TRANSPORTES
www.revistatransportes.org.br
anpet
associação nacional de pesquisa e ensino em transportes
www.anpet.org.br

Effects of the 1978 airline deregulation act on aircraft industry measured by entropy statistics

Efeitos do ato de desregulamentação de 1978 na indústria aeronáutica medidos pelo método da entropia estatística

Bento Silva de Mattos[1], **José Alexandre Tavares Guerreiro Fregnani**[2], **Jesuíno Takachi Tomita**[3]

[1]Aeronautics Institute of Technology, São Paulo – Brazil, bmattos@ita.br
[2]Aeronautics Institute of Technology, São Paulo – Brazil eng.fregnani@hotmail.com
[3]Aeronautics Institute of Technology, São Paulo – Brazil, jtakachi@ita.br

Recebido:
3 de abril de 2021
Aceito para publicação:
21 de dezembro de 2021
Publicado:
11 de agosto de 2022
Editor de área:
Li Weigang

Keywords:
Aicraft design.
Entropy statistics.
Air transportation.

Palavras-chave:
Projeto de aeronaves.
Entropia estatística.
Transporte aéreo.

DOI:10.14295/transportes.v30i2.2594

ABSTRACT
There are plenty of articles and essays concerning the effects of the 1978 Deregulation Act on aviation but almost none of them addresses the impacts on the aircraft industry and the waves of immigration that took place due to cheaper airfares. The present work proposes an innovative methodology, based on the entropy statistics theory, to analyze the impact of the 1978 Airline deregulation effects on jet airliners design. The airplane database used for this task is comprised of 121 jet airliners since the 1950s and it is evident the effect of the Deregulation on design diversity close after it came into force. A special analysis of the regional sector of line aviation, i.e., the regular transportation of passengers is also carried out here. Besides, the application of the Euclidean distance for the calculation of diffusion and convergence indexes and its result is analyzed in the present work.

RESUMO
Há muitos artigos e ensaios versando sobre os efeitos da deregulamentação de aviação civil em 1978, mas quase nenhum deles aborda os impactos na indústria aeronáutica e as ondas de imigração que ocorreram devido às passagens aéreas mais baratas. O presente trabalho propõe uma metodologia inovadora, baseada na teoria da entropia estatística, para analisar o impacto da deregulamentação de 1978 no projeto de aronaves de transporte à jato. O banco de dados usado para esta tarefa é composto por 121 aeronaves desde a década de 1950, evidenciando o efeito da deregulação sobre a diversidade de projetos logo após sua entrada em serviço. Além disso, a aplicação da distância euclidiana para o cálculo dos índices de difusão e convergência e seu resultado é analisado neste trabalho. Uma análise especial do setor regional do transporte regular de passageiros também é realizado aqui.

1. INTRODUCTION

Aircraft manufacturers must offer the right products at the right time. There are plenty of examples of companies that went bankrupt because they marketed inadequate aircraft or did not have the types of airplanes that customers were asking for in their portfolios. This took place during the growth of aviation triggered by the U.S. 1978 Deregulation Act. Not only have traditional aircraft manufacturers gone out of business, but some major airlines faced restructuring, mergers, or simply bankruptcy. The effects of the 1978 Deregulation Act on the aviation industry have been profoundly studied and the experience in air transportation served as a model to encourage deregulation efforts in other areas. However, the impact of the 1978

Deregulation Act on the aeronautical industry must be better understood as it has not been studied enough.

For this purpose, the present work employs an unsupervised learning algorithm based on entropy statistics to measure the impact of Deregulation on the aeronautical industry. In addition, a verification process was carried out to evaluate whether such a tool can help manufacturers to improve their design requirements.

The computational tool developed from entropy statistics principles can be employed to analyze product diversity, perform product classification, capture market trends as well as technological changes when linked to external events, and even evaluate new aircraft concepts, by artificially inserting them in an airplane database. A product under development can be categorized into four classes: classical, breakthrough, failure, or niche product. In addition, by setting product classification as a constraint in a multi-disciplinary design and optimization framework, the aircraft manufacturers can design a breakthrough product or prevent a failure configuration that might emerge from simulations.

1.1. The Deregulation Act of 1978

Before 1978, airlines followed an intricated network of rules elaborated by the North American federal government, which determined whether a new airline could fly to a city and even specified the airfare. With limited competition, airlines had guaranteed profits, issued brochures that offered expensive services covered by high airfare. Aviation was a means of transportation for few as air traveling at that time was significantly expensive. In the 1950s and 1960s, most people could not afford to buy a flight ticket. At that time, over 80% of North American citizens had never had the opportunity to travel by air. Governmental regulation was responsible for expensive flights. Indeed, at the beginning of the 1970s, with the instauration of the oil crisis, ticket prices and operational costs raised significantly for airlines worldwide. In the United States, a commission of economists and senators decided that the government should promote competition among airlines to promote lower fares and improve the quality of services to passengers, to counteract the effects of such a crisis. As consequence, the Airline Deregulation Act was enforced in late 1978 by the American Government. This was a set of economic and operational measures tailored to remove the control of fares, routes and stimulate the entry of new airlines into the market (Cowie & Ison, 2018). As consequence, the civil aviation regulator's power over fares was eliminated, establishing market forces for the first time in the history of the airline industry. The Federal Aviation Administration (FAA), however, continued to enforce strong regulatory power in all fields related to aviation safety, airworthiness, certification, and flight operations.

Significant impacts on many aspects of everyday life followed in the wake of Deregulation. Golich (Golich, 1988) states that overall airline safety has been downgraded due to increased competition among airlines and hub operations. Thanks to cheaper airfares some important changes took place: people were encouraged to immigrate searching for better standards of living; many cities now have regular air services, and major airlines filed for bankruptcy protection. However, other effects like the one affecting airplane manufacturers were not entirely understood and have been poorly studied. Since the 1980s, traditional manufacturers of small-to-medium size airplanes have gone out of business, although those companies had been successful in the past, both in terms of sales records as well as in technology pioneering. Wrong products or good products at the wrong time are some of the reasons that help to explain the shutdown of many aircraft manufacturers. Indeed, the immense competition among airlines thanks to the 1978 Airline Deregulation that took place in the U.S. has led to the evolution of aeronautical technology that generated airplanes with exceptionally low operating costs.

Poole and Butler (Poole & Butler, 1998) describe the deregulation effects in terms of four "waves of change" or periods. The first period is marked by the expansion of the traditional and bigger airlines and the consolidation of the hub-and-spoke model. All this increased the load factor of flights and offered people more and new destinations. The second period was a counterbalancing growth in point-to-point services, largely put into practice by new low-cost carriers, the most notable of which was Southwest Airlines. The third wave is characterized by the growth of regional services provided by regional carriers, filling gaps left by major carriers. The fourth period - which extends up to 2019, before the covid-19 crisis- registers the unbundling of fares with many things that used to be included in a ticket price but are now charged extra or are no longer offered at all. At the same time, there is a reduction in the number of airlines (Poole & Butler, 1998). In part, these waves of changes are explained by another major event that also reshaped the world and, consequently, aviation: the dissolution of the Soviet Union in December 1991. Military budgets suffered big cuts and the technology flow from military programs to civil aircraft and a decrease in innovation can be observed in the airplanes developed in the following two decades.

1.2. Regional aviation

Regional aviation, particularly in the United States, experienced exponential growth after the 1978 Deregulation Act. Most aviation manufacturers around the world realized this market evolution and were able to refine and offer products that airlines wanted. Other manufacturers were slow to react and faced with insufficient sales figures encountered financial difficulties that impacted the development of new products. This was the case for the Dornier 328 and Jetstream 41 turboprops, which came after the successful Saab 340 and EMB-120 Brasilia, all in the 30-34 passenger range. The large competition that has been established in the segment of smaller airliners has led to the bankruptcy of many manufacturers, such as Fokker, Saab, Beechcraft, and Fairchild-Dornier. Despite the great success with the Saab 340, the Saab 2000, a bi-turboprop aircraft with high cruising speed, did not find resonance with the airlines, mainly due to the fact of costing the same as 50-passenger jets. Bombardier Aerospace, which acquired the aircraft manufacturer Canadair, and Embraer performed a leading role in bringing forward regional aviation, noticeably with their CRJ-100/200 and ERJ-145 airplanes, respectively. With the evolution of the market over time, the definition of regional aviation became fuzzy. The Embraer E-jet family was formally launched at the Paris Air Show on June 14th, 1999. Airplanes were designated as the ERJ-170 and ERJ-190, ERJ standing for Embraer Regional Jet designations later changed to Embraer 170 and Embraer 190. However, the concept of a twinjet with an underwing engine configuration in the 7-90 passenger capacity range was initially outlined by MPC 75 GmbH, an Airbus subsidiary (Thomas, 1992). The MPC 75 was designed with a considerable percentage of composite materials to save structural weight. Flight control was based on fly-by-wire with sides-sticks and cathodic ray tube (CRT) displays. However, it was Fairchild-Dornier that began the development of a similar airplane in 1997, when announced its 728/928 Jet family, but with little content of composite in its composition.

Regional carriers are intrinsically low-cost. Regional pilot wages are considerably lower than those of pilots from major carriers (MH Sub I, LLC dba Internet Brands , 2019); maintenance is provided by third parties, and airplanes are purchased in large orders with great discounts, and operation is secondary airports reflect in lower fees. This low-cost operational nature is a reason for the success of many regional carriers after the 1978 Airline Deregulation Act. since competition promoted a race towards lower operational costs in the new Hub-and-Spoke network concept. In the United States, the lower wages earned by cockpit crew led to the establishment of scope clauses for the operation of regional airplanes (de Mattos, et al., 2018). The ERJ 140, a shortened version of the Embraer ERJ 145 was developed to fulfill a request of American Airlines to comply with scope clauses (de Mattos, et al., 2018).

1.3. Literature

Nelson and Winter (1982) pointed out that product trajectories are not only related to periods in which the basic technological level remains unchanged, but they are also related to a stage of incremental scaling of designs. A good example of a series of scaled models in civil aircraft is the piston-powered Douglas airliner series, starting with the introduction of DC-3 in 1936. The scaling of the engine power, wingspan, and fuselage length has led to improvements on a scale of multiple factors. Considering the maximum take-off weight and range of the four-engine DC-7, these values were increased by a factor of five regarding the DC-3 figures.

With the utilization of entropy statistics, Frenken and Leydesdorff (1993) analyzed scaling patterns in terms of changes in the ratios among characteristics of 143 configurations of commercial transport aircraft. In their research, the piston-powered DC-3, and the jet-powered Boeing 707, were revealed to have triggered scaled trajectories. Some configuration characteristics of both airplanes have been scaled at different moments in time, which points to the versatility of a dominant design that allows a firm to react to a variety of customer requirements. Scaling at the level of the industry took off only after subsequently re-engineered models were introduced, like the piston propeller Douglas DC-4 and the twin-aisle Boeing 767. The two scaling trajectories in civil aircraft corresponding to the piston propeller and the turbofan paradigm can be compared with a single, less pronounced scaling trajectory in helicopter technology for data during the period 1940–1996 (Frenken & Leydesdorff, 1999). They state that management and policy implications can be specified in terms of the phases of codification at the firm and the industry level, thanks to the entropy statistics analysis.

Bosquê et al. (2006) surveyed carefully what variables would better represent the aircraft configuration and its embedded technology. Regarding Frenken and Leydesdorff's work, not only the selection of variables was improved. Both works also differ in the range of application based on entropy statistics. Bosquê carried out two analyses with considerably more individuals to validate the methodology. The first is concerned with the evolution of civil aviation transportation in the jet age (1950-2006), and the other with the evolution of fighter aircraft (1914-2009). As mentioned before, a considerable effort was made to select the variables used to describe each aircraft and for the creation of the databank containing the correct figures of the variables. A tool that was developed during this work takes the variables as input to evaluate two important evolutionary indexes: convergence and diffusion. Studies analyzing the combination of the diffusion and the convergence indexes, as well as the critical transition of the airplanes were conducted in his master thesis. A computer tool was also developed, which can be useful in the decision-making process in the conceptual design phase of aircraft.

Many researchers seek similarities between technological development and biological evolution and therefore they consider that evolutionary models are suited to model technological change (Mokyr, 1992). However, as has been repeatedly pointed out by those who endorse the adoption of an evolutionary approach, there are also substantive differences between biological evolution and technological evolution (Nelson, 1995). Therefore, evolutionary models should always be employed with caution, considering the specificities of the processes of mutation and selection under study. Recent evolutionary theorizing in biology and artificial intelligence has stressed that complex entities evolve in ways that are different from non-complex ones in important respects. This claim has also important implications for models of technological evolution, as a technological artifact is a complex evolving entity par excellence (Rosenberg, 1976).

Following Simon's work on the design of artificial systems (Simon, 1969), a technological artifact that can be described as a man-made system is constituted by interconnected components, which are intended to collectively perform some functions. The complexity of an artifact is due to the interdependencies between its components, which causes only some combinations of elements to work well together, in the sense that these combinations can achieve satisfactory levels of performance. According to Simon, good management of innovative activities consists of trying to improve the general performance of the artifact, by finding out progressively better configurations of its constituting elements.

There is a lack of formal approaches of system interdependencies for the understanding of technological innovation. However, thanks to the introduction of complex models from natural sciences in the realm of (evolutionary) economics, new lines of research arose. In this context, Kauffman's NK model of evolutionary biology has proven extremely promising and has already been adopted in many contributions in the innovation and organization literature (Kaufman, 1993). The NK model is a mathematical model considered adequate for a rugged fitness landscape. The attribution of being a tunable rugged model captures the intuition that both the overall size of the landscape and the number of its local hills and valleys can be adjusted via changes to its two parameters, N and K, with N being the length of a string of evolution and K determining the level of landscape ruggedness. Frenken and Nuvolari (2004) applied the NK model to empirical studies of technological change. They considered the examples of the early development of the steam engine (1760-1800), the aircraft evolution (1913-1984), and the development of the helicopter (1940-1983) for studies of technical change. For this purpose, they employed the NK model in combination with entropy statistics and suggested that other historical studies of technology could indeed benefit from the adoption of this type of approach.

1.4. Objectives

Aviation progress is not only triggered by technological advances but also by new business models. The 1978 Deregulation Act open new opportunities for airlines and aircraft manufacturers. Embraer and Canadair – later Bombardier Aerospace – were great beneficiaries of the regional aviation boom and this led to the second jet age of the 1990s. Thus, the present work intends to measure the impact of the Deregulation Act on the aircraft industry with a focus on regional aviation. The entropy statistics, an unsupervised learning algorithm, can also be used to map other regulatory events and therefore to help manufacturers to develop new products.

The present work proposes a methodology, based on the entropy statistics theory, to analyze the impact of the 1978 Airline Deregulation on jet airliners design. The airplane database used for this task is comprised of 121 jet airliners since the 1950s and it is evident the effect of the Deregulation on design diversity closed after it came into force. A classification analysis for some existing airliners according to Frenken criterium (Frenken & Leydesdorff, 1999) was also carried out with interesting and sound results.

A dendrogram of the airplanes used in the entropy statistics computations was carried out using *k-means* clustering. The reason behind this is to make a stress test of the airplane data present in the database mentioned in the preceding paragraph.

2. METHODOLOGY

Entropy is a thermodynamical concept that measures how energy can be distributed in a determined system. A microstate is an instantaneous record describing the energy of each molecule in that system and a given system has many possible microstates. The molecules are constantly interacting and exchanging energy and changing their states, meaning that the system must constantly be observed and measured, and the entropy law states that entropy must increase in time. The Boltzmann entropy logarithmic form.

$$S = k_B \ln W \tag{1}$$

where k_B is the Boltzmann constant and W represents all possible microstates of a system.

The information theory makes use of the entropy concept, and it was originally developed to help to increase the reliability of communication over noisy channels and to enable high data compression with acceptable losses of information. The field of information theory was incepted by Claude Shannon (Shannon, 1948), in his work "A mathematical theory of communication." In this work, Shannon proposes that the mathematical framework that best describes the problems encountered in communications is the same one of thermodynamics. Concepts such as entropy should be applied in broader terms to encompass a whole new class of phenomena. Thus, entropy can be understood as the degree of uncertainty of a transmitted message. Shannon (1948) more specifically described it as the minimum number of bits that must be communicated to as interlocutors for him to know the value of a random variable. Consider, for example, that transmission of data is composed of a series of ASCII characters. If the characters are random, the entropy of the message is seven times the number of characters (since each ASCII character is composed of seven bits, and entropy is given in bits). If, however, we know in advance that this series is a repetition of the same character, then the entropy is zero, since we, by knowing one of the characters will know all the others. On the other hand, if the message is composed of true text, the entropy is not zero but is not its maximum value (as was in the random text). The value of the entropy value is going to assume something in between since written language allows us to infer the rest of a message by knowing only parts of its content. Written language is, in other words, redundant, which reduces its entropy. English language, for example, has average entropy of 1 to 1.5 bits per character, instead of the 7 bits per character in a random message. Shannon (Shannon, 1948) proposed that the entropy of a message, x is a random variable, is given by:

$$H(x) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \quad x \in X \tag{2}$$

where H is the set of all possible messages x and $p(x)$ is the probability mass function.

Note that by using base-2 logarithm, the entropy will be measured in bits. $H(x)$ reaches its maximum when all messages have an equal probability of occurrence. Thus, $p(x) = 1/n$. And consequently, $H(x) = log_2 n$. Eq. 2 was based on the statistical definition of entropy (MTI, 2018).

As it is defined, entropy has a set of properties that are relevant for this work. They are:

- Continuity: The measure of entropy is a continuous function, which allows for easier optimization convergence.
- Symmetry: The measure of entropy is unchanged if the order of parameters p(x) is changed. This allows data for aircraft to be added at will, with no regard for the order in which it is arranged, without a change in the entropy values.
- Additively: Entropy remains the same independently on how the system is divided into subsystems, which in turn allows each aircraft to be compared to others independently.
- Another important parameter is the relative entropy, introduced by Kullback in 1951 (Kullback & Leibler, 1951). According to Cover (Cover & Thomas, 1991), relative entropy is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy or Kullback-Leiber distance between two probability mass functions *p(x)* and *q(x)* is defined by:

$$I(p|q) = \sum_{x \in X} p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{3}$$

Indeed, the relative entropy $I(p|q)$, also known as Kullback–Leibler divergence is a special case of a broader class of divergences called f-divergences as well as the class of Bregman divergences (Wikipedia, The Free Wikipedia, 2019). It is the only such divergence over probabilities that belongs to both classes. Although it is often intuited as a way of measuring the distance between probability distributions, the Kullback–Leibler divergence is not a true metric for this. It does not obey the triangle inequality, and therefore $I(p|q)$ does not equal $I(q|p)$. Nonetheless, it is often useful to consider the relative entropy index as a "distance" between distributions (Cover & Thomas, 1991). In its infinitesimal form, specifically its Hessian, the relative entropy gives a metric tensor known as the Fisher information metric. The relative entropy is always non-negative, and it becomes zero if and only if $p = q$. For applications in the evaluation of product evolution, in Eq. 3, p is a posteriori distribution and q is a priori one. From this, two important parameters can be derived, which are important for product categorization. These parameters are called convergence and diffusion. Convergence is obtained when we consider an individual p and look at a certain period past in time and consider the characteristics from another individual q in that timeframe. For the computation of the diffusion coefficient of a design that appeared in a determined time, designs that appeared after it in a determined timeframe must be selected (Frenken & Leydesdorff, 1999). Here, we consider the service entry of the aircraft as a time variable. The diffusion of a reference product can then be measured by its distance I calculated by Eq. 3 to all the members of the technological population that were selected before. The final I-values are then obtained by dividing the sum of I-values for each product pair by the number of comparisons in the timeframe of observation (Frenken & Leydesdorff, 1999).

The combination of diffusion and convergence indexes is useful for product classification. According to Frenken (Frenken & Leydesdorff, 1999), industrial products can be classified as innovations, scaled trajectories, niches, and failures. Innovative products can be recognized if they present new, redesigned, or substantially improved characteristics that are well distinguished from their predecessors. Innovative products influence their successors and introduce dominant designs. They are categorized by a low value of diffusion and a high value of convergence indexes. Scaled trajectories are projects that follow patterns defined by previous projects and influence their successors, the scaled trajectories are under the influence of dominant designs and are characterized by a low diffusion value as well as a low convergence value.

Products that fit into market niches are those complying with standards defined by previous designs and do not exert a profound influence on designs that come after them. They may be characterized by the combination of high diffusion and low convergence indexes. Failures are concepts that differ from their predecessors and do not influence their successors, being characterized by a high value of its diffusion coefficient and a high value of the convergence index as well. The kind of design classification of industrial products described before was introduced by Frenken (Frenken & Leydesdorff, 1999). Fig. 1 provides a compilation of this classification. Frenken originally designated designs in the Northeast quadrant as failures. Niche products are in the lower-right quadrant, and the scaled trajectories are in the lower-left one. An ellipse with a center at the confluence of the quadrants defines a region where the configurations are undefined, denominated here of fuzzy designs.

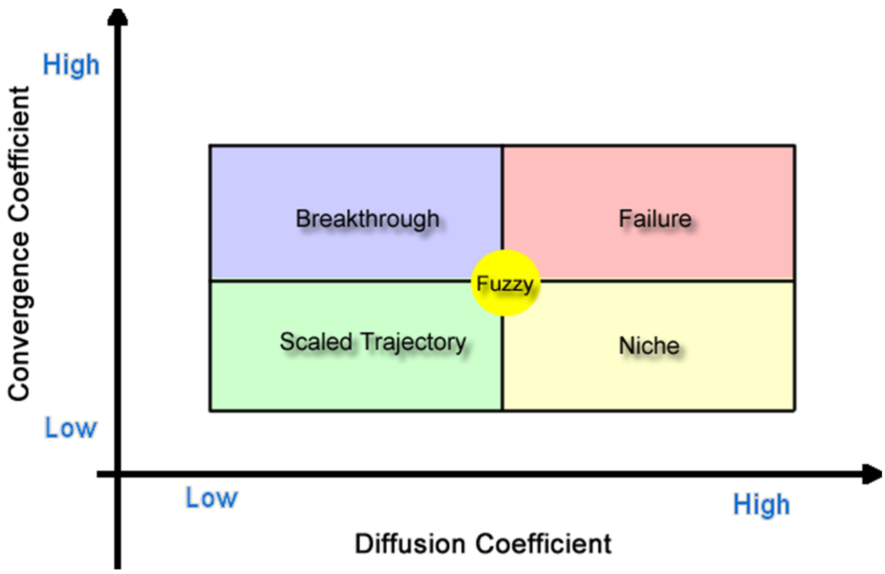

Figure 1. Product classification according to diffusion and convergence indexes

Breakthrough products are pioneers and, in general, they feature advanced technology regarding what is available at the time on the market. Sometimes, the creativity employed to define them rather than technology provides them with unique characteristics. Breakthrough products are intended to fulfill consumer needs to an extended or greater degree than existing products, while most of the time featuring more advanced technology and innovation. Breakthrough products establish a completely new market and revolutionize competition and consumer preferences in an existing market. These products enjoy the advantage of being the first movers in the market and can establish an early market share lead or a higher market share by outperforming existing products. Although all breakthrough products are first movers, they may not succeed in the marketplace in the long run.

Cluster analysis or clustering is the task of grouping a set of elements in such a way that elements inside a given group are more similar among them than those in the remaining groups. Cluster analysis itself is not one specific algorithm, but the general task to be solved. Clustering algorithms build a cluster hierarchy that is commonly displayed as a tree diagram called a dendrogram, a term that is derived from the neuronal connection *dendrite* (MathWorks, 2017). Briefly, a dendrogram is a diagram that shows the hierarchical relationship between objects (Fig. 2).

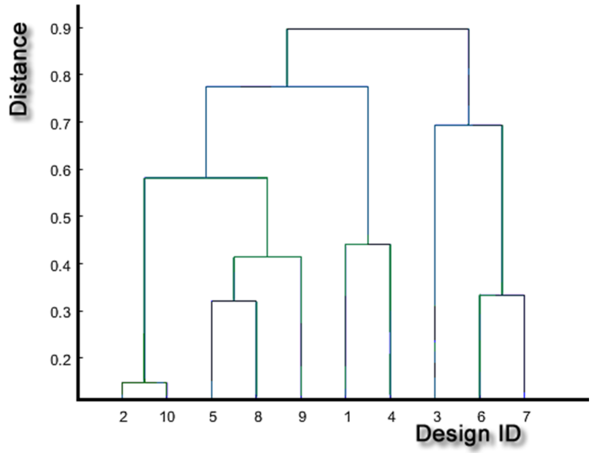

Figure 2. Example of the dendrogram

The determination of the number of clusters in a dendrogram can be also performed according to a user-defined distance threshold. In some techniques, the number of clusters is given. This will be discussed further in the dendrogram of airplanes analysis and its linkage report. There are many clustering techniques. Heuristics for some typical cluster models are given below:

- Connectivity: hierarchical clustering is based on distance connectivity. Regarding the agglomerative technique, there are many algorithms such as single linkage, complex linkage, centroid, median, and others.
- Distribution: clusters are modeled using statistical distributions (Gaussian/Normal). Besides the cluster assignments of objects, distribution-based models provide additional information like the correlation of object attributes. However, it may suffer from overfitting problems if the complexity of the chosen model is not constrained.
- Centroid: the *k-means* algorithm (Wikipedia, The Free Encyclopedia, 2021) represents each cluster by a single mean vector. A solution is then found when no movement of observation from one cluster to another will reduce the within-cluster sum of squares. The algorithm may be iterated several times with different starting configurations. The optimum of these cluster solutions is then taken. *k-means* clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances (Wikipedia, The Free Encyclopedia, 2021).
- Neural: the most well-known unsupervised neural network is the self-organizing map and these models can usually be characterized as similar to one or more of the other models, including subspace models when neural networks implement a form of principal component analysis or independent component analysis (Sureska, 2021).

For the evaluation of the data robustness of the aircraft considered, a dendrogram was elaborated based on the k-means algorithm. The k-means algorithm was developed by J.A. Hartigan and M.A. Wong of Yale University (Blackwell Publishing and Royal Statistical Society, 2012) as a partitioning technique. Its utility lies in the formation of a small number of clusters from many observations. The *k*-means clustering technique aims to partition *n* observations with *p* variables into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. Since the number of possible arrangements is huge, it is not practical to expect the single best solution. Rather, this algorithm finds a local optimum. This is a solution in which no movement of observation from one cluster to another will reduce the within-cluster sum of squares. The algorithm may be repeated several times with different starting configurations. The optimum of these cluster solutions is then selected. One criterion to evaluate the cluster configuration is to use the result that has the largest cophenetic correlation coefficient (Eq. 4). This coefficient has been used as a test for nested clusters. The calculation of the cophenetic correlation coefficient is relatively simple.

For this, consider that Y is the distance matrix and Z is a hierarchical tree, the cophenetic correlation coefficient is then given by:

$$c = \frac{\sum_{i<j}(Y_{ij}-y)(Z_{ij}-z)}{\sqrt{\sum_{i<j}(Z_{ij}-z)^2 \sum_{i<j}(Y_{ij}-y)^2}} \tag{4}$$

Yij is the distance between object i and j; Zij is the dendrogrammatic distance between objects i and j; y and z are the averaged values of Y and Z.

The cophenetic correlation coefficient is the correlation between the original distances and those that result from the cluster configuration. Values above 0.75 are considered acceptable. In statistics, the cophenetic correlation coefficient is a measure of how faithfully a dendrogram preserves the pairwise distances between the original unmodeled data points. Although it has been most widely applied in the field of biostatistics, it can also be used in other fields of inquiry where raw data tend to occur in clumps or clusters.

The interdependency between components of an artifact is a measure of its complexity, which demands only some combinations of elements to work well together, in the sense that these combinations can achieve required levels of performance and cost. Thus, the variables chosen for aircraft representation must provide an acute definition of an airliner configuration, encompassing size, topology, performance, passenger capacity, and, to some extent, design efficiency, the latter with the parameters thrust-to-weight ratio and some weights. Fuel-to-MTOW Empty weight-to-MTOW ratios are considered efficiency factors for a transport aircraft. Naturally, the description employed here (Table I) is not the ultimate one for an airliner. For instance, new variables can be introduced to describe some systems like avionics (cathodic ray tubes, LCD screens, integration with air traffic management), fuselage and wing material, ice protection, and so on.

A MATLAB® application was developed to apply the entropy statistics theory for airplane classification, more specifically, all the jet airliners since 1950. An Excel worksheet containing data for 121 jet airliners was compiled to feed the code with all data it needs. The first column of the worksheet was filled with airplane names; the second one contains the year of service entry of the airliners. The remained columns contain the parameters employed for airplane description and they are shown in Table I.

It is necessary to point out that the number of airplanes of each type that found service with airliner is not a variable used for the entropy statistics computations. This metric is more related to the success or failure of a particular design, indeed, an output of the present classification by entropy statistics. This is another point of interest: how to measure the success of an aircraft program? Some considerations are here suggested: the number of aircraft delivered to operators; years of production of the aircraft encompassing improvements to the configuration; derivatives and variants, including military ones; investment return to the manufacturer; and the profitability for airlines.

The earliest design is the de Havilland Comet 1. Technical data of airliners were normalized before calculating the convergence and diffusion coefficients. A timeframe five years around an airplane selected for the was considered for the calculation of the diffusion and convergence indexes. Fig. 3 displays the MTOW of the airplanes in the databank over the year of service entry. A tendency quadratic line is also shown in the graph. The symbols in Fig. 3 are related to the engine configuration or the variable 21 in Table I.

Table 1 – Airplanes databank variables considered

1) MTOW [kg]	12) Vertical tail taper ratio
2) Maximum Zero-Fuel Weight [kg]	13) Vertical tail quarter-chord sweep angle
3) Operating empty weight [kg]	14) Horizontal tail aspect ratio
4) Fuel capacity [kg]	15) Horizontal tail taper ratio
5) Wing reference area [m²]	16) Horizontal tail sweepback angle
6) Wing aspect ratio	17) Takeoff thrust-to-Takeoff mass ratio
7) Wing taper ratio	18) Maximum operating Mach number
8) Wing quarter-chord sweepback angle [degrees]	19) Maximum range with typical passenger payload [nm]
9) Passenger cabin external width [m]	20) Service ceiling [ft]
10) Fuselage length [m]	21) Engine configuration (1= two underwing; 2= two at rear fuselage; 3=three at rear; 4= four underwing; 5= four at rear fuselage, 6= two overwing;7=two underwing + one at rear;8=two buried in wing;9=four buried in wing)
11) Vertical tail aspect ratio	22) Typical two-class passenger capacity at 32" pitch in the economy class
	23) Seating abreast in the economy class
	24) Wing position (1= low; 2=high)
	25) Tail configuration (1=conventional; 2= "T" tail; 3=Cruciform tail)
	26) Number of decks
	27) Number of corridors in passenger main cabin

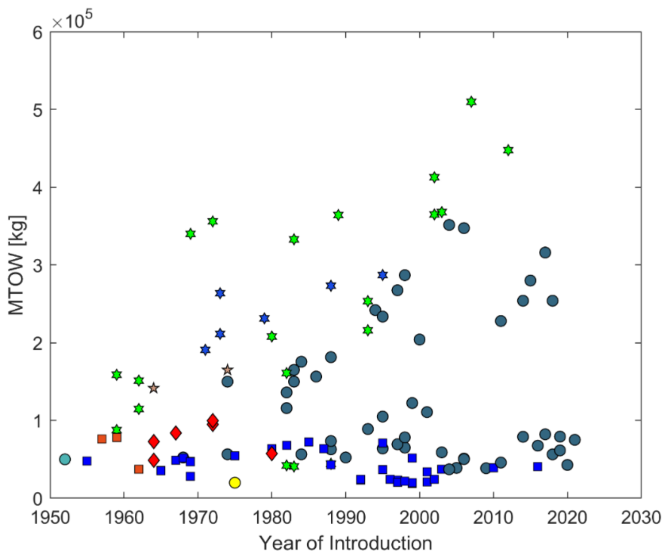

Figure 3. MTOW vs. year of entry into service year

3. ANALYSIS OF RESULTS

The resulting entropy index variation over time is displayed in Fig 4**.** There are three peaks in entropy and that related to the 1979-1984 period is in the middle of the graph.

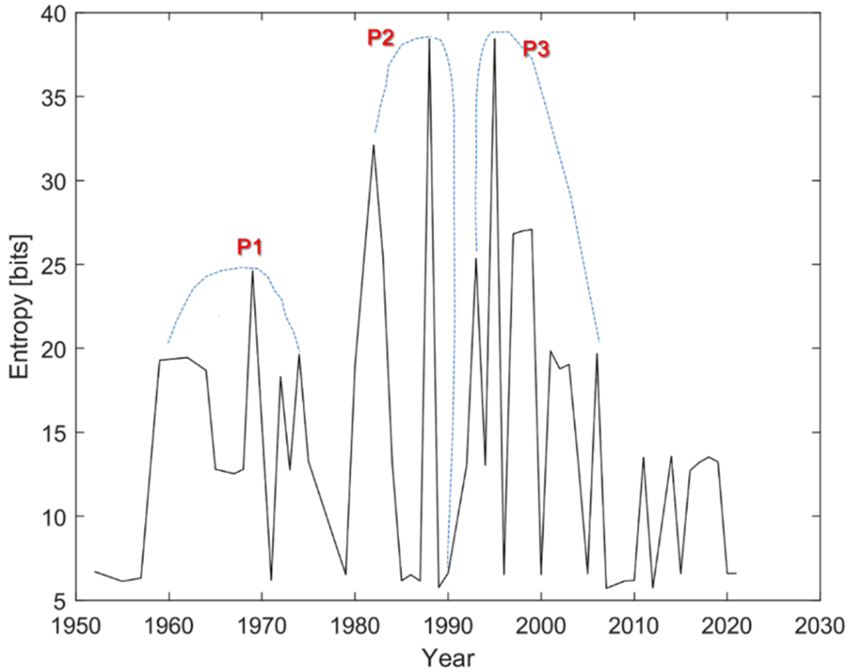

Figure 4. Entropy of the jet airliner system variation over time.

The three peaks that are displayed in Fig. 4 were analyzed and some possible explanations for their appearance have been developed as follows:

- P1 – The 1950s and 1960s are well known for a large diversity of new commercial airplane designs - some of them have never materialized, like the Boeing SST. The Sud Aviation Caravelle, Boeing 747-100, Boeing 737-100, DC-9, and many other remarkable airplanes belong to this creative decade in terms of designs that made history. From the graph, a single curve can be constructed between the several peaks ending before deregulation. Many new designs appeared because of turbofan engines and supercritical airfoils that provided more fuel-efficient airplanes with greater range.
- P2 – The turbofan engine and supercritical airfoil triggered a new generation of high-efficiency airplanes that were needed after the 1973 oil crisis. The Boeing 737-200 underwent a deep redesign from which the -300 version with turbofan engines emerged, and consequently, sales soared. The 1978 Deregulation Act shook the aviation market, and the aircraft manufacturers were agile enough to cope with the new world that appeared thereafter. Thus, the P2 entropy peak is a result of all these factors: technology, Deregulation, and the 1973 oil crisis that triggered innovation among aircraft manufacturers.
- P3 – It took some time after the Deregulation for the introduction of efficient regional jet airplanes. In the 1980s most of them were turboprop airplanes. This peak can be credited to regional jets of the 1990s and early 2000s.
- The P2 and P3 peaks could merge producing one.

Fig. 5 shows two graphs: the number of aircraft that entered service by year, and the calculated entropy represented by a 6-degree polynomial curve. The entropy level rise in the 1960s and late 1990s reveals two jet ages, the second jet age being credited to the regional airplanes. The turboprop airplanes like the EMB-110 Bandeirante, EMB-120 Brasilia, and SAAB 340 were the first airplanes benefited by the 1978 Deregulation Act with the jet airplanes soon following suit.

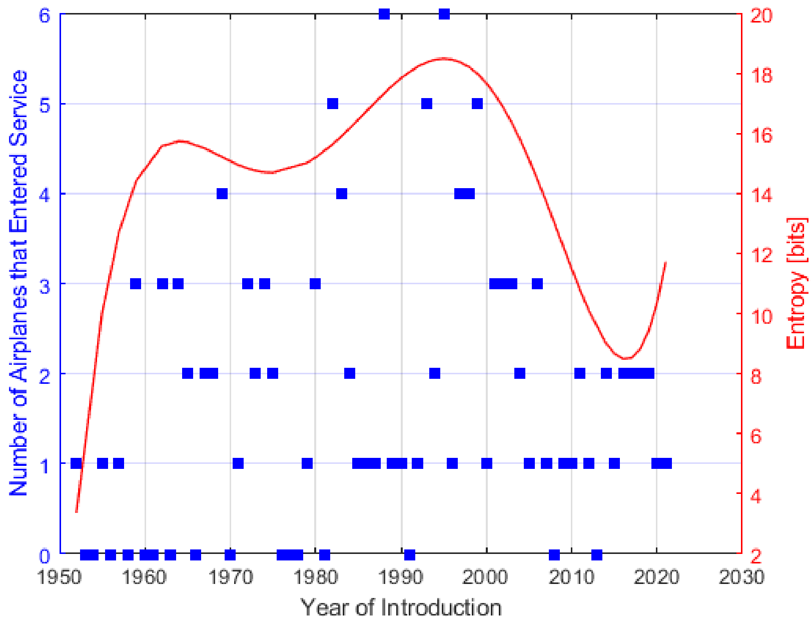

Figure 5. Number of airplanes by year of introduction and Entropy mapped by a 6-degree polynomial

The accumulated entropy over time is shown in Fig. 6. It seems to be stagnating in recent times. This might be a sign that a new revolution in technology is needed.

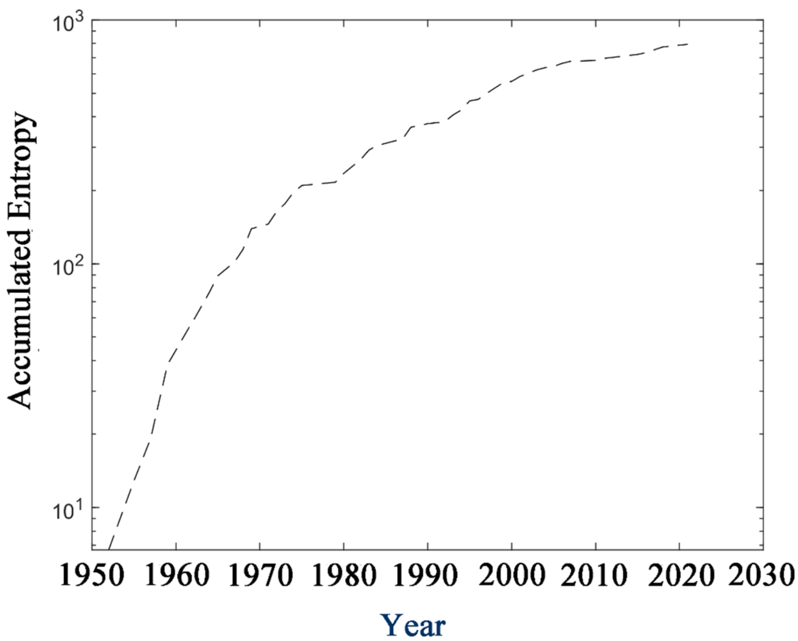

Figure 6. Accumulated entropy over time

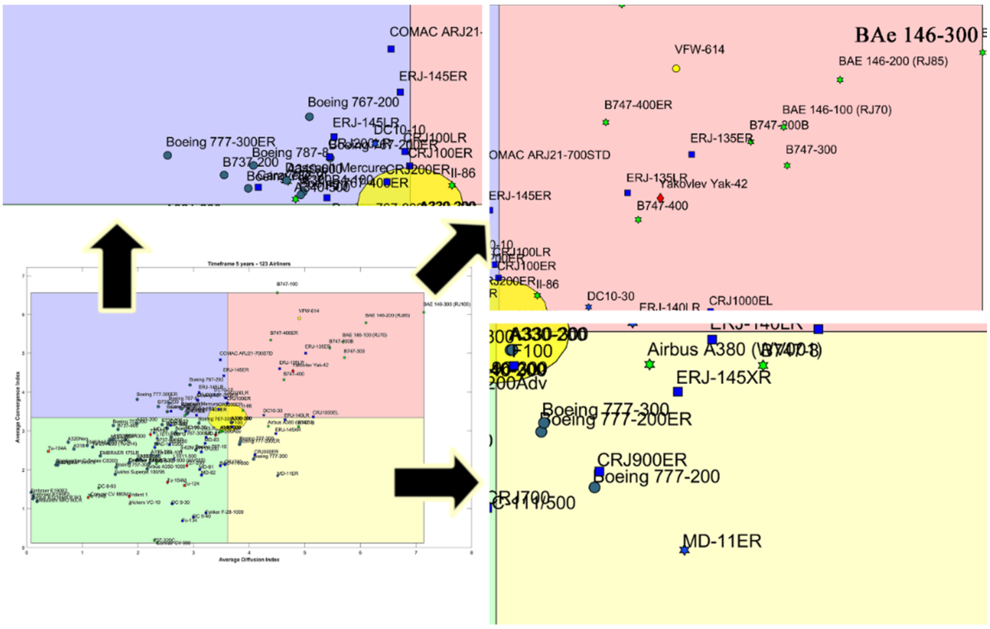

Figure 7. Design classification based on entropy statistics (timeframe of 5 years)

The classification of the designs in the databank according to Frenken's criterium is shown in Fig. 7. A time window of five years was utilized for the calculation of both convergence and diffusion indexes. In fact, instead of defined boundaries between the quadrants, a better approach would be to employ a color gradient for the classification. The normalization of the design variables is in some cases different from those presented by de Mattos *et. al* (de Mattos, et al., 2018). Fig. **7** represents a final output for the airplane classification and sometimes it is hard to visualize where a determined airplane is in the graph. For this reason, the **related computational** code also outputs a list containing the classification of all of them. The entropy statistics methodology results for the three ERJ 145 versions are analyzed as follows:

- ERJ 145ER was positioned in the breakthrough region but remarkably close to that belonging to the failure one. Indeed, this seems to be correct due to the shorter range of this version. The ER was the first ERJ 145 version and entered service in 1997. Embraer soon perceived this shortcoming and quickly developed and offered the LR version with greater range, already in 1998.
- ERJ 145LR. This version was labeled as a breakthrough design which also seems to be correct. Why then was the CRJ100, which entered service earlier in 1992, not also a breakthrough? This can be possibly credited to its shorter range (1,305 nm) and heavier MTOW (23,133 kg). Probably in this category, the ERJ 145LR seems to present the right characteristics to be classified as a breakthrough. Bombardier introduced the CRJ-200LR in 1996 (Airfleets.net, 2017), which features a 1,800-nm range (50 PAX) and better airfield performance when compared to the -100ER version. Alongside the ERJ-145LR, the CRJ-200LR airliner appears in the classification box as a breakthrough concept.
- ERJ 145XR. This airplane seems to have been designed to fulfill a market niche. This version of the ERJ 145 presenting improved performance was purchased by ExpressJet (former Continental Express) only, a subsidiary of Continental Airlines. The Newark airport was the main operational base for this type. Thanks to the better performance when compared to the LR version, the ERJ 145XR enabled ExpressJet to service more distant cities from Newark.

Embraer had the correct perception of the regional aviation market in terms it was demanding even more sophisticated airplanes, such as the ERJ145 family. However, the financial crisis that hit the company in the early 1990s explains the struggles to freeze an optimal ERJ 145 configuration from the time the CRJ-100 entered service. Thus, Embraer due to its inexperience with commercial jet airplanes considered four configurations for the 50-seat airliner, configurations that moved beyond the conceptual phase. This cost Embraer the pioneering in the 50-seat regional jet sector and Canadair was alone until the service entry of ERJ 145 in 1997.

Other airplanes labeled as "fuzzy" are the Boeing 747-300 and Antonov An-148-100. Table II shows some selected cases where we believe the airplanes were correctly labeled. The four-engine Airbus A340-500 airliner was labeled as a breakthrough. Such classification does not mean a sales success, the classical case being that of the Concorde supersonic airliner. Although many Concorde innovative features are present in modern airplanes, the ogival-wing supersonic airliner was a sales disaster. Regarding A340-500 it was introduced by the manufacturer as the world's longest-range commercial airliner. This may be a reasonable explanation for its classification as a breakthrough, even considering that the type records only 34 aircraft delivered (Wikipedia, 2017). Undoubtedly a questionable classification received the Boeing 747-400ER, an improved version of 747 with increased range (Wikipedia, 2017). It was labeled as failure/fuzzy but a total of 694 of the 747-400 series aircraft were delivered.

Table 2 – Airplane classification according to Frenken's criterium (Airfleets.net, 2017)

Airplane	Classification	Units delivered or built*
BAe 146-100	Failure	47
Boeing 747-400ER	Failure	6
Boeing 767-300ER	Scaled trajectory	583
Boeing 767-200ER	Breakthrough	121
Boeing 767-200	Breakthrough	259
Boeing 777-200LR	Scaled trajectory	59
Boeing 777-300ER	Breakthrough	839
Boeing 737-800	Scaled trajectory	4991
Boeing 737-900	Scaled trajectory	52
Boeing 737-900ER	Scaled trajectory	505
Boeing 787-9	Breakthrough	305
Caravelle VI	Breakthrough	109
CRJ-100ER	Failure	173
CRJ-100LR	Failure	47
CRJ-200LR	Breakthrough	298
ERJ145EP/ER/EU/MP	Failure	146
ERJ145LR/LU	Breakthrough	423
ERJ145XR	Niche	110
ERJ135ER	Breakthrough	31
ERJ135LR/LU	Failure	94
ERJ140LR	Failure	74
VFW-614	Failure	19

*Data from Wikipedia

The Boeing 737-800/900 were categorized as classic. This seems to be correct because their configurations turn to be commonplace in the medium-capacity market and the differences among them are not so big. Not only is the Boeing 737 the highest-selling Boeing aircraft, but it is also the highest-selling of any commercial passenger jet series in history up to the year 2018. Several versions and variants were brought to the aviation market since the Boeing 737-100 low-bypass twinjet entered commercial service back in 1968. The Boeing 737-900 was a stretch of the -800 and retained the emergency exit layout of its predecessor, which restricted its maximum seating capacity. The 737-900 also had the same MTOW and fuel capacity as the -800, which limited its range. These shortcomings prevent the 737-900 from effectively competing with the Airbus A321 (Hearn, 2017) and only 52 airplanes have been delivered until early 2021. The Boeing 737-900ER was an improvement against the -900 thanks to an increased passenger capacity and to its uprated engine and other features Boeing Co. incorporated into its design. Thus, it recorded a better sales figure but still far below that numbers achieved by the -800 version. The present framework for airplane classification was not able to properly differentiate the Next-Generation B737 types, all of them being categorized as scaled trajectories. This indicates the necessity of the introduction of new parameters to better define the airplanes in the databank, besides the existing ones.

Using the same databank that was employed in the entropy statistics analysis – variables were again normalized - the Euclidian distance between the airplanes was calculated and a dendrogram of the airplanes was then generated using MATLAB® tools (MathWorks, 2017). Fig. 8 shows the resulting Dendrogram and an enlarged part of it, selected for a deeper analysis. Airbus A321-200 and A321neo have been grouped adequately; the B737-300 and B737-500, which feature two engines in an underwing configuration area feature similar passenger capacity, are part of a selected group. Worthy of mention is the positioning in the dendrogram of the Tu-204 and its stretched version into a unique group. The positioning in the dendrogram enables one to know if a determined airplane is a singular one or its category may fit into one of the following possibilities: a niche airplane; a creator of a new market segment; or a failure design. Thus, the analysis of Fig. 8 reveals a precise grouping of the airplanes. This corroborates that sound parameters were used for the airplane classification.

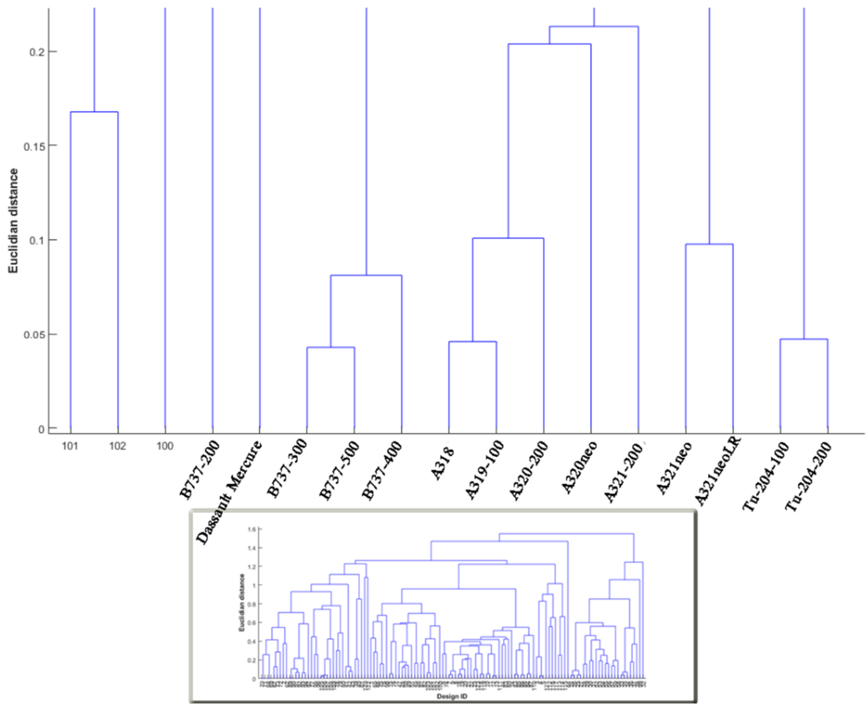

Figure 8. A enlarged part of the airliner dendrogram showing details of airplane grouping

Instead of using Eq. 3 for the calculation of the convergence and diffusion indexes, the Euclidean distance used to construct the Dendrogram was employed. Again, a timeframe of five years before and after the introduction of a determined airplane was considered for the calculation of the new convergence and diffusion indexes, respectively. The distance for each airplane in the timeframe was divided by the sum of all distances that were computed during the calculation of the indexes. The resulting classification quadrants are shown in Fig. 9. It is considerably different from that of Fig. 7. Table III shows a comparison between the two approaches for the jet airliner classification. As can be easily determined, there is a considerable disagreement between the two methods, and the entropy statistics approach is the far more exact one because the classification provided by using Euclidian distance has no match with the market fortune of the airplanes analyzed.

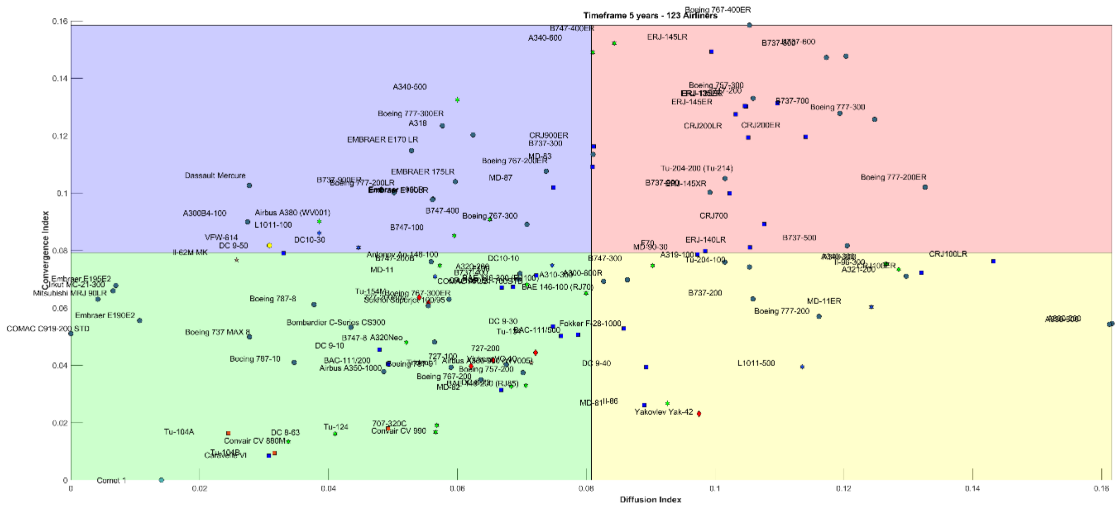

Figure 9. Classification map obtained with Euclidean distance

Table 3 – Airplane classification using Euclidean distance

Airplane	Classification Entropy	Classification with Euclidean Distance
BAe 146-100	Failure	Scaled trajectory
Boeing 747-400ER	Failure	Breakthrough
Boeing 767-300ER	Scaled trajectory	Scaled trajectory
Boeing 767-200ER	Breakthrough	Breakthrough
Boeing 767-200	Breakthrough	Scaled trajectory
Boeing 777-200LR	Scaled trajectory	Breakthrough
Boeing 777-300ER	Breakthrough	Breakthrough
Boeing 737-800	Scaled trajectory	Failure
Boeing 737-900	Scaled trajectory	Failure
Boeing 737-900ER	Scaled trajectory	Breakthrough
Boeing 787-9	Breakthrough	Scaled trajectory
Caravelle VI	Breakthrough	Scaled trajectory
CRJ-100ER	Failure	Niche
CRJ-100LR	Failure	Niche
CRJ-200LR	Breakthrough	Failure
ERJ145EP/ER/EU/MP	Failure	Failure
ERJ145LR/LU	Breakthrough	Failure
ERJ145XR	Niche	Failure
ERJ135ER	Failure	Failure
ERJ135LR/LU	Failure	Failure
ERJ140LR	Failure	Breakthrough
VFW-614	Failure	Breakthrough

4. CONCLUSIONS

An unsupervised learning algorithm was employed to analyze the impact of the 1978 Deregulation Act on aircraft design. The level of this impact was measured by entropy statistics reflecting the aircraft diversity that emerged in the wake of the Deregulation. Two waves of new or improved aircraft variants were captured by the model, in the late 1980s and mid-1990s.

The Airline Deregulation Act of 1978 was one of the most important events in aviation history. It reshaped the airport infrastructure, the aircraft industry, the way airlines do business, and the world population map. It provided a huge boost for low-cost carriers. This characteristic contributed to the resurgence of regional aviation, which, in turn, contributed to regional aircraft manufacturers playing an important role in this new market. Embraer and Bombardier managed to quickly fulfill the market needs at the right time.

The tool based on entropy statistics also enables the classification into four categories: scaled trajectory, failure, breakthrough, and niche. This classification was compared with the market acceptance of the aircraft analyzed and the agreement was revealed to be compatible with what is observed for most aircraft. These good results enable the elaboration of further considerations for the success of some aircraft types like the Embraer ERJ-145 and the Bombardier CRJ-200. Although deregulation of the aviation sector was a key factor for Embraer's success, the deregulation alone would not be enough to keep that company aloft for so many years. The bankruptcy of many manufacturers of regional airplanes, like Jetstream, Fokker, Fairchild Dornier, supports this assertion. The Brazilian company could correctly capture the evolution of the regional aviation market, which was steadily demanding an increase in airplane capacity, performance, and more comfort to passengers. The innovation brought about by Embraer to its product line was the incorporation of technologies that were already mature in larger airplanes.

By the end of the eighties, the regional aviation market was suddenly demanding more comfortable and speedier airplanes. This movement urged the aircraft manufacturers to develop more efficient regional jets (RJs), with a capacity below 100 seats presenting design ranges between 800-2,000 nm and fitted with up-to-date avionics and engines that made them economic to operate with a capacity lower than 100 passengers. These airplanes not only replaced turboprops on most routes but enabled the creation of new routes to more distant destinations. Thanks to their higher cruise speeds, the regional jets present a higher daily utilization rate than one of the turboprops. Besides, their performance enables them to return to home base for overnight maintenance. All this reduces or overcomes the drawback of higher fuel consumption. Another important aspect of RJs is their ability to provide airline service at additional airports in major metropolitan areas. Regional jets are significantly less noisy than medium-capacity airliners such as Boeing 737-600, Airbus A320-200, and MacDonnell Douglas MD-80s. RJs off-airport noise signature is like that of a twin-engine propeller general aviation aircraft.

Finally, considering that the entropy statistics tool correctly could categorize the airplanes that were analyzed, it is now inserted in a design and optimization framework to develop a successor for the Boeing 767-200 airplane. The individuals of a multi-objective optimization task are categorized during an optimization task and a constraint for one of the following outputs is set: breakthrough, failure, or niche designs. The ongoing results are indicating that passenger capacity, number of aisles, and number of engines are the two most relevant design variables that impact classification.

REFERENCES

Airfleets.net (2017). *Production List: Canadair Regional Jet*. Available at: <https://www.airfleets.net/listing/crj-38.htm> (Accessed: 08/12/2019).

Blackwell Publishing and Royal Statistical Society (2012). *Algorithm AS 136: A K-Means Clustering Algorithm*. Available at: https://www.labri.fr/perso/bpinaud/userfiles/downloads/hartigan_1979_kmeans.pdf (Accessed:10/12/2019).

Bosuqe, A. (2006). *Entropy Statistics Applied to Airplane Evolution*. Curitiba, 11th Brazilian Congress of Thermal Sciences and Engineering - ENCIT 2006.

Cover, T. M.and Thomas, J. A. (1991). *Elements of Information Theory.* New York: Wiley&Sons. DOI: 10.1002/0471200611

Cowie, J.and Ison, S. (2018). *The Routledge Handbook of Transport Economics*. 1st ed. New York: Routledge. DOI: 10.4324/9781315726786

Mattos, B.S.; Fregnani, J.A. and Magalhães, P. C. (2018). *Conceptual Design of Green Transport Airplane*. Sharjah: Bentham eBooks. DOI: 10.2174/97816810832781180301

Frenken, K. and Leydesdorff, L. (1999). *Scaling Trajectories in Civil Aviation (1913-1997).* Research Policy, Volume 29, pp. 331-348. DOI: 10.1016/S0048-7333(99)00037-2

Frenken, K. and Nuvolari, A. (2004*). Entropy Statistics as a Framework to Analyse Technological Evolution*. Applied Evolutionary Economics and Complex Systems, Edward Elgar Publishing. DOI: 10.4337/9781845421564.00010

Golich, V. L. (1988). *Airline Deregulation: Economic Boom or Safety Burst?* Transportation Quarterly, pp. 159-179.

Hearn, G. (2017). *Aircraft Comparison: Going Large*. Air Finance Journal, Issue October/November, pp. 38-40.

Kaufman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution.* 1st ed. New York: Oxford University Press. DOI: 10.1142/9789814415743_0003

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. Annals of Mathematical Statistics, Volume 22, pp. 79-86. DOI: 10.1214/aoms/1177729694

Massachusetts Institute of Technology (2018). *A Statistical Definition of Entropy*.

Available at: http://web.mit.edu/16.unified/www/FALL/thermodynamics/notes/node56.html (Accessed 13/09/2018).

MathWorks (2017). *Dendrogram*. Available at: <https://www.mathworks.com/help/stats/dendrogram.html> (Accessed:22/12/2017).