√	√	√	√	√
NEW DELHI (DEL)	1	1	0,557	1,113	√	√	√	√	√
MOSCOW (SVO)	0	0	0,465	-1,130	√	√	√		√
ABU DHABI (AUH)	0	0	0,444	-1,094	√	√	√	√	

** Misclassified cases

$$Y = -20.586 + 0.058\ \mathrm{GATES} + 0.067\ \mathrm{TERMINALAREA} + 0.0006\ \mathrm{RWYLENGHT} + 0.003\ \mathrm{AIRPORTAREA} + 0.509\ \mathrm{CITYGLOB} + 0.052\ \mathrm{GDPPERCAPITA} + 0.191\ \mathrm{POPULATION} \tag{12}$$

$$P = \frac{e^Y}{1 + e^Y} \tag{13}$$

Where Y is the regression Eq., x is the standardized value associated with each of the variables and P is the possibility of an airport being a hub. Eq. 10 refers to the logistic regression. Eq. 11 demonstrates, in descending order of importance, the variables that influence the Y value, ranging from GATES to POPULATION. Eq. 12 presents the simplified Y value considering the variables in their original measurement units presented in the study. Thus, the higher the Factor 1 and the Factor 3, the higher the value of Y, and consequently, the probability P (Eq. 13) of an airport being classified as a hub increases.

3.3. Investigation of airports misclassified by the model

Upon analyzing the results obtained from the BLR, it was found that 20 out of the 300 sampled airports were misclassified by the model. Consequently, these outlier cases were investigated in order to identify any necessary corrections to the tool and to find similar characteristics that may explain the behavior of these airports in the model. Table 5 displays the cases with studentized residuals (SResid) greater than 1; the farther the airport is from its previously observed value, the larger the magnitude of its SResid. It is noteworthy that the last airports on the list were classified correctly in the model; however, they have SResid slightly greater than 1 because the values of their probabilities are very close to the cutoff score of 0.5. This suggests that small changes in the variables could lead to an incorrect classification of these airports. It is important to note that since the GACI score pertains to the year 2022 and the variables of the dominant airline reflect market figures from the end of 2023, there may still be residual effects from the COVID-19 recovery impacting the pre-selected airports as hubs based on the adopted cut-off values, and consequently, the model's classification. The cases with the highest SResid were those which, according to research requirements, should have been predicted as hubs in the model, but had probabilities close to 0, such as PHX, YUL, ZRH, VIE, BRU, GRU, SYD, and MNL. Airports observed as 1 and predicted as 0 are predominantly in North America and Europe, whereas airports observed as 0 and predicted as 1 are mostly in Asia.

Table 5 also indicates whether misclassified airports failed to meet the five requirements of the study and were therefore previously designated as non-hubs. Examining cases where the hub definition determined by the study did not recognize airports considered hubs by the BLR (0->1), it is noted that DTW, HKG, KUL, and PKX were excluded from hub status for not having a GACI score greater than or equal to 2, despite meeting all the requirements of the main carrier. Upon analyzing each case separately, it is observed that the first three airports

Table 6
Comparison of the mean values of hub airports with the values observed in the listed airports

Hub = 1	RWY LENGTH	TERMINAL AREA	GATES	AIRPORT AREA	GDP PERCAPITA	POPULATION	CITY GLOB
Mean	12726	84	108	1796	64,34	10,84	9
Min	5995	29	36	525	20,55	0,93	5
Max	24507	180	195	4230	204,61	37,73	12
1 -> 0							
PHX	9018	50	106	845	50,83	4,58	5
MIA	12643	80	**131**	1045	51,98	6,06	8
MSP	11340	50	**131**	1085	**67,92**	2,96	6
YVR	8548	55	93	1060	49,11	2,46	8
YUL	8413	32	54	900	43,78	3,71	**9**
MUC	8000	54	87	1715	**67,47**	2,04	**9**
ZRH	9500	32	53	890	**72,69**	0,93	**9**
VIE	7100	29	36	915	59,71	1,89	**9**
BRU	9836	44	53	600	55,77	2,20	**10**
GRU	6700	43	54	815	22,43	**23,09**	**10**
DOH	9100	70	47	**2200**	**94,69**	1,89	8
SYD	8930	50	59	770	52,54	4,71	**10**
MNL	5995	42	44	525	22,98	**24,92**	**9**
0 -> 1							
DTW	**17591**	65	**129**	**1860**	53,21	4,22	6
DCA	5295	21	60	275	**204,61**	7,63	8
PKX	**14800**	**95**	82	**2100**	38,46	**18,52**	**11**
NRT	6500	**104**	82	1315	**79,24**	**37,73**	**11**
HKG	11400	**91**	77	1700	62,12	7,45	**11**
KUL	11979	**111**	**130**	**1970**	**69,94**	8,91	**10**
CGK	10260	**112**	94	**1860**	56,40	**33,76**	**10**
1 -> 1							
SFO	12056	65	**115**	800	**77,26**	6,33	**9**
BOS	**12957**	55	102	675	**86,07**	7,44	**9**
YYZ	**15226**	65	104	1615	49,73	6,77	**10**
FRA	**14800**	75	64	**1950**	**64,67**	2,00	**10**
DEL	**15459**	60	79	1600	27,35	**32,23**	**9**
0 -> 0							
SVO	10450	66	84	1125	62,63	**17,33**	**10**
AUH	8200	**108**	82	**1850**	**74,83**	1,37	7

have GACI values very close to 2, and consequently, the cut-off score established ended up restricting these airports from having a value of 1 in the binary variable, leading to misclassification by the model. As for the last one (PKX), it was recently inaugurated with abundant infrastructure leading to a hub prediction by the model, however, it may not have yet achieved the expected international connectivity since Beijing has the older airport (PEK) performing this role of global hub with a GACI greater than 2.

In NRT and CGK, the main carrier operating at these facilities offers 50 or fewer destinations, and therefore, despite meeting all the other requirements, they were removed from the hub group before running the model. DCA is a case where external factors likely classified it as a hub in the regression, but it is a secondary airport in Washington, performing only a possible role as a national hub, while the other airport in the city (IAD) serves as the global connector of the American capital with a GACI greater than 2. As for the airports that were observed and predicted as non-hubs but are close to the cut-off score (0->0), SVO and AUH, it is noted that the former did not meet the code-share agreement requirement due to sanctions on Russian airlines because of the Ukraine War at the time of the analysis, while the latter does not have global connectivity to be recognized as a hub according to the GACI.

On the other hand, to understand why airports previously classified as hubs were not considered as such by the model, it is necessary to refer to Table 6. This table displays the mean, minimum, and maximum values of each variable for the group of airports with a score of 1 in the binary dependent variable and compares them with the observed values for each airport with SResid greater than 1. The values in bold and underlined refer to those that are greater than or equal to the mean of the hub group. When observing Table 6, it is noted that the airports previously observed as hubs and predicted as non-hubs (1->0) have values below the average of hubs in most internal variables. This trend may indicate that, in these cases, there is a shortage of infrastructure found in these facilities compared to their global connectivity and the activities of the main carrier. In some of these cases, the low values in external variables also contributed to a classification with a score of 0. In cases correctly classified as hubs (1->1), it is observed that most of these airports have total runway length greater than the mean of hubs and values for other infrastructure aspects similar to those found at airports predicted as 1, indicating that internal variables were able to sustain a probability above 0.5 for these airports. External aspects such as GDP per capita and Globalization Index in these cases also present substantial values to enhance their prediction of being a hub.

Conversely, the cases previously established with a score of 0 in this paper and labelled as hubs by the model (0->1) present values greater or close to the averages of hubs in most internal variables, with emphasis on terminal area and airport area. A pattern that may be indicated in these cases is that these facilities currently exhibit underutilization of their infrastructure in relation to their global connectivity or the operations of the main airline. Most of the airports listed as 0->1 also achieved good results in external variables. It can be noted that there is a tendency for these facilities to be located in cities considered alpha by the GaWC with a large population and/or a high GDP per capita. These airports, in turn, have potential external forces that may leverage these numbers in the future. Lastly, in cases correctly identified as non-hubs (0->0), SVO presents external variables values comparable to those of a hub airport, however, none of its internal variables are higher than the average of hubs. Although AUH has a considerable terminal area and total area, compared to those found in cases (0->1), the population and globalization index of its city do not correspond to the hub status. If the former improves its infrastructure and the latter boosts the global attractiveness of its city, these airports are strong candidates to be predicted as global hubs by the model in the future.

4. Conclusions

Understanding the key factors that enable an airport to achieve global hub status is crucial for both airlines and airport operators. Airlines need this information to select the most efficient secondary hub to handle excess demand when the primary hub is congested or during network expansion/reorganization for new hubs. Airport operators, on the other hand, can focus their investments and improvements on these critical elements to attract potential carriers to establish a hub at their facilities. Additionally, they can consider these factors when acquiring new facilities for management. This paper aimed to formulate a tool, with the assistance of EFA and BLR, that, based on characteristics present in airport infrastructure and external characteristics of the region where the airport is located, would be able to identify which airports would become hubs and which ones would not. From these results, it was possible to identify which aspects most influence this process, serving as a basis for future studies in the field of air transport hubs. The main limitation of this study was the lack of a reliable database with standardized information on airport infrastructure on a global scale. To address this, a detailed evaluation of each of the 300 sampled airports was conducted using satellite images to determine the passenger terminal and apron area, the number of gates equipped with jet bridges, and the total area of the airport site. For those airport operators that publish such information, the values were cross-checked and generally matched those found by the authors. However, this represented a very small portion of the entire database.

The EFA extracted three factors, of which the first primarily relates to internal airport variables and, to a lesser extent, external variables linked to the airport city´s globalization index and its population. The second factor is also associated with these external variables present in the first factor. Lastly, the third factor only contains significant loading factors of GDP per capita. Conducting the BLR, it was found that the first and third factor showed significance in the analysis, and from this, the Eq. predicting the probability of a certain airport achieving hub status was formulated. Consequently, upon analyzing the model Eq., it becomes evident that internal airport aspects have greater influence on this probability than external aspects. Among the aspects that most influence the definition of an airport's status, those related to the passenger terminal (GATES and TERMINALAREA) rank first. These structures must be designed to optimize the flow of connecting passengers and at the same time accommodate passengers originating from or destined to the airport. While airlines should focus on the terminal's passenger handling capacity, aircraft allocation at gates, and the current utilization of this capacity when choosing a hub, airport operators should concentrate on planning terminal expansions and implementing improvements that enhance hub airlines' utilization of their facilities. Following closely, with slightly less importance, are aspects related to the airside (RUNWAY LENGTH and AIRPORT AREA). Airports with larger areas have the advantage of being able to build more runways and have an airside layout that enhances aircraft operations, such as independent parallel runways. Larger areas also mean greater opportunities for future expansion of facilities, as well as the construction of new maintenance hangars and buildings for the dominant airline's operations. Among the external aspects that most affect the establishment of a global hub at a particular airport, CITYGLOB stand out. Thus, it is not enough for an airport to have quality infrastructure if it is not located in a globally interconnected city with economic attractiveness. Finally, GDPPERCAPITA and POPULATION are the variables in this paper that exhibit the least influence in this airport status selection process.

The definition of a global hub in this study focused on two premises. The first was that an airport with this status has extensive global connectivity in its network of flights and destinations, thereby inhibiting potential misclassification by the model due to small or regional airports predominantly served by only one airline. Such airline may not necessarily use this airport as a hub for passenger distribution between countries or continents. The second premise is that a global hub airport is one where its main carrier offers a wide range of destinations, especially intercontinental ones, provided from this facility and also through code-share agreements with other partner airlines. Thus, the analysis sought to avoid errors in identifying hubs in cases of large global airports that mainly serve as bases for low-cost airlines, without the massive presence of a full-service carrier with a hub-and-spoke network, and, therefore, do not enjoy from a considerable share of connecting passengers. Hence, the GACI score was used to assess the global connectivity of airports, combined with restrictive factors related to the main carrier´s activity at each airport in the sample. This ensured that only airports with a GACI value equal or greater than 2, and with strongly established full-service carrier within their facilities, were designated as hubs in advance, before performing the tool. Future studies could further explore the definition of the term "global hub" to standardize research in aviation literature. This includes establishing the precise threshold values for airport connectivity, the market share of the dominant airline, and the percentage of connecting passengers required to distinguish hub airports from non-hubs.

The tool proposed by this paper achieved an overall predicted accuracy of 93.3 % in identifying the status of airports in the sample. In the case of the group classified as non-hubs, the accuracy reached 97.3 %, while in the group identified as hubs, it was 68.3 %. After investigating the airports misclassified by the model, it was found that the regression proved to be vulnerable in cases where the hub airport exhibits obvious scarcity or under exploitation of its infrastructure. Hence, it can be inferred that these airports are hubs when, in reality, they should not have this status given the lack of infrastructure, or they are non-hubs when internal and external aspects indicate that they should be, due to their abundant infrasctruture. Future research could examine the relationship between the over- or underutilization of infrastructure at specific hubs and their slot levels and market concentration, as portrayed in Milioti and Odoni (2024). Likewise, other aspects beyond those analyzed in the study may possibly be responsible for an airport achieving or not achieving hub status, such as political factors, lack of competition or demand in the region, multiple airport systems, market choices by airlines, among others. Subsequent studies may deepen this analysis and introduce these new aspects with the aim of reducing the model's susceptibility to the percentage of airport infrastructure utilization. Similarly, aspects related to the number of connecting passengers at each airport can be calculated and incorporated into the binary dependent variable, in order to better explain the behavior of hubs and their common characteristics that enhance this status.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Gabriel Bordeaux: Writing – original draft, Methodology, Investigation, Conceptualization. **António Couto:** Writing – review & editing, Methodology.

Acknowledgements

The authors were funded through the programmatic funding - UIDP/04427/2020 and the research grant UI/BD/153356/2022 awarded by the Fundação para a Ciência e a Tecnologia (FCT) of Portugal to the Research Centre for Territory, Transports and Environment

(CITTA). The authors would like to thank researchers Tommy Cheung, Collin Wong and Anming Zhang for providing the GACI data regarding the 2022 scores of the sample airports.

Appendix A. Database airports

Database Airports
North America, Latin America and the Caribbean Albuquerque (ABQ), Anchorage (ANC), Atlanta (ATL), Austin (AUS), Baltimore (BWI), Belo Horizonte (CNF), Bogota (BOG), Boise (BOI), Boston (BOS), Brasilia (BSB), Buenos Aires (AEP), Buenos Aires (EZE), Burbank (BUR), Cancun (CUN), Calgary (YYC), Charleston (CHS), Charlotte (CLT), Chicago (MDW), Chicago (ORD), Cincinnati (CVG), Cleveland (CLE), Columbus (CMH), Curitiba (CWB), Dallas (DAL), Dallas/Fort Worth (DFW), Denver (DEN), Detroit (DTW), Edmonton (YEG), Fort Lauderdale (FLL), Fort Myers (RSW), Fortaleza (FOR), Guadalajara (GDL), Hartford (BDL), Honolulu (HNL), Houston (HOU), Houston (IAH), Indianapolis (IND), Jacksonville (JAX), Kahului (OGG), Kansas City (MCI), Las Vegas (LAS), Lima (LIM), Los Angeles (LAX), Medellin (MDE), Memphis (MEM), Mexico City (MEX), Miami (MIA), Milwaukee (MKE), Minneapolis/St Paul (MSP), Monterrey (MTY), Montreal (YUL), Nashville (BNA), New Orleans (MSY), New York (JFK), New York (LGA), Newark (EWR), Oakland (OAK), Omaha (OMA), Ontario (ONT), Orlando (MCO), Palm Beach (PBI), Panama City (PTY), Philadelphia (PHL), Phoenix (PHX), Pittsburgh (PIT), Portland (PDX), Porto Alegre (POA), Punta Cana (PUJ), Raleigh/Durham (RDU), Recife (REC), Rio De Janeiro (GIG), Rio De Janeiro (SDU), Sacramento (SMF), Salt Lake City (SLC), Salvador (SSA), San Antonio (SAT), San Diego (SAN), San Francisco (SFO), San Jose (SJC), San Juan (SJU), Santa Ana (SNA), Santiago (SCL), Sao Paulo (CGH), Sao Paulo (GRU), Sao Paulo/Campinas (VCP), Seattle/Tacoma (SEA), St Louis (STL), Tampa (TPA), Tijuana (TIJ), Toronto (YYZ), Vancouver (YVR), Washington (DCA), Washington (IAD)
Europe, the Middle East and Africa Abu Dhabi (AUH), Abuja (ABV), Addis Ababa (ADD), Alicante (ALC), Algiers (ALG), Amman (AMM), Amsterdam (AMS), Ankara (ESB), Antalya (AYT), Athens (ATH), Bahrain (BAH), Barcelona (BCN), Basel/Mulhouse (BSL), Belfast (BFS), Belgrade (BEG), Bergen (BGO), Berlin (BER), Bilbao (BIO), Birmingham (BHX), Bari (BRI), Bologna (BLQ), Bordeaux (BOD), Bristol (BRS), Brussels (BRU), Brussels/Charleroi (CRL), Budapest (BUD), Bucharest (OTP), Cairo (CAI), Cagliari (CAG), Cape Town (CPT), Casablanca (CMN), Catania (CTA), Cologne (CGN), Copenhagen (CPH), Cyprus/Larnaca (LCA), Dammam (DMM), Dubai (DXB), Dubai/Sharjah (SHJ), Dublin (DUB), Dusseldorf (DUS), Durban (DUR), Edinburgh (EDI), Eindhoven (EIN), Faro (FAO), Frankfurt (FRA), Geneva (GVA), Gdansk (GDN), Glasgow (GLA), Gothenburg (GOT), Gran Canaria (LPA), Hamburg (HAM), Helsinki (HEL), Heraklion (HER), Hurghada (HRG), Ibiza (IBZ), Istanbul (IST), Istanbul (SAW), Izmir (ADB), Jeddah (JED), Johannesburg (JNB), Katowice (KTW), Kuwait (KWI), Krakow (KRK), Lagos (LOS), Lisbon (LIS), London (LGW), London (LHR), London/Luton (LTN), London/Stansted (STN), Luxembourg (LUX), Lyon (LYS), Madrid (MAD), Malaga (AGP), Malta (MLA), Manchester (MAN), Marrakesh (RAK), Marseille (MRS), Milan (LIN), Milan (MXP), Milan/Bergamo (BGY), Munich (MUC), Moscow (DME), Moscow (SVO), Moscow (VKO), Muscat (MCT), Nairobi (NBO), Nantes (NTE), Naples (NAP), Nice (NCE), Oslo (OSL), Palermo (PMO), Palma de Mallorca (PMI), Paris (CDG), Paris (ORY), Paris/Beauvais (BVA), Pisa (PSA), Porto (OPO), Prague (PRG), Reykjavik (KEF), Rhodes (RHO), Riga (RIX), Riyadh (RUH), Rome (FCO), Saint Petersburg (LED), Salt Lake City (SLC), San Diego (SAN), San Francisco (SFO), San Jose (SJC), Santa Ana (SNA), Santiago (SCL), Sao Paulo (CGH), Sao Paulo (GRU), Sao Paulo/Campinas (VCP), Seattle/Tacoma (SEA), Sevilla (SVQ), Sofia (SOF), Stockholm (ARN), Stuttgart (STR), Tel Aviv (TLV), Tenerife (TFS), Thessaloniki (SKG), Tirana (TIA), Toulouse (TLS), Tunis (TUN), Valencia (VLC), Vancouver (YVR), Venice (VCE), Vienna (VIE), Warsaw (WAW), Zurich (ZRH)
South/Southeast/Far East Asia and Oceania Adelaide (ADL), Ahmedabad (AMD), Auckland (AKL), Bangalore (BLR), Bangkok (BKK), Bangkok (DMK), Beijing (PEK), Beijing (PKX), Brisbane (BNE), Busan (PUS), Cebu (CEB), Changchun (CGQ), Changsha (CSX), Chengdu (CTU), Chengdu (TFU), Chennai (MAA), Chiang Mai (CNX), Christchurch (CHC), Chongqing (CKG), Kochi (COK), Dalian (DLC), Da Nang (DAD), Denpasar (DPS), Fukuoka (FUK), Fuzhou (FOC), Gold Coast (OOL), Goa (GOI), Guangzhou (CAN), Guiyang (KWE), Haikou (HAK), Hanoi (HAN), Hangzhou (HGH), Harbin (HRB), Hefei (HFE), Ho Chi Minh City (SGN), Hohhot (HET), Hong Kong (HKG), Hyderabad (HYD), Jakarta (CGK), Jeju (CJU), Jinan (TNA), Kolkata (CCU), Kuala Lumpur (KUL), Kunming (KMG), Lanzhou (LHW), Manila (MNL), Makassar (UPG), Melbourne (MEL), Medan (KNO), Mumbai (BOM), Nagoya (NGO), Nanjing (NKG), Nanchang (KHN), Nanning (NNG), New Delhi (DEL), Ningbo (NGB), Osaka (ITM), Osaka (KIX), Okinawa (OKA), Perth (PER), Phuket (HKT), Pune (PNQ), Qingdao (TAO), Sanya (SYX), Seoul (GMP), Seoul (ICN), Shanghai (PVG), Shanghai (SHA), Shenyang (SHE), Shenzhen (SZX), Shijiazhuang (SJW), Singapore (SIN), Sapporo (CTS), Sydney (SYD), Surabaya (SUB), Taipei (TPE), Taiyuan/Jinzhong (TYN), Tianjin (TSN), Tokyo (HND), Tokyo (NRT), Urumqi (URC), Wenzhou (WNZ), Wuhan (WUH), Xiamen (XMN), Xi'an/Xianyang (XIY), Zhengzhou (CGO)

Appendix B. Categories of Airport Status Variables

Variable	Units of Measurement	Categories				
		1	2	3	4	5
CARRIERDEST	destinations	0-25	26-50	51-75	76-100	100+
CARRIERFLIGHTS	nautical miles (thousand)	0-5	6-10	11-15	16-25	25+
INTERCONTDEST	destinations	0	1-5	6-15	16-30	30+
CODESHARE	agreements	0	1-5	6-10	11-20	20+

References

ACI – Airports Council International (2023). Global passenger traffic expected to recover by 2024 and reach 9.4 billion passengers. Advisory Bulletins. URL: https://aci.aero/2023/09/27/global-passenger-traffic-expected-to-recover-by-2024-and-reach-9-4-billion-passengers/

Akhavan, M., Ghiara, H., Mariotti, I., & Sillig, C. (2020). Logistics global network connectivity and its determinants. A European City network analysis. *Journal of Transport Geography, 82*. 10.1016/j.jtrangeo.2019.102624.

Albayrak, M. B. K., Özcan, I. Ç., Can, R., & Dobruszkes, F. (2020). The determinants of air passenger traffic at turkish airports. *Journal of Air Transport Management, 86*. 10.1016/j.jairtraman.2020.101818.

Alumur, S., & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research, 190*(1), 1–21. 10.1016/j.ejor.2007.06.008.

Ashford, N., Stanton, H. P., Moore, C., Coutu, P., & Beasley, J. R. (2013). *Airport Operations* (3rd Ed). New York: McGraw-Hill.

Bilotkach, V., & Bush, H. (2020). Airport competition from airports' perspective: evidence from a survey of European airports. *Competition and Regulation in Network Industries, 21*(3), 275–296. 10.1177/1783591720937876.

Bonnefoy, P. A., De Neufville, R., & Hansman, R. J. (2010). Evolution and development of multiairport systems: Worldwide perspective. *Journal of transportation engineering, 136*(11), 1021–1029. 10.1061/(ASCE)0733-947X(2010)136:11(1021).

Boonekamp, T., Zuidberg, J., & Burghouwt, G. (2018). Determinants of air travel demand: the role of low-cost carriers, ethnic links and aviation-dependent employment. *Transportation Research Part A Policy and Practice, 112*, 18–28. 10.1016/j.tra.2018.01.004.

Budd, L., & Ison, S. (2017). *Air Transport Management: An International Perspective*. London, New York: Routledge.

Burghouwt, G., & Hakfoort, J. (2001). The evolution of the European aviation network, 1990–1998. *Journal of Air Transport. Management, 7*, 311–318. 10.1016/S0969-6997(01)00024-2.

Button, K. (2002). Debunking some common myths about airport hubs. *Journal of Air Transport Management, 8*, 177–188. 10.1016/S0969-6997(02)00002-9.

Button, K., & Lall, S. (1999). The economics of being an airline hub city. *Research in Transport Economics, 5*, 75–106. 10.1016/S0739-8859(99)80005-5.

Ceccato, V., & Masci, S. (2017). Airport environment and passengers' satisfaction with safety. *Journal of Applied Security Research, 12*, 356–373. 10.1080/19361610.2017.1315696.

Chang, Y. C., Lee, W. H., & Hsu, C. J. (2020). Identifying competitive position for ten Asian aviation hubs. *Transport Policy, 87*, 51–66. 10.1016/j.tranpol.2020.01.003.

Chaouk, M., Pagliari, R., & Moxon, R. (2020). The impact of national macro-environment exogenous variables on airport efficiency. *Journal of Air Transport Management, 82*. 10.1016/j.jairtraman.2019.101740.

Chen, X., Xuan, C., & Qiu, R. (2021). Understanding spatial spillover effects of airports on economic development: New evidence from China's hub airports. *Transportation Research Part A: Policy and Practice, 143*, 48–60. 10.1016/j.tra.2020.11.013.

Cheung, T. K., Wong, C. W., & Zhang, A. (2020). The evolution of aviation network: global airport connectivity index 2006–2016. *Transportation Research Part E Logistics and Transportation Review, 133*. 10.1016/j.tre.2019.101826.

Cook, G. N., & Goodwin, J. (2008). Airline networks: a comparison of hub-and-spoke and point-to-point systems. *Journal of Aviation/Aerospace Education and Research, 17*(2). 10.15394/jaaer.2008.1443.

Costa, T. F. G., Lohmann, G., & Oliveira, A. V. M. (2010). A model to identify airport hubs and their importance to tourism in Brazil. *Research in Transportation Economics, 26*, 3–11. 10.1016/j.retrec.2009.10.002.

D'Alfonso, T., Daraio, C., & Nastasi, A. (2015). Competition and efficiency in the Italian airport system: New insights from a conditional nonparametric frontier analysis. *Transportation Research Part E Logistics and Transportation Review, 80*, 20–38. 10.1016/j.tre.2015.05.003.

Demographia. Demographia world urban areas: Built up urban areas or world agglomerations. 19th Ed. 2023. URL: http://www.demographia.com/db-worldua.pdf.

De Neufville, R., Odoni, A. R., Belobaba, P. P., & Reynolds, T. G. (2013). *Airport Systems, Second Edition: Planning, Design and Management* (2nd Ed). New York: McGraw-Hill.

Fragoudaki, A., Giokas, D., & Glyptou, K. (2016). Efficiency and productivity changes in Greek airports during the crisis years 2010–2014. *Journal of Air Transport Management, 57*, 306–315. 10.1016/j.jairtraman.2016.09.003.

Globalization and World Cities Research Network. GaWC. The World According to GaWC – Globalization and World Cities. 2020. URL: https://www.lboro.ac.uk/microsites/geography/gawc/world2020t.html.

Hansen, M., & Weidner, T. (1995). Multiple airport systems in the United States: current status and future prospects. *Transportation Research Record, 1506*, 8–17.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd Ed). New York: John Wiley and Sons Inc.

Ivy, R. L. (1993). Variations in hub service in the US domestic air transportation network. *Journal of Transport Geography, 1*, 211–218. 10.1016/0966-6923(93)90045-2.

Jiang, Y., Lu, J., Feng, T., & Yang, Z. (2020). Determinants of wave system structures of network airlines at hub airports. *Journal of Air Transport Management, 88*, Article 101871. 10.1016/j.jairtraman.2020.101871.

Kidokoro, Y., & Zhang, A. (2023). Effects of non-aeronautical service on airports: A selected review and research agenda. *Journal of the Air Transport Research Society, 1*(1), 40–53. 10.59521/5372A44CC8A85136.

Maertens, S., Grimme, W., & Bingemer, S. (2020). The development of transfer passenger volumes and shares at airport and world region levels. *Transportation Research Procedia, 51*, 171–178. 10.1016/j.trpro.2020.11.019.

Martín, J. C., & Román, C. (2003). New potential hubs in the South-Atlantic market. A problem of location. *Journal of Transport Geography, 11*(2), 139–149. 10.1016/S0966-6923(02)00062-5.

Matsumoto, H., & Domae, K. (2018). The effects of new international airports and air-freight integrator's hubs on the mobility of cities in urban hierarchies: a case study in East and Southeast Asia. *Journal of Air Transport Management, 71*, 160–166. 10.1016/j.jairtraman.2018.04.003.

Merkert, R., & Assaf, A. G. (2015). Using DEA models to jointly estimate service quality perception and profitability–Evidence from international airports. *Transportation Research Part A: Policy and Practice, 75*, 42–50. 10.1016/j.tra.2015.03.008.

Merkert, R., & Mangia, L. (2014). Efficiency of Italian and Norwegian airports: a matter of management or of the level of competition in remote regions? *Transportation Research Part A Policy and Practice, 62*, 30–38. 10.1016/j.tra.2014.02.007.

Milioti, C., & Odoni, A. R. (2024). Airport "level" and market concentration in Europe. *Journal of the Air Transport Research Society*, (2), Article 100007. 10.1016/j.jatrs.2024.100007.

O'Connell, J. F., & Bueno, O. E. (2018). A study into the hub performance Emirates, Etihad Airways and Qatar Airways and their competitive position against the major European hubbing airlines. *Journal of Air Transport Management, 69*, 257–268. 10.1016/j.jairtraman.2016.11.006.

O'Kelly, M. E. (1998). A geographer's analysis of hub-and-spoke networks. *Journal of Transport Geography, 6*(3), 171–186. 10.1016/S0966-6923(98)00010-6.

Organisation for Economic Co-Operation and Development Statistics. OECD. Regional GDP Per Capita, $, current prices, current PPP. 2023. URL: https://stats.oecd.org.

Örkcü, H. H., Balikçi, C., Dogan, M. I., & Genç, A. (2016). An evaluation of the operational efficiency of Turkish airports using data envelopment analysis and the Malmquist productivity index: 2009-2014 case. *Transport Policy, 48*, 92–104. 10.1016/j.tranpol.2016.02.008.

Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. CreateSpace Independent Publishing.

Ozdemir, A. (2011). Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *Journal of Hydrology, 405*(1), 123–136. 10.1016/j.jhydrol.2011.05.015.

Pauwels, J., Buyle, S., & Dewulf, W. (2024). Regional airports revisited: Unveiling pressing research gaps and proposing a uniform definition. *Journal of the Air Transport Research Society*, Article 100008. 10.1016/j.jatrs.2024.100008.

Redondi, R., Malighetti, P., Paleari, S. (2010). De-Hubbing cases and recovery patterns. Working Papers 1008, department of management, information and production engineering, university of Bergamo.

Rodrigue, J. P. (2020). *The Geography of Transport Systems* (5th Ed.). Routledge. 10.4324/9780429346323.

Rodríguez-Déniz, H., Suau-Sanchez, P., & Voltes-Dorta, A. (2013). Classifying airports according to their hub dimensions: an application to the US domestic network. *Journal of Transport Geography, 33*, 188–195. 10.1016/j.jtrangeo.2013.10.011.

Ryerson, M. S., & Kim, H. (2013). Integrating airline operational practices into passenger airline hub definition. *Journal of Transport Geography, 31*, 84–93. 10.1016/j.jtrangeo.2013.05.013.

Sun, X., Wandelt, S., & Zhang, A. (2023). Price discrimination through hidden city options? A data-driven study on the extent and evolution of skiplaggability in the global aviation system. *Journal of Air Transport Management, 108*, Article 102373. 10.1016/j.jairtraman.2023.102373.

Tian, Y., Sun, M., Wan, L., & Hang, X. (2020). Environmental impact analysis of hub-and-spoke network operation. *Discrete Dynamics in Nature and Society, DOI*. 10.1155/2020/3682127.

Tsui, W. H. K., Gilbey, A., & Balli, H. O. (2014). Estimating airport efficiency of New Zealand airports. *Journal of Air Transport Management, 35*, 78–86. 10.1016/j.jairtraman.2013.11.011.

Uher, M (2005). *Airport Dubai - Evaluation of Dubai as a First Choice Hub for International Travellers*. Institut für Transportwirtschaft und Logistik, WU Vienna University of Economics and Business. 10.57938/82721381-a961-4b3b-b9ee-3a6116c4acf9.

Wandelt, S., Dai, W., Zhang, J., & Sun, X. (2022). Toward a reference experimental benchmark for solving hub location problems. *Transportation Science, 56*(2), 543–564. 10.1287/trsc.2021.1094.

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2011). *Statistical and Econometric Methods for Transportation Data Analysis* (2nd Ed). CRC Press.

World Bank. GDP Per Capita, PPP (current international $). 2023. URL: https://data.worldbank.org.

Yoshida, Y., & Fujimoto, H. (2004). Japanese-airport benchmarking with the DEA and endogenous-weight TFP methods: testing the criticism of overinvestment in Japanese regional airports. *Transportation Research Part E: Logistics and Transportation Review, 40*(6), 533–546. 10.1016/j.tre.2004.08.003.