



Airline route profitability analysis on Bigdata using Firefly Algorithm

¹E. Kasturi, ²Dr. S. Prasanna Devi, ³S. Vinu Kiran

¹Research Scholar MSUniversity, Dept. of CSE, SSE, Saveetha University, Chennai

²HOD, Dept. of CSE, Apollo Engineering College, Chennai

³PG Scholar, Dept. of CSE, Apollo Engineering College, Chennai

ARTICLE INFO

Article history:

Received 28 January 2015

Accepted 25 February 2015

Available online 6 March 2015

Keywords:

Aviation, BigData analytics, Firefly Algorithm, Decision analysis

ABSTRACT

In this paper we discuss about the usage of Big data techniques on data mining and data analytics principles over a large-scale structured, semi-structured and unstructured data sets gathered from various aviation sources. We are using those data with Firefly algorithm to find out the most profitable routes based on the gathered data sets. We have implemented for a very less specific data which can be further extended in future.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: E. Kasturi, Dr. S. Prasanna Devi, S. Vinu Kiran., Airline route profitability analysis on Bigdata using Firefly Algorithm. *Aust. J. Basic & Appl. Sci.*, 9(10): 313-320, 2015

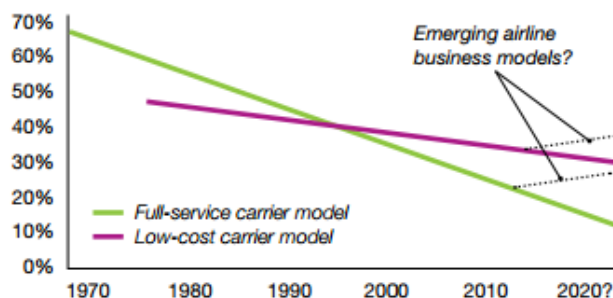
INTRODUCTION

To ascend over choppy weather, it is decisive for airlines to economize every single cent and favorably employ their assets to cut down expenses and enhance profitability.

These involve meticulous analysis of information and Intelligent judgments. One of the fundamental elements which define the future of an airline is Route Profitability.

Airline business model evolution

Percentage of global airlines at which returns exceed cost of capital



Source: IBM Institute for Business Value analysis, Carrier business model analysis based on IATA estimates in 2009.

Fig. 1: Almost all carriers will see the cost of capital increase

Route profitability systems determine whether a specific course of journey is rewarding or not during a given time period. Route profitability analysis authorizes the management to make choices on

whether to alter, supplement or exclude routes from the airline's schedule. All major aviation industrial experts are seeking for route profitability measures on upcoming years as shown in Fig 1.

For each airline, success or failure is determined by a unique mix of circumstances and decisions. The question is which of these matter the most? Precise answers are specific to each carrier, but understanding how these factors impact carriers reveals many important truths about viability of the prevailing airline business models (Airport authority of India)

I. Mining And Analyzing of Large Scale Data:

A. Large scale data (or) Big Data:

In the distributed systems world, "Big Data" started to become a major issue as these data can be of unstructured, semi-structured or structured, in the late 1990's due to the impact of the world-wide Web and a resulting need to index and query its rapidly mushrooming content. Database technology (including parallel databases) was considered for the task, but was found to be neither well-suited nor cost-effective for those purposes. The turn of the millennium then brought further challenges as companies began to use information such as the topology of the Web and users' search histories in order to provide increasingly useful search results, as well as more effectively-targeted advertising to display alongside and fund those results.

B. Data Mining:

Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. It is actually the process of finding the hidden information/pattern of the repositories.

C. Data Analysis:

Data analysis by Web-scale information extraction or the problem of creating structured tables using extraction from the entire web, is gathering lots of research interest. We perform a study to understand and quantify the value of web-based survey on educational institutions. We believe this is the first study of its kind, and gives us new insights for information extraction over the Web.

II. Problem Formulation:

A. Objective function:

We are using objective function for aircraft routing problem (Xin-She Yang, 2011). The objective function of this work is to find the optimal settings of route profitability control variables which minimizes the real route-cost loss and route deviation. Hence, the objective function can be expressed as:

$$f = \min(P_L) \dots \dots \dots (1)$$

The total real cost of the system can be calculated as follows

$$P_L = \sum_{k=1}^{N_L} G_k [V_i^2 + V_j^2 - V_i V_j \cos(\delta_i \delta_j)] \dots \dots (2)$$

Where, N_L is the total number of lines(routes) in the system; G_k is the optimality (profitable route) of the line „k“; V_i and V_j are the speed of the source and destination air-fleet capacities of the line; δ_i and δ_j are angles of the aircraft. Minimization of the function will give the minimized cost.

III. Proposed Systems:

A. Firefly Algorithm:

This paper is used to solve non-linear design problems. This technique is used to reduce revenue loss and improve the airline route profile by analyzing and optimizing the route for aviation. These are achieved by the help of statistical analysis to adjust the control parameters for parameter tuning (Yang, X.S., 2008) such as airline route, decision (Zhe L., et al., 2011) settings and variable outputs from previous data set are the control parameters. And this algorithm is used to determine better global optimal solutions that are given below. The global optimal solutions are achieved by the help of two test functions one is singularity (or) stochastic the other is deterministic. Stochastic methods produce different solution events with the same starting point. Deterministic methods produce the same set of solution events with the same starting point. These deterministic algorithms are used to find the efficient local optima.

It is difficult to find the global optimal solution. So, stochastic methods are used to find global optimal solution. Most stochastic algorithms can be considered as metaheuristic (APMS Research Overview, 2012), and good examples are genetic algorithms (GA) and particle swarm optimization (PSO) (Kennedy, J., R.C. Eberhart, 1995). Many modern meta-heuristic algorithms were developed based on the swarm intelligence in nature. Stochastic method has a deterministic component and a random component. Stochastic methods can take many forms such as simple randomization by randomly sampling the search space or by random walks.

B. Fireflies Algorithm Characteristics:

The fireflies characteristics are governed by three rules which are given below:

1) All fireflies are unisex so that one firefly is attracted to remaining fireflies regardless of their sex.

2) Attractiveness is proportional to their brightness thus for any two flashing fireflies the low brighter one will move towards the higher brighter one. The attractiveness is directly proportional to the brightness and they both decrease as their distance increases. If no one is brighter than a specified firefly, it moves randomly.

3) The brightness or light intensity of a firefly is determined by the landscape of the objective function which is to be optimized.

C. Attractiveness and Light Intensity:

In the Firefly algorithm, there are two important issues: the variation of the light intensity and the formulation of the attractiveness. We know, the light intensity varies according to the inverse square law i.e.:

$$l(r) = \frac{l_s}{r^2} \dots \dots \dots (9)$$

Where $l(r)$ is the light intensity at a distance r and l_s is the intensity at the source. When the medium is given the light intensity can be determined as follows:

$$l(r) = l_0 e^{-\gamma r} \dots \dots \dots (10)$$

To avoid the singularity at $r=0$ in the equations, it can be approximated in the following Gaussian form.

$$l(r) = l_0 e^{-\gamma r^2} \dots \dots \dots (11)$$

As we know, that a firefly's attractiveness is proportional to the light intensity seen by adjacent fireflies and thus the attractiveness β of a firefly is determined by β_0 and the β_0 is the attractiveness at $r=0$

$$\beta = \beta_0 e^{-\gamma r^m} \quad (m \geq 1) \dots \dots (12)$$

D. Distance:

The distance between any two fireflies i and j at x_i and x_j respectively, the Cartesian distance is determined by equation (13) where $X_{i,k}$ is the k^{th} component of the spatial coordinate x_i of the i^{th} firefly and d is the number of dimensions.

$$r_{ij} = \sqrt{\sum_{k=1}^d (x_{i,j} - x_{j,k})^2} \dots \dots \dots (13)$$

E. Movement:

The movement of a firefly i is attracted to the other more attractive (brighter) firefly j is determined by equation (14)

Where the second term is due to the attraction while the third term is randomization with α being the randomization parameter r [3] and β being the vector of random numbers drawn from a Gaussian distribution.

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha \epsilon \dots \dots (14)$$

It is worth pointing out that is a random walk partial towards the brighter fireflies and becomes a simple random walk if $\beta_0 = 0$. The tuned parameter (Xin-She Yang, 2011 set used in this work is described later.

F. Algorithm Steps:

- Step 1: Start the program
- Step 2: Enter the load system input data
- Step 3: Initial population of fireflies x_i ($i = 1, 2, \dots, n$)
- Step 4: To determine Light intensity l_i at x_i is given By $f(x_i)$
- Step 5: Set the iteration count $iter=1$
- Step 6: To calculate i^{th} firefly for $i = 1: n$ all n fireflies
- Step 7: To calculate j^{th} firefly for $j = 1:n$ all n fireflies

- Step 8: To check if $(I_j > I_i)$, Move firefly i towards j in dimension; end if
- Step 9: To calculate attractiveness, when Attractiveness varies with distance r .
- Step 10: To evaluate new solutions and update light Intensity
- Step 11: end for j
- Step 12: end for i
- Step 13: Rank the fireflies and find the current best
- Step 14: To evaluate $Iter=Iter+1$
- Step 15: Check $Iter > Iter_{max}$; if on no condition go to step 4.
- Step 16: Print the results
- Step 17: Stop the program.

IV. Algorithm Implementation For Route Profitability:

We have used the airline data downloaded from aviation authority of India and flight operation details from Australia.

Using two datasets we have analyzed the route and their performance using parameters such as passengers, operating cost, travel charges, freight, mail and the route distance (IBM Institute for Business Value analysis et 2011). We have correlated the route with the other parameters to compute the optimum results. We have also forecasted the freight and passenger patterns based on the used data set.

A. Analysis of Australian flight data:

We took flight data for only 3 locations as a case study to derive the output.

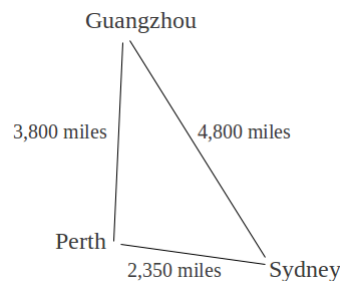


Fig. 2: Case study for 3 Australian air ports

1. Market:

- At this point there is no other airline flying from Guangzhou to Perth, and there is only one competitor flight is flying from Guangzhou to Sydney. Sydney to Perth is a domestic route with one competitor.
- The market grows with GDP.
- 15% of passengers from Guangzhou to Sydney (or return) have Perth as the origin or final destination
- China Southern Airlines' current load factor in Guangzhou to Sydney flight is 80%
- Ticket price to fly from Guangzhou to Sydney or Guangzhou to Perth is \$600 one way

2. Flight distance:

- Guangzhou – Sydney = 4,800 miles
- Guangzhou – Perth = 3,800 miles
- Sydney – Perth = 2,350 miles

3. Aircraft type limitations:

- Airbus 319 (120 seats) has a maximum fly range of 4,000 miles
- Airbus 320 (165 seats) has a maximum fly range of 6,000 miles
- Airbus 330 (240 seats) has a maximum fly range of 7,000 miles

4. Costs:

- Airbus 319 (120 seats) has a fixed cost of \$41,000
- Airbus 320 (165 seats) has a fixed cost of \$62,000

We analyze this case which requires us to estimate the potential market size of route Guangzhou-Perth, the growth opportunity, the competition as well as our client's resources. So, we worked to find out the number of flights per day, the number of passengers per flight and the percent of these passengers that actually have Perth as the origin or destination. One of the Airlines currently flies twice a day from Guangzhou to Sydney and back. It currently operates Airbus 320 in this route, with 80% load factor. 15% of the passengers have Perth as the origin or destination.

By considering the aircraft flight limitations, this gives us $165 * 80\% = 132$ passengers per flight or

$132 * 2 = 264$ passengers per day (one way), of which 15% or 40 have Perth as the origin (and final destination). We should have in mind that if the client offers the new Guangzhou-Perth direct flight (predicted new route), it will reduce the load factor of the Guangzhou-Sydney route from 80% to 68%.

There is only one competitor flying from Guangzhou to Sydney, but this competitor has a code-share agreement with a domestic airline that flies from Sydney to Perth. From Guangzhou to Sydney the competitor flies an A330 (240 seats) daily with the same 80% load factor. Also, 15% of its passengers have Perth as the origin or final destination. This gives us an additional $240 * 80\% * 15\% = 30$ passengers/day. For eg: all airlines under this route charge \$600 per one way from Guangzhou to Sydney or Guangzhou to Perth.

We have observed that the passenger pays the same fare to fly from Guangzhou to Sydney (4,800 miles) and from Guangzhou to Perth (4,800 + 2,350 = 7,150 miles). So basically we have the same price and potentially a faster trip. Based on analysis of passengers travel requirement, we know the potential demand; although there is an expected growth in the industry. Let's assume the total fixed cost of operating A319 and A320 are \$41,000 per flight and \$62,000 per flight respectively (<http://www.airservicesaustralia.com>). There are some variable costs, but they are marginal. Basically for route Guangzhou-Sydney the client must use an A320, but A319 is an option for route Guangzhou-Perth. Let's analyze the daily revenue, cost, and profit for each route:

Table 1: Analysis of To and Fro Flights Costs

Guangzhou to	Sydney	Perth
Revenue	\$67,320	\$42,000
Cost	\$62,000	\$41,000
Profit	\$5,320	\$1,000
Profit Margin	8%	2.4%

We are assuming from the analysis outcome that the A320 for the main Guangzhou to Sydney route and A319 for the new Guangzhou to Perth route. As we can see, the former route will remain profitable (8%) and the new route will give a 2.4% margin. Taking into consideration that a 2.4% margin is not a very bad number for the airline industry and

that our assumption (optimized route decision) does not take into consideration an additional demand generated by the new faster route from Guangzhou to Perth, based on our output, at least launch a trial of this new route. The former route, from Guangzhou to Sydney, will remain profitable.

B. Outcome based on current flight datasets

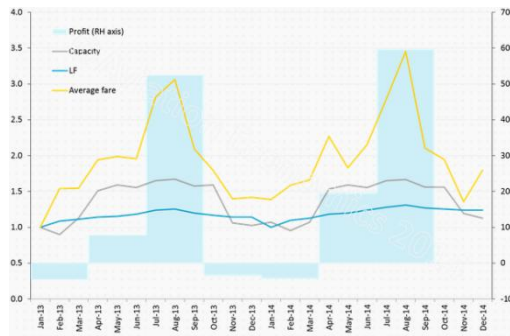


Fig. 3: Graph representing average profit calculated based on capacity(freight, passengers, and mails) to the average fare for each operated routes.

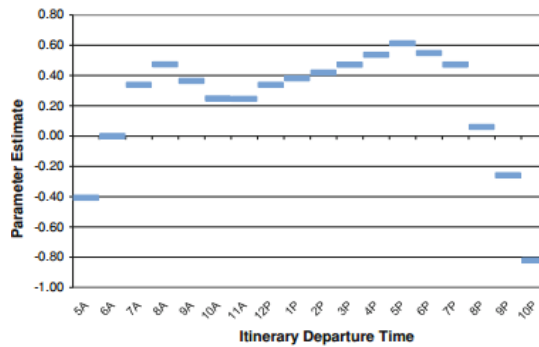


Fig. 4: Passenger departure preferences graph

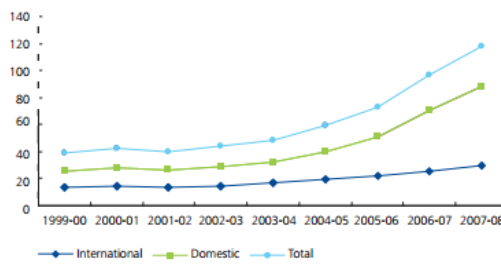


Fig. 5: Passenger growth trends (Airport authority of India)

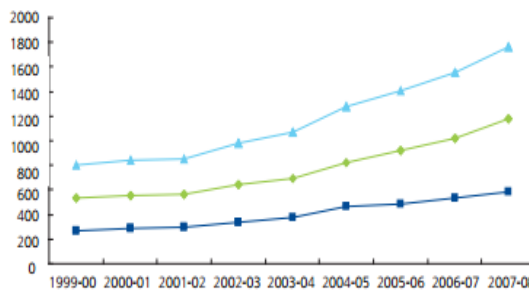


Fig. 6: Air cargo growth trends (Airport authority of India)

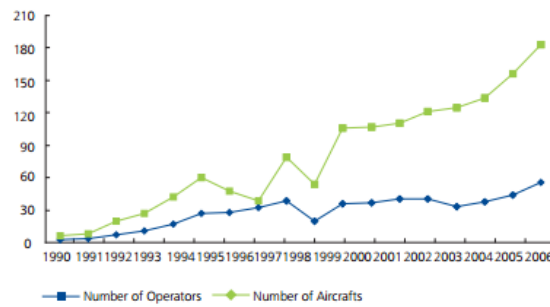


Fig. 7: Growth trends in the non-scheduled operators and aircraft (Airport authority of India)

C. Forecast from past data set:

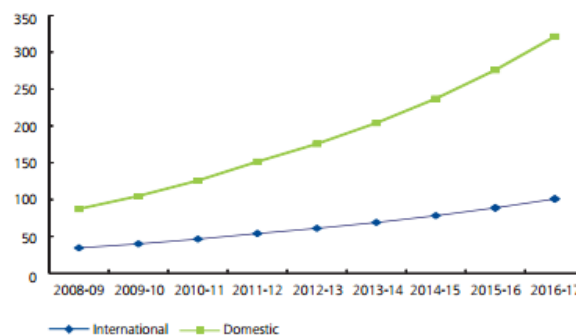


Fig. 8: Passenger growth forecast

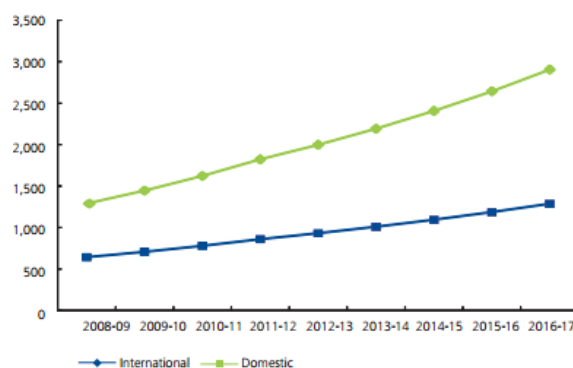


Fig. 9: Cargo growth forecast

V. Key Challenges In Airline Business:

A. Lost Opportunities:

In the dynamic environment that airlines operate in, internal and external factors have a direct impact on cost as well as revenue. Consolidating information from various sources to analyze performance leads to time delays and this often results in opportunities being missed which will reap very high loss (Quantitative Problems Solving methods in Airline Industry).

B. Standardization:

Airlines struggle to measure route performance in a reliable, standard, consistent and timely manner. This can lead to the deployment of the wrong aircraft

and the implementation of schedule changes that compromise profitability.

C. Constant Pressure on Margins:

The airline industry has always been under constant pressure on margins because of their high fixed operating costs. Visibility into the performance of key KPIs such as ASK, RPK, load factor and eventually into route profitability would help carriers optimize their network (Zhe L., *et al.*, 2011) to maximize profits.

VI. Key Outcomes On Analysis Of Large-Scale Flight Data:

A. Key functions that will be provided by the system:

Route Planning decision (Zhe L., et al., 2011) making,Route Profitability analysis,Budgeting,Actual vs. estimated/budgeted analysis.Network (Zhe L., et al., 2011) contribution analysis, Profitability ROI,Daily route performance reports Full Airline statistics, Operational Statistics,Traffic reports (Yang, X.S., 2008) Statistics,Revenue Statistics,Invoice reconciliation & discrepancies,Route Benchmarking

B. Complete Flexibility as to how costs are allocated (Campbell, N.H., 2007):

Flights,Landing,Block hours,Flight hours,ASK/ATK, RPK/RTK,Percentage of Revenue,Pax/Cargo/Mail Revenue,Cycles,Passengers, Additional allocation methods will be used as we are using flight data (APMS Research Overview, 2012; Campbell, N.H., 2007) for analysis in a highly parameterized manner.

C. Warehouse Data usage is used for statistics:

Based on the warehouse data, standard reports will be available with statistical data such as:

Kilometers Flown, Available seat Kilometers (ASK),Cargo Available ton-Kilometers(CATK),Revenue passenger Kilometers(RPKM), RevenueTonKilometers(RTKM), Passenger Load Factor, Overall load factor, Fuel Statistics, Available ton-Kilometers (ATK),Available Freight ton-Kilometers (AFTK),Available Mail Ton-Kilometers (AMTK),Available Dead Load Ton-Kilometers (ADTM),Aircraft utilization

D. Result oriented data-sets:

Besides a detailed breakdown of each routes revenues and costs, a number of Key performance indicators will be presented and benchmarked against pre-defined standards:Break Even Point, Route costs/RPK, RouteRevenues/RPK, Route Revenues/ASK, Ticket Revenues/ASK, Route Costs/ASK, Ticket Revenues/RPK.

VII. Benefits of Using Bigdata Analytics In Aviation:

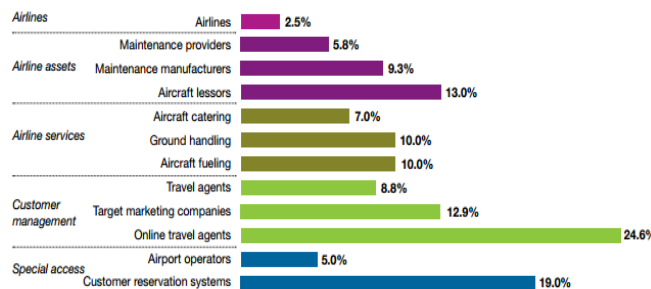


Fig. 10: Estimated ROI on selected airline sectors (IBM Institute for Business Value analysis) based on route profitability analysis

Make corrective decisions adjust non-performing routes (Zhe L., et al., 2011), Locate poorly performing routes, Sector capacity assessments, Future performance, route evaluation, capacity planning etc.Nadditional implementation costs, Nfunctionality modifications additional costs, Easy user graphical interface, for data presentation, Risk Management Strategy support decision making data, Execute what-if scenarios test and tune route parameter changes (Xin-She Yang, 2011), Productivity improvements, as invoice automation.Quick ROI realization, Improve company profitability, Networkmodel (Zhe L., et al., 2011) analysis, as itgenerates most revenue with cost effective use of resources.Assists in Crew scheduling, Aircraft assignment, Fleet

planning/replacement, Yield Management, Flight Scheduling.

Conclusion:

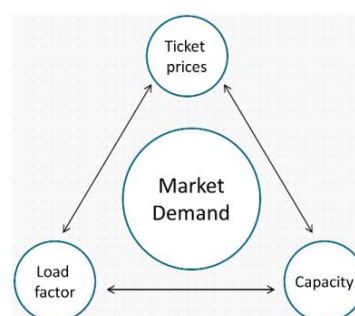


Fig. 11: Interdependency between primary factors of aviation decision making

Computation of the airline route profitability is extremely complex given the revenue and cost structures in the industry. But we can gather some notable business insights by analyzing large-scale data sets of airline data (Campbell, N.H., 2007). In today's world where airlines face stiff competition from other airlines and alternate modes of transportation, a solution that provides flight leg-wise profitability and highlights the strengths and weaknesses of their network would play a vital role in making strategic decisions as all the operational parameters are interdependent as shown in Fig : 11. This information can be used to increase profits by optimizing the aircrafts used and routes flown. In future this can be extended in all the decision making analysis were in hybrid algorithms can be triggered to get the sensible data set for the input to be into the Big data analytical tools with multi-objective criteria.

VIII. Appendix:

A. Future work:

We can gather business insights over several airline datasets (Campbell, N.H., 2007) and by using statistical analysis over necessary data; we could expand the project in mind with future development, by planning to continue with the following modules in the future:

- Fuel Management System
- Marketing Contribution Analysis
- Business Intelligence
- Airline Performance Analyzer
- Demand/Route Forecasting

B. Nomenclature:

ASK = Available Seat Kilometers
 ASM = Available Seat Miles
 ATK = Available Tonne Kilometers
 CASK = Cost per Available Seat Kilometer
 DOC = Direct Operating Cost
 LCC = Low Cost Carriers
 LP = Linear Program
 Nm = Nautical Mile
 RASK = Revenue per Available Seat Kilometer
 RPK = Revenue Passenger Kilometer
 RTK = Revenue Tone Kilometer

REFERENCES

(Eds), *Advances in Human Aspects of Road Transportation*, pp: 777-785). CRC Press.

Xueyan, S., Q. Mingliang and G. Mingang, 2012. Safety Risk Analysis in Flight Operations Quality Assurance. *Systems Engineering Procedia*, 5: 81-86.

Xin-She Yang, 2011. Chaos-Enhanced Firefly Algorithm with Automatic Parameter Tuning, *International Journal of Swarm Intelligence Research*.

Ananda, C.M., and R. Kumar, 2008. Configurable Flight Safety System for Trends and Statistics Analysis Of Avionics Systems-An Embeeded Perspective Of An Efficient Flight Safety Tool. In: *International Conference on Avionics systems*, 22-24 Feb 2008, RCI Hyderabad.

APMS Research Overview, 2012. Retrieved from <http://apms.arc.nasa.gov/index.html>

Yang, X.S., 2008. "Nature-Inspired Metaheuristic Algorithms", Luniver Press.

Campbell, N.H., 2007. The Evolution of Flight Data Analysis. Regional Seminar, Australia. Retrieved from [http://asasi.org/papers/2007/The_Evolution_of_Flight t_Data_Analysis_Neil_Campbell.pdf](http://asasi.org/papers/2007/The_Evolution_of_Flight_Data_Analysis_Neil_Campbell.pdf).

Yemenia Yemen Airways, 2008. "Monthly Traffic Reports" Sana,a.

Zhe L., W.A. Chaovalitwongse, H.C. Huang, E.L. Johnson, 2011. Network model for aircraft routing problem. *Transp Sci.*, 45(1): 109-120.

Kennedy, J., R.C. Eberhart, 1995. "Particle swarm optimization", *IEEE International Conference on Neural Networks*, Piscataway, NJ., pp: 942-1948.

Mitchell, K., B. Sholy, and A.J. Stolzer, 2007. General Aviation Aircraft Flight Operations Quality Assurance: Overcoming the Obstacles. *Aerospace and Electronic Systems Magazine*, IEEE, 22(6): 9-15.

IBM Institute for Business Value analysis, ROIC data from IATA Value Chain Profitability report 2005, Motlyfool "Do These Airlines Pass Buffett's Test" 2010, Seeking Alpha "Expedia Is Insanely Cheap" 2012, Morgan Stanly "Internet 2012 Outlook: Worlds Collide" 2011.

Airport authority of India Movements at Australian airport <http://www.airservicesaustralia.com/publications/reports-and-statistics/movements-at-australian-airports/> Quantitative Problems Solving methods in Airline Industry, A Modeling Methodology Handbook. *International Series in Operation Research & Management Science*.