University of Texas at Arlington

PhD Dissertation in Industrial Engineering

"STATISTICAL META-MODEL FOR AIR TRAFFIC FLOW AND CAPACITY MANAGEMENT BASED ON AIRSPACE OPTIMIZATION-SIMULATION: THE CONTINUOUS CHALLENGE OF THE HUB OF THE AMERICA CONGESTION"

By: Juan Marcos Castillo Guerra

Presented to the Thesis Committee members from the Industrial, Manufacturing and System Engineering Department:

Dr. Erick C. Jones
Dr. Shouyi Wang
Dr. John W. Priest

August 2018

TABLE OF CONTENTS

I. ACKNOWLEDGEMENT 9

II. EXCECUTIVE SUMMARY 10

III. INTRODUCTION 11

- A. PROBLEM DEFINITION 11

IV. LITERATURE REVIEW 15

- A. AIR TRAFFIC FLOW AND CAPACITY MANAGEMENT 15
- B. TRADITIONAL MODELING AND OPTIMIZATION OF AIR TRAFFIC 18
- C. SIMPLE LINEAR REGRESSION (SLR) 20
  - 1) Coefficient of Determination (R-square) 21
- D. MULTIPLE LINEAR REGRESSION (MLR) 21
- E. MODELING INTERACTIONS BETWEEN QUANTITATIVE AND QUALITATIVE PREDICTORS 23
- F. STUDY OF DESIGN OF EXPERIMENTS (DOE) 24
  - 1) Types of Experimental Design 24
- G. ENGINEERING ECONOMICS 27
- H. STATISTICAL PROCESS CONTROL 28

V. BACKGROUND 29

- A. RESEARCH BACKGROUND 29
- B. PRELIMINARY STUDIES 29
  - 1) First DACE 29
  - 2) Second DACE 53
  - 3) Data Exploration and Classification 54

VI. RESEARCH GOALS, OBJECTIVES OR AIMS 56

- A. RESEARCH QUESTIONS 56
- B. OBJECTIVES 57
- C. HYPOTHESIS 57

VII. METHODOLOGY 59

- A. PHASE 1: PLAN 60
  - 1) Define: Methodology Definition 60
  - 2) Measure 62
- B. PHASE 2: PREDICT 63
  - 1) Analyze: 63
  - 2) Design: 64
  - 3) Identify: 66
- C. PHASE 3: PERFORM 66

1) Optimize: 66
2) Verify: 67

VIII. RESULTS 68
A. DATA COLLECTION AND DATA EXPLORATION 68
B. ISHIKAWA ANALYSIS AND LINEAR MODELS 70
1) Approach 70
C. FIRST ISHICAWA MODEL 71
D. SECOND ISHIKAWA MODEL 76
E. FIRST LINEAR MODEL 79
F. THIRD ISHIKAWA MODEL 84
G. SECOND LINEAR MODEL 86
1) Analysis of Variance 87
2) Residual Analysis 89
3) Analysis of the Linear model assumptions for the new model logy. 101
H. OPTIMIZATION AND SPC CHARTS 108
1) Optimization 108
2) Sensitivity Analysis 112
I. ECONOMIC ANALYSIS AND FORECASTING 120
1) Economic Forecasting Analysis based on Scenarios 120
2) Impact of the Flight Miles over the Jet Fuel 123
3) Summary of the Economic Analysis 125
J. RESULTS SUMMARY 127

IX. CONCLUSIONS 128

X. FUTURE STUDIES 129

XI. LIMITATIONS 130

XII. PROJECT PLAN 130

XIII. CONTRIBUTION TO THE BODY OF KNOWLEDGE 132

XIV. APENDIX A 133

XV. REFERENCES 134

LIST OF TABLES

Table 1. Literature Review Chart 19
Table 2. Example of RBD 26
Table 3. DOE Layout 34
Table 4. Interaction Plot Summary 45
Table 5.Interaction Plot Summary 46
Table 6. Interaction Plot Summary 47
Table 7. Model Notation 71
Table 8. Equation 8 explanation 72
Table 9. Equation 9 explanation 73
Table 10. Equation 10 explanation 74
Table 11. Equation 11 explanation 75
Table 12. Equation 12 explanation 75
Table 13. Second Ishikawa Model Notation 78
Table 14. Third Ishikawa a Model Notation 85
Table 15. Residual Analysis plots (Yhat vs Residuals and Y vs Residuals) 89
Table 16. Residual Analysis plots (x_1 vs Residuals and x_2 vs Residuals) 90
Table 17. Residual Analysis plots (x_3 vs Residuals and x_4 vs Residuals) 91
Table 18. Residual Analysis plots (x_5 vs Residuals and x_6 vs Residuals) 92
Table 19. Residual Analysis plots (x_5 vs Residuals and x_6 vs Residuals) 93
Table 20. Residual analysis plots (log $\hat{y}$ vs residuals and logy vs residuals) 102
Table 21 Residual analysis plots (x_1 vs residuals and x_2 vs residuals) 103
Table 22. Residual analysis plots (x_3 vs residuals and x_4 vs residuals) 104
Table 23. Residual analysis plots (x_5 vs residuals and x_6 vs residuals) 105
Table 24. Residual analysis plots (residuals vs x_7) 106
Table 25. Mixed-Integer Linear Programming solver settings in MATLAB 110
Table 26 . Arrivals SPC at 3NM distance between aircraft per route for February 2015 112
Table 27. Arrivals SPC at 3NM distance between aircraft per route for March 2015 113
Table 28. Arrivals SPC at 3NM distance between aircraft per route for June 2016 113
Table 29. Arrivals SPC at 3NM distance between aircraft per route for July 2016 114
Table 30. Arrivals SPC at 3NM distance between aircraft per route for November 2016 114
Table 31. Arrivals SPC at 3NM distance between aircraft per route for November 2016 115
Table 32. Departure SPC at 3NM distance between aircraft per route for February 2015 115

Table 33. Departure SPC at 3NM distance between aircraft per route for March 2015 116
Table 34. Departure SPC at 3NM distance between aircraft per route for June 2016 116
Table 35. Departure SPC at 3NM distance between aircraft per route for July 2016 117
Table 36. Departure SPC at 3NM distance between aircraft per route for November 2016 117
Table 37. Departure SPC at 3NM distance between aircraft per route for November 2016 118
Table 38. Sensitivity Analysis per month of the Arrivals 118
Table 39. Sensitivity Analysis per month of the Departures 118
Table 40. Summary of Economic Scenario 1 121
Table 41. Summary of Economic Scenario 2 122
Table 42. Summary of Economic Scenario 3 122
Table 43. Impact of the economic Scenarios over the Jet Fuel price 123
Table 44. Summary of the Economic Analysis Scenario 1 / Average monthly fuel burn per year 125
Table 45. Summary of the Economic Analysis Scenario 2 / Average monthly fuel burn per year 125
Table 46. Summary of the Economic Analysis Scenario 3 / Average monthly fuel burn per year 125
Table 47. Net Present Value of each scenario per record. 126
Table 48. Summary of the results 127
Table 49. Gantt Chart 131

LIST OF FIGURES

Figure 1 The Six Bank-Hub Operation Schedule of Copa Airlines (52) 12
Figure 2. Forecast Air Cargo Growing. (www.boeing.com 10) 13
Figure 3. Merge Point (53).......... 16
Figure 4. Historical growth of passengers using the Tocumen International Airport. 17
Figure 5. Example of a SLR scatter plot (51) 21
Figure 6. Example of an MLR scatter plot (52) 22
Figure 7. Example of hierarchical Nested Design. 26
Figure 8. Example of Crossed Design. 27
Figure 9. Final Draft to Optimize the Airspace at Panama (22) 31
Figure 10. ANOVA Table for Sequence Actions as a Dependent Variable. 41
Figure 11. ANOVA Table for Number of Conflicts as a Dependent Variable. 41
Figure 12. ANOVA Table for Flight Time as a Dependent Variable. 42
Figure 13. ANOVA Table for Track Mile Distance as a Dependent Variable. 42
Figure 14. ANOVA Table for Fuel Burn as a Dependent Variable. 43
Figure 15. Pairwise Tukey Comparison for Sequence Actions 49
Figure 16. Pairwise Tukey Comparison for Number of Conflicts 50
Figure 17. Pairwise Tukey Comparison for Flight Time 51
Figure 18. Pairwise Tukey Comparison for Track Mile Distance 51
Figure 19. Pairwise Tukey Comparison for Fuel Burn. 51
Figure 20. Behavior of the Operation in Tocumen – Worst Case Scenario- Dec 2015. 54
Figure 21. Sum of Number of Operations during December 2015. 55
Figure 22. Contrast between Number of Departures and Number of arrivals per day. 55
Figure 23. Contrast between Number of Arrivals and Number of Departures per hour. 56
Figure 24. Design for Six Sigma-Research 59
Figure 25. Analysis Process 64
Figure 26. Theme Diagram 66
Figure 27. Approach for Ishikawa model with Linear model testing 70
Figure 28. Cause and Effect Diagram 71
Figure 29. Second Cause and Effect Diagram 76
Figure 30. Analysis of Variance of the Multiple Linear Regression model 80
Figure 31. Analysis of Variance of the Multiple Linear Regression model 81
Figure 32. Analysis of Variance of the Multiple Linear Regression model 82

Figure 33. Correlation Matrix 83
Figure 34. Third Cause and Effect Diagram 84
Figure 35. ANOVA of the Second Linear Model 87
Figure 36. Type 1 and type 3 regression 88
Figure 37. The Normal Probability Plot 93
Figure 38. Correlation Procedure. 94
Figure 39. ANOVA of the transformation of the y to Log y 95
Figure 40. ANOVA of the transformation of the y to 1/y 95
Figure 41. ANOVA of the transformation of the y to lny 96
Figure 42. ANOVA of the transformation of the y to $y^{1/2}$ 97
Figure 43. ANOVA of the transformation of the y to log y 97
Figure 44. ANOVA of the model without x_7*x_1 interaction 98
Figure 45. ANOVA of the model without x_6*x_1 interaction 99
Figure 46. ANOVA of the model without x_5*x_1 interaction 99
Figure 47. ANOVA of the model without x_4*x_1 interaction 100
Figure 48. ANOVA of the model without x_3*x_1 interaction 100
Figure 49. ANOVA of the model without x_2*x_1 interaction 101
Figure 50. Normal Probability Plot 106
Figure 51. Normal Probability Plot 107
Figure 52. Comparison between optimization iterations and the reported daily 111
Figure 53. Comparison between optimization iterations and the reported 111
Figure 54. Jet fuel forecasting over the next 5 years 124
Figure 55. Air Traffic Mix by type of Aircraft 124

LIST OF EQUATIONS

Equation 1: Simple Linear Regression Examplel 20
Equation 2. Example of Multiple Linear Regression Model 22
Equation 3. VIF Formula 23
Equation 4. Example of Modeling Interactions between Quantitative and Qualitative Predictors 23
Equation 5. DoE formulation 35
Equation 6. DoE re-formulation without full interaction effect. 36
Equation 7 Tentative Formulation for cost minimization 67
Equation 8. First Ishikawa Model. 71
Equation 9. Sub-Equation for Regulations 72
Equation 10. Sub-Equation for Stakeholders. 73
Equation 11. Sub-Equation of Investment. 74
Equation 12. Sub-Equation of the Nature Impact on the operation. 74
Equation 13. Sub-Equation of Operation efficiency. 75
Equation 14. Congestion Uncertainty Factor 77
Equation 15. Multiple Regression Model for the Second Ichicagua Model 78
Equation 16. MLR 79
Equation 17. Third Ishikawa Model Formulation 84
Equation 18. ANOVA linear model 86
Equation 19. Mixed Integer Linear Programming to minimize Flight Miles. 109
Equation 20. Hurwicz Criterion 126

I. ACKNOWLEDGEMENT

I cannot believe that this moment has arrived. My family and I took the decision to start a Doctorate in Industrial Engineering as a step of faith, since I had 10 years without formal studies in Engineering. The first acknowledgement is for God, for Jesus my Lord and Savior, He is the main reason I am finishing my doctorate in Industrial Engineering. The second acknowledgement is for my wife Kira Guerra, she is not just my wife, she is my support and my solid ground through the different crisis that I have during this journey. The third acknowledgement is for my father and my mother. It was my mother who took the risk before me to take a PhD in Industrial Engineering, the first female with that degree in my Country. The fourth acknowledgement is for my parents in law, both were for my wife to support her during the longlines of the large times of studies at the University. The Fifth acknowledgment is for my three kids, Sofia, Marcos y Juan Alejandro, they are my inspiration to keep pushing forward and the valley of relaxation when everything is going down. The sixth acknowledgment is for those students and colleges that support me during the journey and the GTAs and RAs that were my partners of craziness during the PhD.

The seventh acknowledge is for my Dissertation board and the professors that took the time to listen to me and explain me the different topics and areas of the Industrial Engineering during my journey. The last acknowledgment is for my friend and mentor Dr. Erick Jones, he is the best Dissertation Advisor and professor ever, almost as great speaker as Dr. Priest.

August 27, 2018

II. EXCECUTIVE SUMMARY

STATISTICAL META-MODEL FOR AIR TRAFFIC FLOW AND CAPACITY MANAGEMENT BASED ON AIRSPACE OPTIMIZATION-SIMULATION: THE CONTINUOUS CHALLENGE OF THE HUB OF THE AMERICAS CONGESTION

Juan Marcos Castillo, PhD.

The University of Texas at Arlington, 2018

Supervising Professor: Erick C. Jones

Panama is only Country on the Americas with a Canal and the hub of Logistics that include the interaction of the Atlantic and the Pacific in a less than a day. This Logistics growth after the Panama Canal Expansion resulted in an overwhelming growth in Aviation. Furthermore, the economic and Logistic growth at Panama is increasing the demand of air transportation and it is creating potential for Air Logistics. Thus, the air traffic congestion is one of the greatest challenges that the Aviation Industry is seeking to address.

The objective of this research is to understand if the Air Traffic Congestion in Panama can be reduced through minimizing the impact of historical congestion variables. In order to meet this objective, three specific criteria are investigated as follows:

- ***Specific Objective #1:*** *Determine which variables are most relevant to minimize Air Traffic Congestion.*
- ***Specific Objective #2:*** *Determine the significance of the variables and their impact on the Air Traffic Congestion.*
- ***Specific Objective #3:*** *Identify the cost effectiveness of the variables on Air Traffic Congestion.*

A statistical Meta-Model that includes Cause and Effect Analysis, Design and Analysis of Computer Experiments, Linear Regression, Mixed Integer Linear Programming, and Engineering Economics is used to address Air Traffic Flow and Capacity Management Uncertainty over the Congestion at the Airspace in Panama.

III. INTRODUCTION

A. Problem Definition

The main problem of the Air Traffic Flow and Capacity Management in Panama is the risk reduction in the Panamanian Air Space Operation during the landing and the departure operations at the Tocumen International Airport without decreasing the productivity of the system. The safety of the passengers is at risk when the congestion in the landing affects the operation of the Tocumen airport, so the airport received penalties from ICAO in 2015, after the audit with a qualification of 36.4%, falling from 85.7% in 2005. Also, FAA changed the Tocumen International Airport from Category 1 to Category 2, in other words, there are more restrictions in the allowance of carriers, new routes and codeshare agreements with USA. (1)

Thus, the problem can be described as the seeking of an optimum number of aircrafts that can land per hour in a time frame of 12 hours where the six bank hub operational model of the Tocumen International and COPA Airlines takes place, while minimizing the Fuel Burn, the number of Interactions between pilots and air traffic controllers, the number of air conflicts, the delays, the flight miles, and the flight time.

The first challenge is operation constraints of the Airport. For example, the two runways work like one runway, since the design of the runways is not parallel. The airport reported a maximum of 40 operations per hour as a service rate (2). The Tocumen International Airport ("PTY" which is the IATA code for the airport) is expanding the gate availability from 34 gates in critical day to 54 gates (2). Consequently, the number of runways is going to be the next bottleneck in the system. In contrast, the standard capacity per airport with a single runway in the USA lies between 35 to 60 operations per hour (3). Consequently, even with the actual runway layout, Tocumen may achieve 60 operations per hour using the actual runway capacity.

The second challenge is the collaboration between stakeholders. There are three main stakeholders, The Autoridad de Aeronautica Civil (AAC), Tocumen International Airport (PTY) and Copa Airlines (COPA). These stakeholders are collaborating together during the six-bank-hub operation in a daily basis, as Figure 1 shows. Therefore, the main stakeholders can manage the air traffic and landing operation to move from 40 operations per hour to 60 operations per hour.

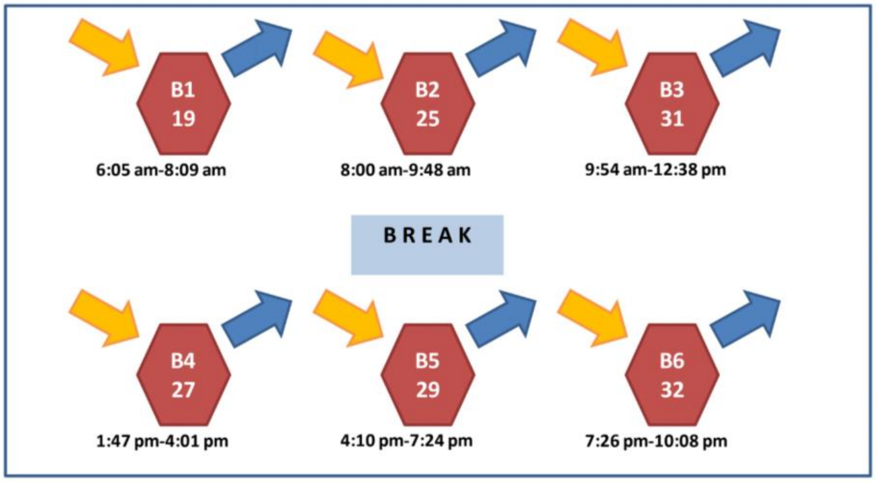


Figure 1 The Six Bank-Hub Operation Schedule of Copa Airlines (52)

Even though, there are three main stakeholders who directly impact on the daily operation, there are more than 20 organizations involved indirectly. These other 20 organizations are trying to find the way to optimize the airspace of Panama. One of the efforts is the implementation of Airport Collaborative Decision Making (ACDM). The second effort is the Air Space Optimization (ASO) which includes the evaluation of the airspace layout.

There is a specific conflict between the air traffic controllers' union, the airlines and the Government. The opinions are different between stakeholders, and those different point of views affect the operation in a daily basis. So, when the representatives of each sector meet there are disputes about the KPIs and the weight each stakeholder address. In other words, the safety is the most important factor to address, however sometimes it conflicts with fuel consumption and controller workload.

The third challenge is the economic impact of the congestion over the country. The aviation performance in Panama produce a des-acceleration of the economic growth. (1). For example, in 2016

American Airlines eliminate a direct fly from Dallas-Forth Worth to Panama, which means at least $10,000 less in daily taxes perceived by the Tocumen International Airport, based on the tax per passenger (2). The lost income described in this example it does not include flight fees from the Civil Aviation Authority and side services from suppliers of fuel, food and cleaning to the aircraft. Consequently, each time the air space is not well managed there are economic consequence such as close of a regular operation or a cancel route that cannot be sustainable. The opportunity cost can be even worse than the closing of an existed operation. (2).

The air transportation is growing fast in the whole world, which means that the evolution of the industry is demanding changes in the air traffic flow and airports efficiency. For example, the Airbus´ Global Market Forecast for 2016-2035 anticipates that air traffic will grow at 4.5 percent annually, in other words, 33,000 new passengers. (12)

In addition, the last 40 years the volume of air logistic growth 7%, so the average in the last 5 years was 3.8% against an average of 3% in the rest of the world. The Forecast of growth in air logistic for the next 20 years is 5-6%. The quote of air logistic is 2% of the world commerce in weight (t) and 35% in (USD). (1)

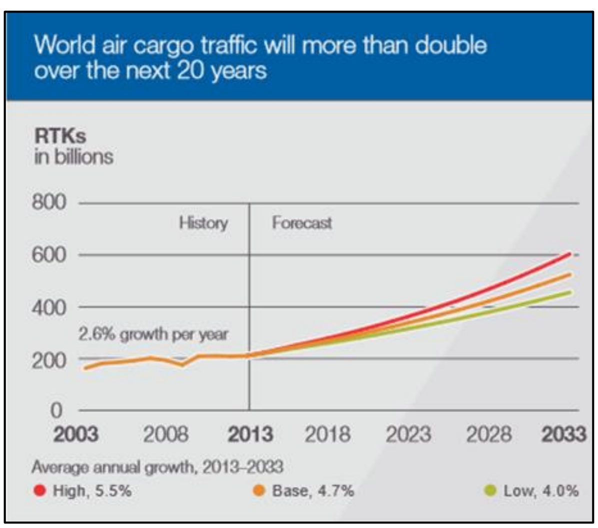


Figure 2. Forecast Air Cargo Growing. (www.boeing.com 10)

As figure 2 shows, the world air cargo traffic is increasing 2.6% per year, and the expected growth in passenger air transportation is "4.9% over the period of 2010-2030" (12)

Consequently, this constant increment in air traffic has not found an adequate expansion of airport facilities and flight assistance (6). These situations presented in an airport, in everyday life, are very unpredictable because those depends on several factors, such as holidays, peak hours, weather conditions, the number of flights and the increment of passengers (7). Those factors have a strong impact in the performance of the airports operations.

There are limitations in capacity at specific ranges of time during the day leading to several issues in the operation, such as large queues in the airspace, congestion in the taxi flow in the ground, waiting lines to depart. Consequently, these air queues are producing delays, cost impact, and more pollution.

Furthermore, the fail in landing create congestion in the air traffic affecting the fuel consumption, which means that if the landing is not safe, the air traffic controller sends the aircraft to make circles in the Balboa Bay until there is another space for landing. That means an increment in the fuel consumption per aircraft and increase the risk of the operation. The frequency founded by empirical interviews with air traffic controllers in 2015 was 7 per day.

The level of the actual congestion in landings procedures to the Tocumen International Airport increases the level of risk in the Air Traffic Management.

Thus, the main problem that we enhance to address is to reduce the actual uncertainty of the decision-making process in real time from the air traffic controllers to enhance a safe Air Traffic Flow and Capacity Management, since the Airspace itself is capable of manage the actual and future demand (2), but it does have issues in the arrival to the main International Airport.

IV. LITERATURE REVIEW

The literature review includes the description of Air Traffic Flow and Capacity Management (ATFCM), Panama Context in relationship with ATFCM, the Traditional modeling and optimization of Air Traffic, and a literature review about Data Analytics, specifically about Cause and Effect Diagram, Statistical Process Control, Linear Regression and Design of Experiments.

A. Air Traffic Flow and Capacity Management

The most popular Air Traffic Flow and Capacity Management methods are the Point Merge from Euro Control, Next Gen from FAA and the Collaborative Actions Renovation of Air Traffic Systems from Japan (4,18, 20, 21, 23).

The capacity and traffic issues in the aviation is a continuous challenge for United States, Europe and Japan, since the 80s. (12) Therefore, the FAA, Euro Control and the Civil Aviation of Japan has different approaches to redesign the airspace and to adequate the airport facilities. Then, the Air Traffic Flow and Capacity Management was born as a research concept to improve the aviation in Europe, USA and Japan. In 2015, when this research started the concept was just Air Traffic Management in the 80s (12), then change to Air Traffic Flow Management (12) in the last decade; nowadays the concept changed to Air Traffic Flow and Capacity Management (11, 12).

The first approach to understand is the Euro Control Model, the Merge Point as shown in figure 3. The concept of the point merge comes from Queueing Theory, specifically it addresses one line for landing merging the different operations in one point.

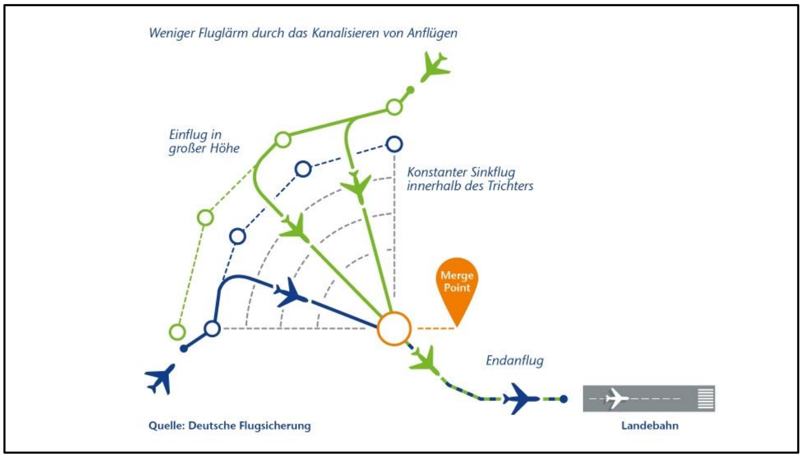


Figure 3. Merge Point (53)

The second approach to understand is the Next Gen, which includes aspects such as ATM planning modes and Collaborative Decision Making (CDM) as a way to increase the Reliability of the System (Decision Support Service, FAA, 2014). As a part of the research, one of the empirical data collections was the visit to the FAA Air Traffic Control System Command Center at Washington D. C. The FAA model includes the Common Situational Awareness that enhance the same information for all parties or stakeholders involved in the Air and Ground operation, aspect that is still absent in the interaction between AAC, Tocumen International Airport, and the Airlines (29).

There are some facts to take in consideration to understand the context of Panama. The main industry of Panama is the transportation ever since the discovery of the Pacific by Balboa. Therefore, the growing of Logistics and Transportation service is 24.3% of the GDP of the Country. (1). This growth is challenging the air traffic system due to the increment of flights. The changes in the Demand are making several issues in the actual air traffic management operation.

Figure 2 shows the growth of the Tocumen International Airport (2), the main airport of Panama. The Passenger movement of Tocumen International Airport has grown steadily during the last few years, with a growing rate of 11.8% during the 2013 comparing with previous years. This study is based on the arrivals and departures operations of the Tocumen International Airport.

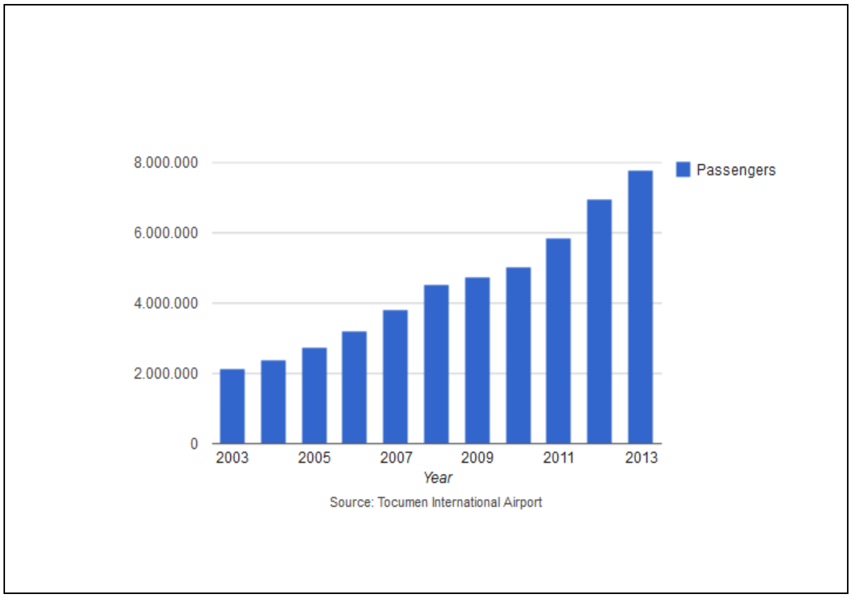


Figure 4. Historical growth of passengers using the Tocumen International Airport.

(Tocumen International Airport 25)

Furthermore, this growing in the air traffic industry in Panama leads to the necessity of new technology and new knowledge to respond. As we stand before, the FAA is working with NextGen in the United States (FAA 20) and the EUROCONTROL is implementing the Point Merge as solution in the air traffic flow management in Europe (EUROCONTROL 23, Ivanescu, et al. 16, Invanescu, et. al. 17, Ozlem, M. 21). Those two are the main sources of ATFCM models to enhance a better aviation in Panama.

Back in 2015, the AAC was taking in consideration the POINT MERGE, which is the concept to merge all the traffic in a single point/line to land in the airport, since the queuing theory assumption is that a single line per server is always more productive than several lines (13). The figure 3, shows an example of the behavior of the air traffic flow using the POINT MERGE approach. EURO CONTROL, the agency that manages the air traffic in Europe, developed this methodology with the following objectives in mind:

In contrast, the FAA has a great influence over Panama aviation, since before 1999 they did share the airspace (4,18, 20, 21, 23). The Federal Aviation Administration (FAA) of the United States of America (USA) and the Autoridad de Aereonatica Civil (AAC) has long relationship, since the USA Government constructs the Panama Canal in 1914, and the USA Army has several bases to around the canal that gave a bound between the two Aviation Agencies. However, after the sign of the Torrijos-Carter Agreement Panama started changing the way they manage all the transportation system, including aviation (28). In the 80s, the ICAO started changing the standards of the AAC from FAA standards to ICAO standards. Consequently, the ICAO was pushing the effort to Air Traffic Flow Management over the Tocumen International Airport and the AAC (preliminary Diagnostic, 2015). Furthermore, the FAA is still a model to the AAC, so the Next Gen methodology includes practical principles that the AAC may apply in the pursuing of a better Air Traffic Management (ATM).

B. Traditional modeling and optimization of Air Traffic

The traditional modeling and optimization of Air traffic includes Optimization Mathematical tools applied to Aviation to achieve a better air traffic. (30). There are several approaches to enhance air space optimization and better flow of the air traffic that include queuing theory (30, 31, 32, 33, 34), simulation models (35, 36, 37, 38), dynamic optimization (30), network analysis (30, 34), scheduling approaches (30, 39), and combinatorial optimization (30). However, the application of Design and Analysis of Computer Experiments as a statistical method that is useful in conducting computer experiments (9, 40, 41, 42, 43, 44) is still not widely applied in Air Traffic Management (ATM) simulation. There are some computer experiments for ATM simulation since 2000, such as the computer experiments for ATM simulation to determine the impact of distributed air-ground traffic management on safety and procedures (45), experiments of the designing for safety: The ´Free Flight ‘air traffic management concept (46), Distributed agent-based air traffic flow management (47), and Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload (48). Even those experiments did not include Statistical Meta Models as Chen, et. al. proposes in 2006. Although, the DACE is a concept from the latest 80s when Sacks, et. al. (49) started discussing about this, the development of DACE is more evident in the 2000s by Deng, et. al. (50).

Consequently, the table 1 shows the summary of the literature review of different traditional approaches of modeling and the optimization of the Air Traffic Flow and Capacity Management.

Table 1. Literature Review Chart

Name	**Model Type**	**Reference**	**Problem**	**Relationship to the model of Panama**
A model of inbound air traffic	Queuing Model	Caccavale, et. al. (2014)	Focus in Arrivals	Airspace Optimization
Data and queuing analysis of a Japanese Air Traffic Flow	Queuing Model	Gwiggner & Nagaoka (2014)	Computational time expensive	Airspace Optimization
Design and Simulation of Airport Congestion Control Algorithms	Monte Carlo Simulation, Dynamic Programming and Queuing Model	Simaiakis & Balakrishnan (2014)	Computational time expensive	Airport Management
A simulation Model for Airport Runway Capacity Estimation	Discrete Evernt Simulation Model	Zou, Cheng & Cheng (2014)	Focus in Runway Optimization	Airport Management
Simulation-based Capacity Analysis for a Future Airport	Discrete event Simulation Model	Mota, et. al. (2014)	Focus in Airport not Airspace	Focus on commercial Aviation, focus in metropolitan region and just one international airport
Capacity and Delay analysis	Discrete event Simulation Model	Celeb, et. al. (2014)	Computational time expensive	Airspace Optimization
Airport Runway Scheduling	Dynamic Programming Model	Bennell, et. al. (2013)	Focus in Runway Optimization	Airport Efficiency
Airport Capacity Constraints impact in future development of air traffic	Simulation Model	Gelhausen, et. al. (2013)	Forecasting/Empirical	Global impact about Aviation
Multi-Objective stochastic Supply Chain Modeling to Evaluate Tradeoffs between Profit and Quality	Stochastic -Supply Chain model	Franca et. al. (2010)	Computational time expensive	There are more than one optimization in the Airspace of Panama. For example, Capacity vs Congestion, or Congestion vs Economic Impact of Congestion Reduction
Probability Airspace Congestion Management	Stochastic model	Zobell et. al. (2010)	Limits the spectrum on the air traffic management	Airspace Optimization
Data Mining of Air Traffic Control Operational Errors	Data Mining model (Attribute Focusing Technique)	Nazeri (2006)	Limits the spectrum on the air traffic management	Airspace Optimization
Accident Risk Assessment for Advance Air Traffic Management	Risk Management model using Markov Chain	Blom, et. al. (2001)	Computational time expensive	Airspace Optimization

A comparison of Aircraft Trajectory-Based and aggregate queue-based control of Airport Taxi Processes	Queuing Model	Lee, et. al. (2010)	Taxi ways	Airport Operations
Modeling Delay Propagation within an airport Network	Queuing Model	Pyrgiotis, et. al. (2013)	Focus in USA	Delay Analysis
Air Traffic of an airport using discrete event simulation method	Discrete event Simulation model	Bevilacqua, et.al. (2012)	Focus just in Aiport	Airport Management
Fast-Time Simulations of Detroit Airport Operations	Mote Carlo Simulation Model	Lee & Balakrishnan (2012)	Focus in Airport Facilities	Airport Efficiency

C. *Simple Linear Regression (SLR)*

SLR is used to determine the relationship between a response variable and a single predictor. The scatter plot of the two variables would contain two-dimensional sample points which can eventually be represented in the form of a true line which helps in predicting the response as a function of the predictor. A basic SLR model would be represented as equation 1.

Equation 1: Simple Linear Regression Examplel

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Where, i = 1, 2,…., n observations

x_i's are fixed (nonrandom) and known variables

Y_i's are the corresponding response observations

β_0 = y-intercept of the true line

β_1 = Slope of the true line

ε_i = Random error

Figure5 shows an example of an SLR plot

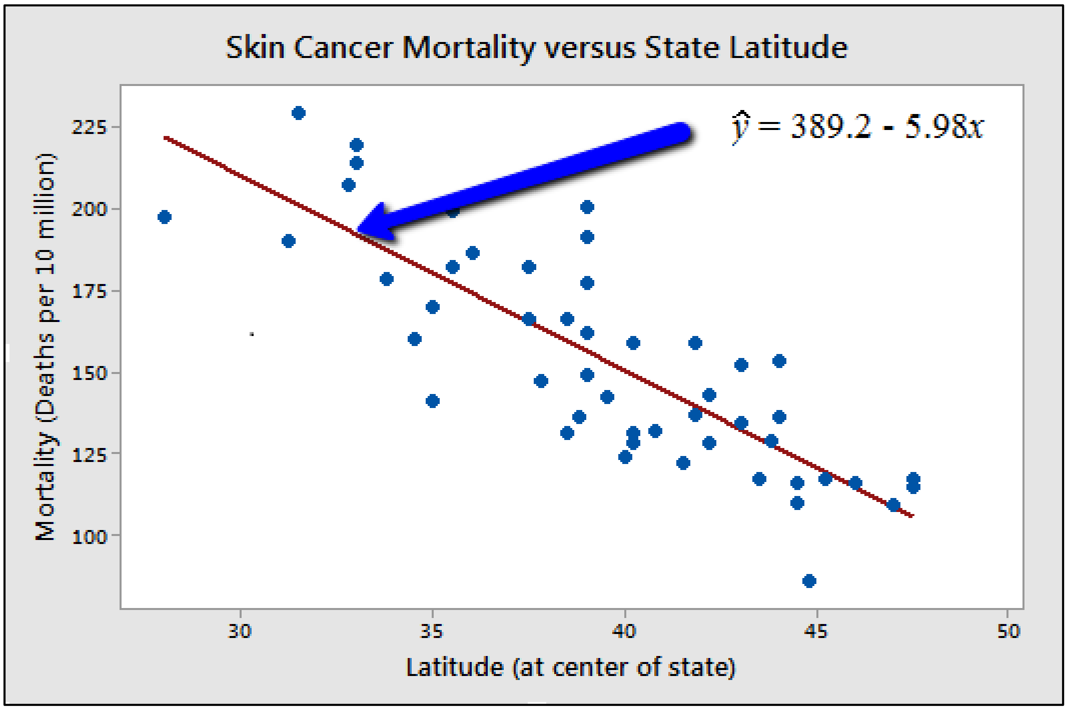


Figure 5. Example of a SLR scatter plot (51)

In this case, the predictor if the latitude and the response is the mortality rate. The points are plotted, and the true line is determined whose equation is given as 389.2 – 5.98x.

If the mortality rate for a latitude of 40 is to be predicted, the procedure is a very simple substitution of 40 in place of x in the equation of true line. Hence, the predicted mortality rate would be 389.2 – (5.98*40) = 150 deaths/10 mn.

1) Coefficient of Determination (R-square)

R^2 is the proportion of the variance in the response that is predictable from the predictors. R^2 value tells how well the model fits the data. Closer the value of R^2 is to 1, better the model fits the data.

D. Multiple Linear Regression (MLR)

MLR is another type of linear regression model with two or more independent variables (predictors). The MLR model with *p-1* predictors is showed at equation 2.

Equation 2. Example of Multiple Linear Regression Model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots.. + + \beta_{p-1} x_{i,\,p-1} + \varepsilon_i$$

Where, i = 1, 2,…., n observations

Y_i = Response when the p-1 predictors are set to $(x_{i1}, x_{i2}, \ldots.\ x_{i,p-1})$

$\mathrm{x_i} = (x_{i1}, x_{i2}, \ldots.\ x_{i,p-1})^T$ = vector of p-1 predictors

$\beta = (\beta_0, \beta_1, \ldots.\ \beta_{,p-1})^T$ = vector of unknown parameters

ε_i = Random error

An example of a MLR model with two predictors is shown at figure 6.

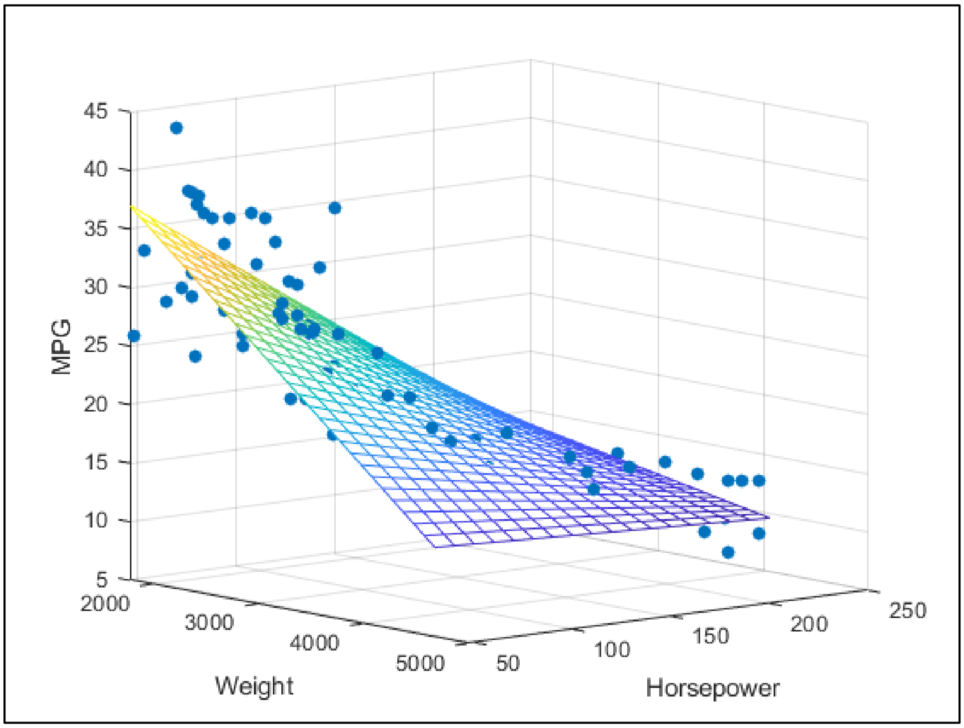


Figure 6. Example of an MLR scatter plot (52)

Some MLR models may have huge number of predictors. While it is perfectly fine to have several predictors in a model, one of the most important things to analyze is the correlation between predictors. A high correlation between the response and the predictors is desirable. But on the other hand, if two or more predictors are highly correlated with each other, one predictor can be used to predict the other with a substantial degree of accuracy. This phenomenon in statistics is called multicollinearity. It is an undesirable property of predictors which in practical applications would require huge calculations and would consume time. Presence of multicollinearity means that there are one or more redundant predictor variables which does not explain the model.

Variance Inflation Factor (VIF) quantifies the severity of multicollinearity. For practical applications, models with a VIF greater than 5 would be considered to have highly correlated predictors. VIF can be calculated by equation 3.

Equation 3. VIF Formula

$$VIF_i = \frac{1}{1-R_i^2}$$

Where R_k^2 is the coefficient of determination of the MLR equation with X_i on the left hand side, and all other predictors on the right hand side (53)

E. *Modeling Interactions between Quantitative and Qualitative Predictors*

The models with qualitative and quantitative predictors can be evaluated using a formulation with interactions between the qualitative variable and the quantitative variable. The meaning of the regression coefficients in response function can best be understood by examining the nature of this function. For Example, the equation 4 shows the regression model with one quantitative variable x_1 and one qualitative variable x_2. Also, this example equation includes the interaction effect between the qualitative variable and the quantitative variable as x_1x_2.(86)

Equation 4. Example of Modeling Interactions between Quantitative and Qualitative Predictors

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$$

An advantage of this type of model with indicator variable is that one regression run will yield both fitted regressions. Another advantage is that tests for comparing the regression functions for the different classes of the qualitative variable can be clearly seen to involve tests of regression coefficients in a general linear model.

F. Study of Design of Experiments (DoE)

Generally, any process under consideration is affected by one or more factors. These input factors may influence the output in any way. Hence, there is a need to understand the relationships between these factors and the output. The first step towards achieving that goal is to create a dataset with response to all the other input factors. Secondly, the interactions between multiple factors must be considered at each level, for example, if there are two factor is just one interaction between them, but if there are three factors there are three interactions with two factors and one full interaction with all three factors. This is where Design of Experiments (DoE) proves to be useful. By manipulating multiple inputs at the same time, DoE can identify important interactions that may be missed when experimenting with one factor at a time (54). This chapter explores the classification and applications of DoE and also provides a framework of which type of DoE must be used in certain cases.

1) Types of Experimental Design

a) Factorial Design

When there are several factors in an experiment, a factorial design can be used (55) In this type of design, factors are varied simultaneously and all possible combinations of the levels are investigated. The two classifications of factorial design are discussed below.

b) Full Factorial Design

Full factorial design includes all possible combinations of all the levels of all factors. In this way, it is impossible to miss any interactions. This type of design requires at least one observation for every combination of factors and levels.

c) Fractional Factorial Design

In a fractional factorial design, not all possible treatments are considered. Taguchi and Latin Squares are some examples of fractional factorial design models.

d) Orthogonal Design

Two vectors are orthogonal when the sum of the products of corresponding elements is zero.

For instance, let a = [2 3 5 0]’ and b = [-4 1 1 4]’

$$\sum_{i=1}^{4} a_i b_i = -8 + 3 + 5 + 0 = 0$$

If vectors are orthogonal, they are independent and does not affect other factors.

e) Randomized Block Design (RBD)

RBD is the design where similar factors are grouped/arranged together to minimize the number of experiments. Using RBD can decrease the unexplained variability which is otherwise known as the Sum of Squared Errors (SSE) of the model. (56)

Let L_n be the number of levels in the n^{th} factor. The number of runs for the respective multi-factor models can be calculated as the table 2 shows.

Table 2. Example of RBD

Number of factors	Number of runs
2	$L_1 * L_2$
3	$L_1 * L_2 * L_3$
k	$L_1 * L_2 \ldots.* L_k$

Two further types of Randomized Block Design are discussed below:

(1) Complete Randomized Block Design (CRBD)

It is the type of design in which every block receives all the treatments. The defining feature of the CRBD is that each block sees each treatment exactly once.

(2) Incomplete Randomized Block Design (IRBD)

In the IRBD, each block receives only some of the selected treatments and not all the treatments.

f) ***Nested Design***

Also called as hierarchical design, nested design is mainly used in experiments in which there is an interest in a set of treatments and the experimental units are sub-sampled. An example of a nested design of a biologist collecting 3 seeds from 3 superior trees in each of three forests A, B and C would look something like this:

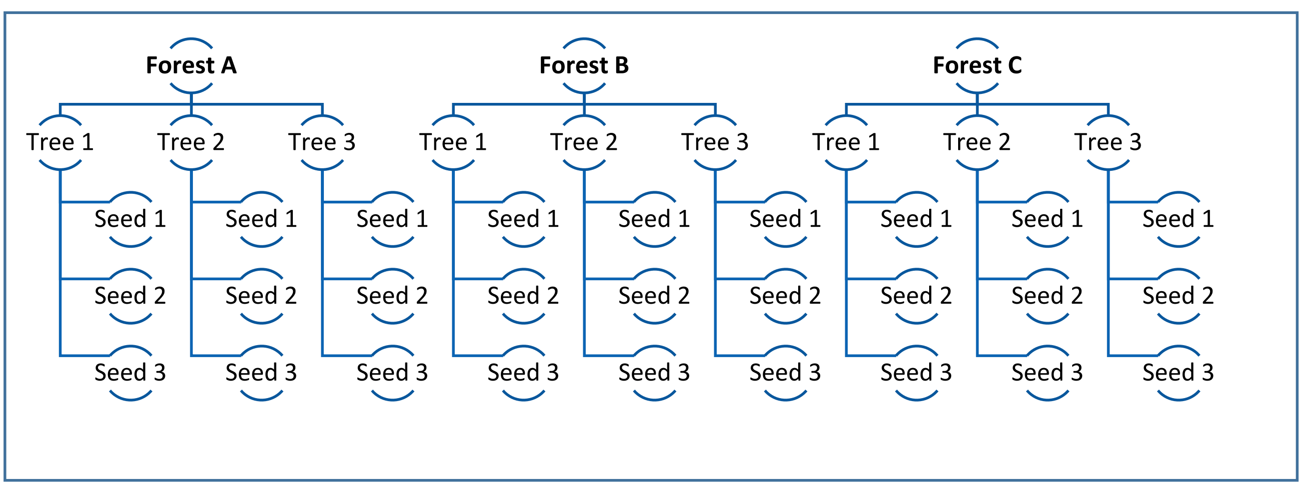


Figure 7. Example of hierarchical Nested Design.

g) Crossed Design

This type of design is faintly similar to nested design. In a crossed design, each level of one factor occurs in combination with each level of another factor. For example, a crossed factor design of 2 machines which can be serviced by any of the three engineers will look somewhat like:

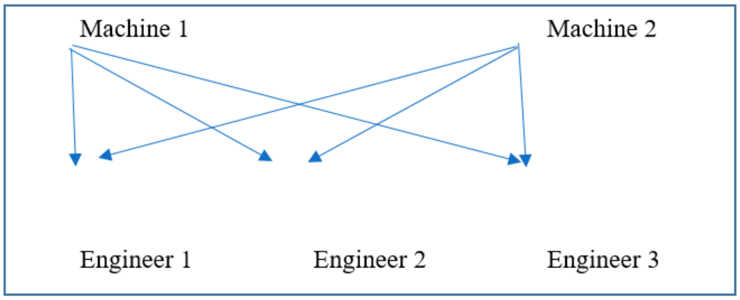


Figure 8. Example of Crossed Design.

h) Latin Hypercube Design

Latin Hypercube design is a random design in which the model parameter values for the experiments are assigned on the basis of a random process (57). This method is mostly used in computer experiments.

i) Design and Analysis Computer Experiments (DACE)

In this method, codes are used to design experiments with no random error. A computer model is used to make inferences about the system it replicates. For example, climate models are often used because experimentation on an earth sized object is impossible (58)

G. Engineering Economics

The study of the time value of money is the concern of the Engineers while performing projects that include project of investments and cost evaluation over time. The Engineering Economics use statistics, mathematics and cost accounting (60) to establish a logical and analytical framework that seeks to find solutions of technical problems viability. (60). Typically, the Engineering Economics convers the analysis of inflexion point, tipping points, depreciation and valuation, capital budget, taxes, interest and money, sensitivity analysis and reliability. The basic indicators used to analyze time value of money are Net Present Value, Future Worth and IRR.

H. Statistical Process Control

The statistical process control is a methodology for monitoring a process to identify special causes of variation and signal the need to take corrective action when appropriate (62). Some of the most utilized tools are the Pareto Diagram, the Cause-Effect Diagram, the Check Sheets, the Process Flow Diagram, the Scatter Diagram, the Histogram and the Control Charts.

For the purpose of the study we use the tools of Cause and Effect Diagrams to understand the factors that affect directly and indirectly the air traffic congestion. Also, we use the Control Charts to evaluate the performance of the optimization as a control, specifically the Moving Average Charts for individual values, which is a special type of control charts. There are other control charts, such as Exponential Weighted Moving Average, non-acceptance limits, control charts for nonconformities. Usually the control charts are to measure the variability of the system, and find ways to keep it in control. (77, 81)

V. BACKGROUND

A. Research Background

The original challenge presented by "The Autoridad de Aereonuatica Civil" (AAC), which is the Government Agency to manage the Aviation in Panama, was to optimize the airspace to allow better flow of the aircrafts.

In 2015, the AAC was looking forward to implement new methodologies to expand the capacity of the actual airspace. The first intention was to apply the Euro Control model, which is the agency that manage the aviation in Europe and which includes a research center. The methodology of airspace optimization of Euro Control is the Point Merge. They found a way to merge the air routes with a standard approximation to minimize fuel burn consumption, conflict and numbers of interactions between pilot and air traffic controller; also, the methodology aims to maximize accuracy in the flight plan.

Consequently, the AAC worked with COPA Airlines, which is the main Airline in Panama and the Leader in Latin America, to find a solution using the Point Merge approach. COPA constructed a simulation model in the simulation Software from Jepessen named TAAM (26). They constructed 5 models. The last optimization version was the preferred by COPA Airlines, but the pilot's union refuse it. The summary of the timeline is at the appendix A.

B. Preliminary Studies

1) First DACE

a) General Description

There is a preliminary study called "Design and Analysis of Computer Experiments based on a Simulation Model of Air Traffic Flow Optimization in Panama" from summer 2016 that had

3 Factors and 5 Key Performance Indicators. This study was made it between UT Arlington and The Universidad Tecnológica de Panamá.

Copa Airlines and AAC were trying to use a combination of the FAA approach and EURO CONTROL approach combining NEXTGEN and the Point Merge in their air traffic flow management. In addition, the Autoridad de Aereonautica Civil (AAC), COPA Airlines and the Tocumen International Airport (Tocumen) are working in a continuous collaboration between FAA and Panama in order to improve the actual system using COLLABORATIVE DECISION MAKING (CDM), specifically Airport CDM or A-CDM. The objective of the simulation model was to select an air traffic alternative that would be able to improve the actual situation. In other words, COPA and AAC were looking to minimize the numbers of conflicts, the number of sequence actions, the flight time, the track flight distance and the fuel burn.

A preliminary study was conducted based on the COPA and AAC simulation model to understand the factors that can affect the air traffic flow in Panama. The objective of the simulation model is to improve the air operation efficiency. The software used in the simulation was Total Airspace and Airport Modeler (TAAM) by Jeppesen (26). It is important to mention that they create 5 scenarios with 5 different airspace layouts. These simulation models are based on some rules in terms of airport description and geographical location of the airport, the layout of the airport, the itinerary of the flights and the airways. The simulations were all based on the old layout with 32 fixed gates. However, there is an expansion of the airport with more gates, taxiways and runways. However, for the purpose of the preliminary study, those factors were not part of the simulation itself.

The main experiments conducted by COPA and AAC were five models of air traffic flow (22), Actual situation, an Alternative based on Vectoring, a Point Merge version 1, a Point Merge version 2, and the Final Draft

The Final Draft, as figure 9 shows, was made it by COPA as a mix of the testing models. However, this experiment was not constructed with an experimental design and it does not include the weather seasoning.

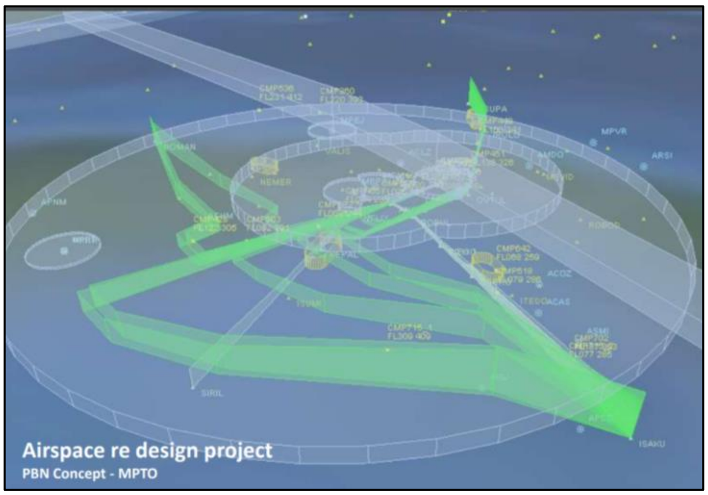


Figure 9. Final Draft to Optimize the Airspace at Panama (22)

Based on COPA analysis, there are other factors to take in consideration, and important for future evaluation. Those factors, that needs to be considered, are the wind, the weather events, the aircraft weight, domestic flights, over flights, aircraft speed and Altitude. In addition, the demand seasons, the Air Traffic Rules and the Ground Traffic Rules, which are important when Point Merge is considered.

The most important KPIs from COPA standpoint are the Fuel Burn, the Track mile distance and the Flight time. On the other hand, the most important KPI´s for the AAC are the number of Sequencing Actions and the airborne conflicts. Therefore, the goal of both organizations is to optimize the five KPI´s.

In Contrast, the majority of this studies using simulation models for air traffic flow management does not use any methodology to understand the impact of the factors at certain levels. The preliminary study includes Design and Analysis of Computer Experiments in order to understand how the factors at certain levels can impact the Key performance indicators or response variables. So, the objective of this study is to analyze how the Itinerary by Season (Low Season of Demand or High Season of Demand), the Ground Traffic Rules and the Air Traffic Rules can affect the Air Traffic Management KPI´s.

b) Model Definition

(1) Factors

(a) Itinerary

The Itinerary is a data base which include Type of aircraft, License plate, Origin, Destiny, Departure Time and Arrival Time. Each row of the data base is a flight.

(b) Ground traffic rules

The Ground traffic rules is a time distance between aircraft during the arrival, which is between 1 minute to 2 minutes.

(c) Air traffic rules

The Air traffic rule is the distance in nautical miles between aircraft during the approximation to the airport which range lies between 3 NM as a minimum and 10 NM.

Consequently, the dataset of itinerary is a factor with two levels (high season data set, low season data set), the Ground traffic rules is a factor with three levels (1 minute, 1.5 minutes and 2 minutes) and the Air traffic rules is a factor with four levels (3NM, 5 NM, 7 NM, and 10 NM)

(2) Response Variables

When Copa Airlines run a simulation, they obtain five output as response variables for each model. The Key performance indicators for the Air Traffic Management in Panama are the Sequencing actions (number of interactions per day), the Airborne conflicts (number of conflicts per day), the Flight time (hours per day), the Track mile distance (Nautical Miles per day) and the Fuel Burn (gallon per day).

(3) Experimental Design and Linear Model

The first experiment conducted by COPA used a fixed ground traffic rule, a fixed air traffic rule, and a data set from the high season.

Therefore, this experimental design is a Three Factor Complete Factorial Experiments. The table 1 shows the coded layout of the experiment. The following are the factors description with their levels:

Factor 1: Itinerary (1-high season, 2-low season)

Factor 2: Ground Traffic rules (1- 1 min, 2-1.5 min., 3-2 min.)

Factor 3: Air Traffic rules (1- 3NM, 2 - 5 NM, 3 - 7NM, 4- 10NM)

Table 3 shows the coded layout that was used to conduct the 24 experiments. These experiments were conducted directly in the COPA office, since the limited license in place that they have. In addition, there is one replication made in each experiment.

In order to achieve flexibility and efficiency, it is better to select the full factorial design to run the experiments. This kind of design was originally used in Design of

Experiments for physical experiments, but it is suitable to apply in Computer experiments as well. (Chen et. al., 4). Table 3. Layout coded of the Three Factor Complete Factorial Design.

Table 3. DOE Layout

Factor 1	Factor 2	Factor 3
1	1	1
1	1	2
1	1	3
1	1	4
1	2	1
1	2	2
1	2	3
1	2	4
1	3	1
1	3	2
1	3	3
1	3	4
2	1	1
2	1	2
2	1	3
2	1	4
2	2	1
2	2	2
2	2	3
2	2	4
2	3	1
2	3	2
2	3	3
2	3	4

The linear model formulation per each response variable is as follows:

Equation 5. DoE formulation

$$Y_{ijkt} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkt}$$

for i = 1,…, a, j=1, …., b, k=1, …., c, and t=1, ….., r.

Where ε_{ijkt} are iid $N(0,\sigma^2)$

Y_{ijkt} = t-th response observed for trt (i,j,k).

$\mu_{...}$ = is the overall mean.

α_i = is the effect on the response due to the fact that ith level of factor 1.

β_j = is the effect on the response due to the fact that jth level of factor 2.

γ_k = is the effect on the response due to the fact that kth level of factor 3.

$(\alpha\beta)_{ij}$ = is the interaction effect in ith and jth of factors 1 and 2.

$(\alpha\gamma)_{ik}$ = is the interaction effect in the ith and kth of factors 1 and 3.

$(\beta\gamma)_{jk}$ = is the interaction effect in the jth and kth of factors 2 and 3.

$(\alpha\beta\gamma)_{ijk}$ = is the interaction effect in ith, jth and kth of factors 1, 2 and 3.

COPA mentioned that there is a way to obtain the probabilistic data, but the analyst asked us to run the model without stochastic data, since they made the previous experiments using deterministic output. Consequently, the mathematical model is going to suffer a modification, since there is not going to consider any interaction effect with the three factors in conjunction.

The linear model formulation per each response variable is going to be as follow:

Equation 6. DoE re-formulation without full interaction effect.

$$Y_{ijkt} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijkt}$$

for i = 1,…, a, j=1, …., b, k=1, …., c, and t=1, ….., r.

Where ε_{ijkt} are iid $N(0,\sigma^2)$

Y_{ijkt} = t-th response observed for trt (i,j,k).

$\mu_{...}$ = is the overall mean.

α_i = is the effect on the response due to the fact that ith level of factor 1.

β_j = is the effect on the response due to the fact that jth level of factor 2.

γ_k = is the effect on the response due to the fact that kth level of factor 3.

$(\alpha\beta)_{ij}$ = is the interaction effect in ith and jth of factors 1 and 2.

$(\alpha\gamma)_{ik}$ = is the interaction effect in the ith and kth of factors 1 and 3.

$(\beta\gamma)_{jk}$ = is the interaction effect in the jth and kth of factors 2 and 3.

c) The Simulation Experiments

(1) The Simulation Model

(a) The Simulation Software

The Total Airspace and Airport Modeler (TAAM) is a fast-time gate-to-gate simulator of airport and airspace operations. This software can simulate 4D and 3D. TAAM enables the analyst to identify the system benefits of such changes in the airport layout for gates, taxi ways and runways. In addition, other air space requirements.

Some of the features are the 3D multi-color models of airports and aircrafts. 4D full airspace & flight profile calculations, detailed ground functionality, detailed airside functionality, a flexible rule base to accommodate different modelling requirements, statistical data generated in a wide variety of report forms, direct output to spreadsheet and database tools for further in-depth analysis.

(b) The Simulation Model

The simulation model consists in set the static files (in our case the itineraries), the parameter setting, and the rules (Air Traffic Rules and Ground Traffic Rules for this experiment)

There exist other parameters that most keep standard, such as Airport layout (32 gates), 2 runways, 3 taxi ways, and the airport geolocation (COPA Airlines 7). The Airport with the specifications most be drawing in AutoCAD and uploaded in the software. In addition, there is an airspace design, so the regions of the airspace and the air ways must be drawing.

(c) The Itinerary Samples

The department of Operation Efficiency of COPA analyzed the air traffic flow from 1st of January to July 7th and took to days one from the high season and another from the low season to obtain the sample. Then, COPA took, using another software called AIMS, the itinerary for each day. However, the procedure says that it is required to take at minimum of 3 days. This is necessary since they need to take from the 05:00 a.m. of the actual day to the 05:00 a.m. of the day after the actual day. This is necessary to keep the continuity of the simulation in terms of time. As an explanation, COPA takes the 05:00 as a reference, since is the hour zone of Panama based on the Greenwich Meridian.

As we mention before, the itineraries contain the type of aircraft, the license plate, the origin, the destiny, the departure time and the arrival time. This information is per flight.

(d) Simulation Output

In order to obtain the output, it is necessary to use the sample itinerary which is part of the input information. Then, it is important to change some of the Air Traffic Rules and the Ground Traffic Rules in 2 windows and in the map of the air space. The areas of the map changed are for approximation to the Tocumen Airport. In other words, this rules affect in some ways the departure, depending if the runways have not conflict in the departure, and this rules affect all the arrival queue in the air space of Panama.

Therefore, the simulation is going to run per 5 days, just to check any outlier and maintain the continuity. However, the model has a rule to stop at some point (which is 05:00 a.m. as we mention before) to record the information for the main in study. There is another rule, the clock must stop at 05:00 a.m. the next day to stop the recording, this process recording is manually. After the model stop, it is necessary to run the three different queries; two of them were customized by COPA for the previous analysis.

In addition, the output of time is in seconds and the fuel consumption is Kilograms, so it is necessary to convert those. The flight time is converted in hours of flight and the fuel burn is change in gallons.

(2) Statistical Analysis

For the statistical analysis, it is presented each response variable separately in order to analyze the effect of each factor which their levels. The objective is to know how the factors and the levels affect each response variable separately. For this analysis the software used is SAS. Therefore, the analysis shows the ANOVA table with the main factors and the interaction effects. However, we do not include the full interaction with the three factors since we don’t have replications. The model does not include the full interaction effect between the three factors, since the simulation model is deterministic. Furthermore, the interaction plot and the “Tukey” comparison per each model is presented with the followed discussion.

(3) Analysis of Variance

The analysis of Variance conducted present the results per each response variable. Therefore, we are working with five different models and 5 different analyses. Significance level used for the ANOVA is 0.1 as an alpha value.

Consequently, the hypothesis for the linear model stands as:

H_o: There is no difference in the treatments/full model is not statistically significant.

H_a: at least two treatments are different/full model is statistically significant.

So, the p-value must be less than the alpha value 0.1 to reject Ho. Then the model is statistically significant.

The hypothesis analyzed based on the ANOVA tables for interaction effects are:

H_{012}: factor 1 and factor 2 interaction is negligible.

H_{112}: factor 1 and factor 2 interaction is not negligible.

H_{013}: factor 1 and factor 3 interaction is negligible.

H_{113}: factor 1 and factor 3 interaction is not negligible.

H_{023}: factor 2 and factor 3 interaction is negligible.

H_{123}: factor 2 and factor 3 interaction is not negligible.

The decision rule for those hypotheses is that the p-value must be less than the alpha value 0.1 to reject Ho.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	22519.16667	1324.65686	195.44	<.0001
Error	6	40.66667	6.77778		
Corrected Total	23	22559.83333			

R-Square	Coeff Var	Root MSE	Sact Mean
0.998197	1.577032	2.603417	165.0833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Iti	1	7561.50000	7561.50000	1115.63	<.0001
GTRules	2	4.33333	2.16667	0.32	0.7380
ATRules	3	14345.83333	4781.94444	705.53	<.0001
Iti*GTRules	2	37.00000	18.50000	2.73	0.1436
Iti*ATRules	3	517.83333	172.61111	25.47	0.0008
GTRules*ATRules	6	52.66667	8.77778	1.30	0.3808

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Iti	1	7561.50000	7561.50000	1115.63	<.0001
GTRules	2	4.33333	2.16667	0.32	0.7380
ATRules	3	14345.83333	4781.94444	705.53	<.0001
Iti*GTRules	2	37.00000	18.50000	2.73	0.1436
Iti*ATRules	3	517.83333	172.61111	25.47	0.0008
GTRules*ATRules	6	52.66667	8.77778	1.30	0.3808

Figure 10. ANOVA Table for Sequence Actions as a Dependent Variable.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	3835.833333	225.637255	26.72	0.0003
Error	6	50.666667	8.444444		
Corrected Total	23	3886.500000			

R-Square	Coeff Var	Root MSE	Cflict Mean
0.986963	7.597209	2.905933	38.25000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Iti	1	864.000000	864.000000	102.32	<.0001
GTRules	2	27.000000	13.500000	1.60	0.2776
ATRules	3	2723.500000	907.833333	107.51	<.0001
Iti*GTRules	2	3.000000	1.500000	0.18	0.8415
Iti*ATRules	3	192.333333	64.111111	7.59	0.0182
GTRules*ATRules	6	26.000000	4.333333	0.51	0.7815

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Iti	1	864.000000	864.000000	102.32	<.0001
GTRules	2	27.000000	13.500000	1.60	0.2776
ATRules	3	2723.500000	907.833333	107.51	<.0001
Iti*GTRules	2	3.000000	1.500000	0.18	0.8415
Iti*ATRules	3	192.333333	64.111111	7.59	0.0182
GTRules*ATRules	6	26.000000	4.333333	0.51	0.7815

Figure 11. ANOVA Table for Number of Conflicts as a Dependent Variable.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	216612.1440	12741.8908	3714.94	<.0001
Error	6	20.5794	3.4299		
Corrected Total	23	216632.7234			

R-Square	Coeff Var	Root MSE	FTime Mean
0.999905	0.151742	1.851999	1220.496

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Iti	1	181381.1840	181381.1840	52882.3	<.0001
GTRules	2	2.0598	1.0299	0.30	0.7511
ATRules	3	32100.5313	10700.1771	3119.67	<.0001
Iti*GTRules	2	1.9931	0.9965	0.29	0.7578
Iti*ATRules	3	3102.3106	1034.1035	301.50	<.0001
GTRules*ATRules	6	24.0652	4.0109	1.17	0.4271

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Iti	1	181381.1840	181381.1840	52882.3	<.0001
GTRules	2	2.0598	1.0299	0.30	0.7511
ATRules	3	32100.5313	10700.1771	3119.67	<.0001
Iti*GTRules	2	1.9931	0.9965	0.29	0.7578
Iti*ATRules	3	3102.3106	1034.1035	301.50	<.0001
GTRules*ATRules	6	24.0652	4.0109	1.17	0.4271

Figure 12. ANOVA Table for Flight Time as a Dependent Variable.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	34949883849	2055875521	3248.03	<.0001
Error	6	3797769	632961		
Corrected Total	23	34953681618			

R-Square	Coeff Var	Root MSE	TMDist Mean
0.999891	0.159594	795.5888	498508.5

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Iti	1	29912584945	29912584945	47258.1	<.0001
GTRules	2	474151	237076	0.37	0.7026
ATRules	3	4560621424	1520207141	2401.74	<.0001
Iti*GTRules	2	287703	143852	0.23	0.8033
Iti*ATRules	3	471737542	157245847	248.43	<.0001
GTRules*ATRules	6	4178084	696347	1.10	0.4554

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Iti	1	29912584945	29912584945	47258.1	<.0001
GTRules	2	474151	237076	0.37	0.7026
ATRules	3	4560621424	1520207141	2401.74	<.0001
Iti*GTRules	2	287703	143852	0.23	0.8033
Iti*ATRules	3	471737542	157245847	248.43	<.0001
GTRules*ATRules	6	4178084	696347	1.10	0.4554

Figure 13. ANOVA Table for Track Mile Distance as a Dependent Variable.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	1.6263486E13	956675650912	2099.05	<.0001
Error	6	2734594642	455765773.67		
Corrected Total	23	1.6266221E13			

R-Square	Coeff Var	Root MSE	FuelB Mean
0.999832	0.177819	21348.67	12005824

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Iti	1	1.3020771E13	1.3020771E13	28569.0	<.0001
GTRules	2	206192953.04	103096476.52	0.23	0.8041
ATRules	3	2.9608878E12	986962604521	2165.50	<.0001
Iti*GTRules	2	291163925.63	145581962.82	0.32	0.7382
Iti*ATRules	3	278828016005	92942672002	203.93	<.0001
GTRules*ATRules	6	2502124108	417020684.67	0.91	0.5416

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Iti	1	1.3020771E13	1.3020771E13	28569.0	<.0001
GTRules	2	206192953.04	103096476.52	0.23	0.8041
ATRules	3	2.9608878E12	986962604521	2165.50	<.0001
Iti*GTRules	2	291163925.63	145581962.82	0.32	0.7382
Iti*ATRules	3	278828016005	92942672002	203.93	<.0001
GTRules*ATRules	6	2502124108	417020684.67	0.91	0.5416

Figure 14. ANOVA Table for Fuel Burn as a Dependent Variable.

The evaluation of the 5 models using the ANOVA concludes that, at 0.1 level of significance, all the linear models are statistically significant. So, we reject H_0 in our first hypothesis analysis. However, the interactions between factor 2 and the others factors are greater than 0.1 as an alpha value, which means we fail to reject Ho in the interaction hypothesis. In contrast, the interaction between factor 1 and factor 3 is significant and we can reject H_0.

The hypothesis analyzed based on the ANOVA table for the main effects is:

H_{02}: main effect for factor 2 is negligible.

H_{12}: main effect for factor 2 is not negligible.

The decision rule for this hypothesis is that the p-value must be less than the alpha value of 0.1 to reject Ho.

The Ground Traffic Rules (GT Rules) or factor 2 is not significant at 0.1 level, since the three-way ANOVA shows that the p-value of GT Rules (factor 2). So, we fail to reject Ho and the main effect of factor 2 negligible. There is not necessity to test the other main effects since the interaction between factor 1 and 3 is not negligible. Figure 3 shows that the p-value of GT Rules is 0.738 when the number of sequence actions as a response variable. Figure 4 shows that the p-value of GT Rules is 0.2776 when the response variable is the number of conflicts. Figure 5 shows that the p-value of GT Rules is 0.7511 when the response variable is the flight time. Figure 6 presents that the p-value of GT Rules is 0.7026 when the track mile distance is the response variable. Figure 7 presents that the p-value of GT Rules is 0.7382 when the fuel burn is the response variable. In other words, the GT Rules has not significant effect in the dependent variables or Air Traffic KPIs.

As it is mentioned before, the “Iti” or Itinerary and the AT Rules or Air Traffic Rules are statistically significant at 0.1 level. So, AT Rules and Itinerary have an effect over the Air Traffic KPIs. Therefore, the following analysis of interaction plots and Tukey pairwise comparison is going to be considering only between those two factors.

(4) Interaction Plots

The objective of the interaction plots is to understand how the interaction can affect each response variable.

Table 4. Interaction Plot Summary

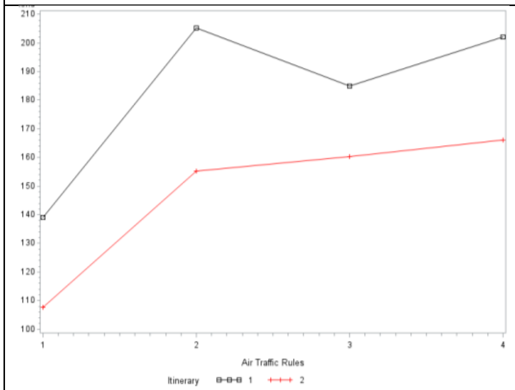
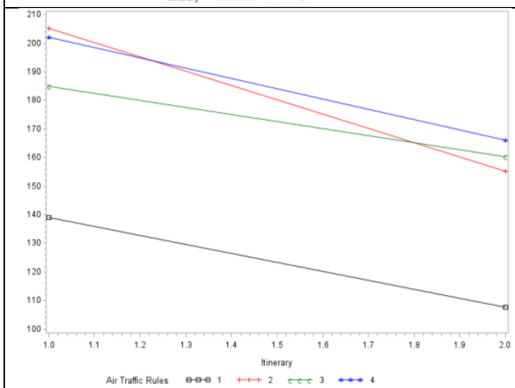
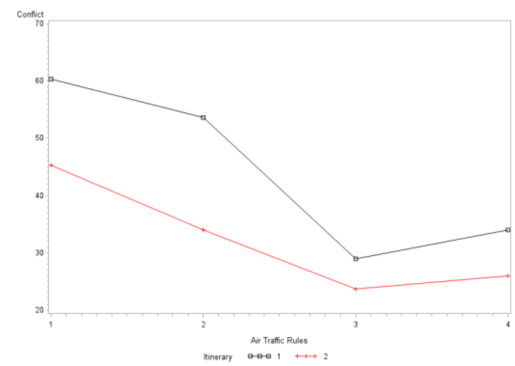
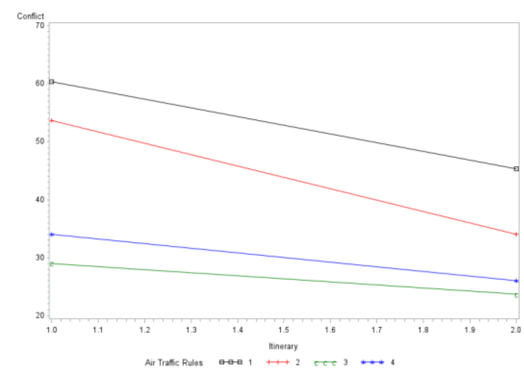
<table>
<tr><th>Plot</th><th>Plot Description</th></tr>
<tr><td>
</td><td>Interaction plot of Sequence Actions as a response variable by Air Traffic Rule</td></tr>
<tr><td>
</td><td>Interaction plot of Sequence Actions as a response variable by Itinerary</td></tr>
<tr><td>
</td><td>Interaction plot of Number of Conflicts as a response variable by Air Traffic Rule</td></tr>
<tr><td>
</td><td>Interaction plot of Number of Conflicts as a response variable by Itinerary</td></tr>
</table>

Table 5.Interaction Plot Summary

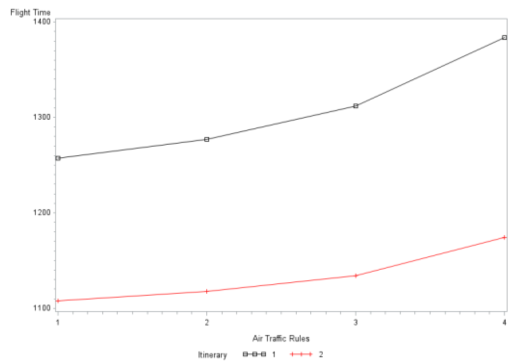
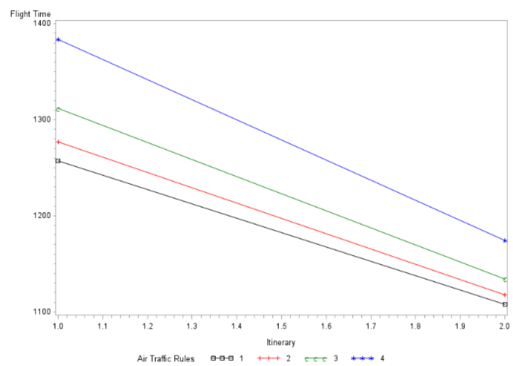
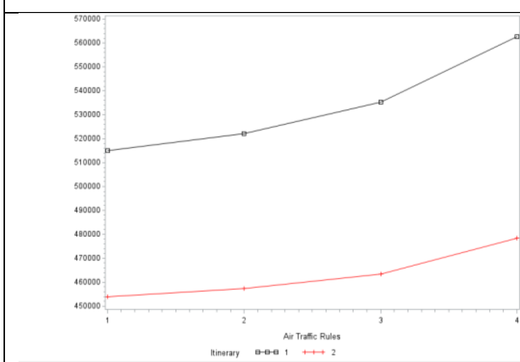
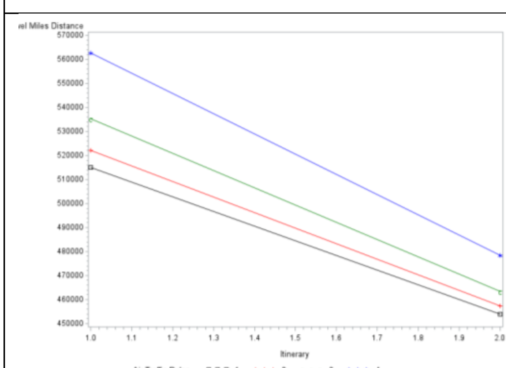
Plot	Plot Description
	Interaction plot of Flight Time as a response variable by Air Traffic Rule
	Interaction plot of Flight Time as a response variable by Itinerary
	Interaction plot of Track Mile Distance as a response variable by Air Traffic Rule
	Interaction plot of Track Mile Distance as a response variable by Itinerary

Table 6. Interaction Plot Summary

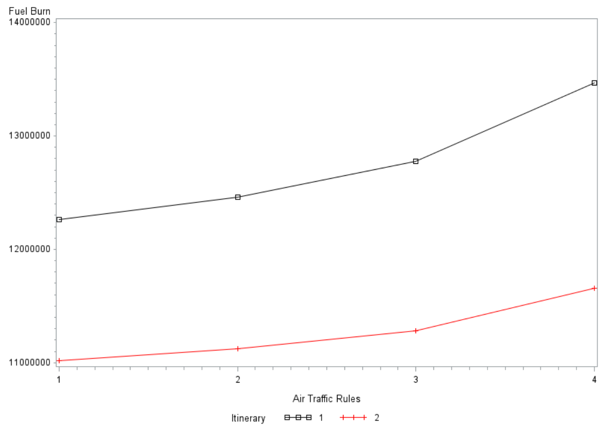
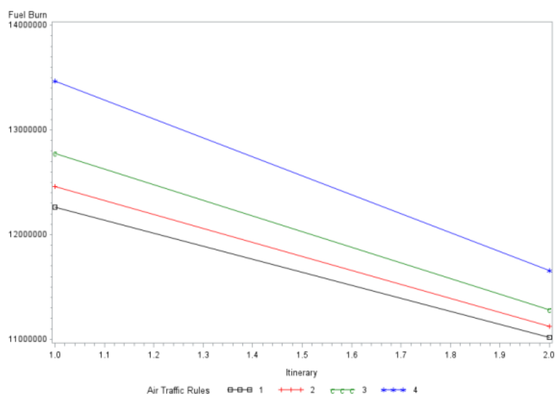
Plot	Plot Description
	Interaction plot of Fuel Burn as a response variable by Air Traffic Rule
	Interaction plot of Fuel Burn as a response variable by Itinerary

Table 4 shows the interaction plot for Itinerary and Air Traffic Rules using the response variable as sequence actions, and the same type of plots using the response variable the number of conflicts. The plots of sequence actions show that the Air Traffic Rule level 1, which is 3NM miles, minimize the numbers of sequence actions. In contrast, the plots of numbers of conflicts present that the Air Traffic Rule level 3, which is 7NM miles, minimize the numbers of conflicts and the level 1 of “ATRules” is the worst for this purpose.

In addition, table 5 shows the interaction plot for Itinerary and Air Traffic Rules using the response variable the Flight Time and the same kind of plot using the response variable the Track Mile Distance. The plots of Flight Time and Track Mile Distance show the same. The Air Traffic Rule at level 1 minimizes both response variables.

Finally, the table 6 shows the interaction plot for Itinerary and Air Traffic Rules using the response variable the Fuel Burn. This plots shows that the level 1 of Air Traffic Rules minimizes the Fuel Burn. Therefore, there is an issue between the interaction plot results from the number of conflict and the others interaction plots, since the level 1 of Air Traffic Rules minimize all the response variables except the number of conflict, which is maximized.

A. Pairwise Tukey Comparison

In order to conduct the corresponding family of tests of the form:

H_o: D=0

H_1: D≠0

The objective is to find the significance of the comparison. So, if 0 is included in the confidence interval that means that is not statistically significant.

(a) Sequence Actions

Least Squares Means for Effect Iti*ATRules				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-66.333333	-74.101787	-58.564879
1	3	-46.000000	-53.768454	-38.231546
1	4	-63.000000	-70.768454	-55.231546
1	5	31.333333	23.564879	39.101787
1	6	-16.333333	-24.101787	-8.564879
1	7	-21.333333	-29.101787	-13.564879
1	8	-27.000000	-34.768454	-19.231546
2	3	20.333333	12.564879	28.101787
2	4	3.333333	-4.435121	11.101787
2	5	97.666667	89.898213	105.435121
2	6	50.000000	42.231546	57.768454
2	7	45.000000	37.231546	52.768454
2	8	39.333333	31.564879	47.101787
3	4	-17.000000	-24.768454	-9.231546
3	5	77.333333	69.564879	85.101787
3	6	29.666667	21.898213	37.435121
3	7	24.666667	16.898213	32.435121
3	8	19.000000	11.231546	26.768454
4	5	94.333333	86.564879	102.101787
4	6	46.666667	38.898213	54.435121
4	7	41.666667	33.898213	49.435121
4	8	36.000000	28.231546	43.768454
5	6	-47.666667	-55.435121	-39.898213
5	7	-52.666667	-60.435121	-44.898213
5	8	-58.333333	-66.101787	-50.564879
6	7	-5.000000	-12.768454	2.768454
6	8	-10.666667	-18.435121	-2.898213
7	8	-5.666667	-13.435121	2.101787

Figure 15. Pairwise Tukey Comparison for Sequence Actions

Figure 15 shows the 36 pairwise comparison of Tukey. Consequently, the figure 8 shows the following information.

All the comparisons are statistically significant, except:

1. The comparison between sequencing actions when the interaction is high season itinerary and 5NM as Air Traffic Rule and the sequencing actions when the interaction is high season itinerary and 10 NM.
2. The comparison between sequencing actions when the interaction is low season itinerary and 5NM as Air Traffic Rule and the sequencing actions when the interaction is low season itinerary and 7 NM.
3. The comparison between sequencing actions when the interaction is low season itinerary and 7NM and the sequencing actions when the interaction is low season itinerary and 10NM as Air Traffic Rule.

(b) Number of Conflicts

Least Squares Means for Effect Iti*ATRules

i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	6.666667	-2.004479	15.337812
1	3	31.333333	22.662188	40.004479
1	4	26.333333	17.662188	35.004479
1	5	15.000000	6.328854	23.671146
1	6	26.333333	17.662188	35.004479
1	7	36.666667	27.995521	45.337812
1	8	34.333333	25.662188	43.004479
2	3	24.666667	15.995521	33.337812
2	4	19.666667	10.995521	28.337812
2	5	8.333333	-0.337812	17.004479
2	6	19.666667	10.995521	28.337812
2	7	30.000000	21.328854	38.671146
2	8	27.666667	18.995521	36.337812
3	4	-5.000000	-13.671146	3.671146
3	5	-16.333333	-25.004479	-7.662188
3	6	-5.000000	-13.671146	3.671146
3	7	5.333333	-3.337812	14.004479
3	8	3.000000	-5.671146	11.671146
4	5	-11.333333	-20.004479	-2.662188
4	6	0	-8.671146	8.671146
4	7	10.333333	1.662188	19.004479
4	8	8.000000	-0.671146	16.671146
5	6	11.333333	2.662188	20.004479
5	7	21.666667	12.995521	30.337812
5	8	19.333333	10.662188	28.004479
6	7	10.333333	1.662188	19.004479
6	8	8.000000	-0.671146	16.671146
7	8	-2.333333	-11.004479	6.337812

Figure 16. Pairwise Tukey Comparison for Number of Conflicts

Figure shows 16 that the majority of the comparison are significant because they are not including 0 in the Tukey Confidence Interval. The following are the exceptions:

1. High Season Itinerary with 3NM vs High Season Itinerary with 5NM.
2. High Season Itinerary with 5NM vs Low Season Itinerary with 3NM.
3. High Season Itinerary with 7NM vs High Season Itinerary with 10NM.
4. High Season Itinerary with 7NM vs Low Season Itinerary with 5NM.
5. High Season Itinerary with 7NM vs Low Season Itinerary with 7NM.
6. High Season Itinerary with 7NM vs Low Season Itinerary with 10NM.
7. High Season Itinerary with 10NM vs Low Season Itinerary with 5NM.
8. High Season Itinerary with 10NM vs Low Season Itinerary with 10NM.
9. Low Season Itinerary with 5NM vs Low Season Itinerary with 10NM.
10. Low Season Itinerary with 5NM vs Low Season Itinerary with 10.

(c) Flight Time, Track Mile Distance and Fuel Burn

Least Squares Means for Effect Iti*ATRules

i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-19.803333	-25.329599	-14.277068
1	3	-54.110000	-59.636266	-48.583734
1	4	-125.900000	-131.426266	-120.373734
1	5	149.456667	143.930401	154.982932
1	6	139.676667	134.150401	145.202932
1	7	123.150000	117.623734	128.676266
1	8	83.376667	77.850401	88.902932
2	3	-34.306667	-39.832932	-28.780401
2	4	-106.096667	-111.622932	-100.570401
2	5	169.260000	163.733734	174.786266
2	6	159.480000	153.953734	165.006266
2	7	142.953333	137.427068	148.479599
2	8	103.180000	97.653734	108.706266
3	4	-71.790000	-77.316266	-66.263734
3	5	203.566667	198.040401	209.092932
3	6	193.786667	188.260401	199.312932
3	7	177.260000	171.733734	182.786266
3	8	137.486667	131.960401	143.012932
4	5	275.356667	269.830401	280.882932
4	6	265.576667	260.050401	271.102932
4	7	249.050000	243.523734	254.576266
4	8	209.276667	203.750401	214.802932
5	6	-9.780000	-15.306266	-4.253734
5	7	-26.306667	-31.832932	-20.780401
5	8	-66.080000	-71.606266	-60.553734
6	7	-16.526667	-22.052932	-11.000401
6	8	-56.300000	-61.826266	-50.773734
7	8	-39.773333	-45.299599	-34.247068

Figure 17. Pairwise Tukey Comparison for Flight Time

Least Squares Means for Effect Iti*ATRules

i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-7177.333333	-9551.327122	-4803.339545
1	3	-20115	-22489	-17741
1	4	-47727	-50101	-45353
1	5	61289	58915	63663
1	6	57632	55258	60006
1	7	51735	49361	54109
1	8	36755	34381	39129
2	3	-12938	-15312	-10564
2	4	-40550	-42924	-38176
2	5	68467	66093	70841
2	6	64809	62435	67183
2	7	58912	56538	61286
2	8	43932	41558	46306
3	4	-27612	-29986	-25238
3	5	81404	79030	83778
3	6	77747	75373	80121
3	7	71850	69476	74224
3	8	56870	54496	59244
4	5	109017	106643	111391
4	6	105359	102985	107733
4	7	99462	97088	101836
4	8	84482	82108	86856
5	6	-3657.333333	-6031.327122	-1283.339545
5	7	-9554.666667	-11929	-7180.672878
5	8	-24535	-26909	-22161
6	7	-5897.333333	-8271.327122	-3523.339545
6	8	-20877	-23251	-18503
7	8	-14980	-17354	-12606

Figure 18. Pairwise Tukey Comparison for Track Mile Distance

Least Squares Means for Effect Iti*ATRules

i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-199774	-263477	-136071
1	3	-516002	-579705	-452299
1	4	-1209758	-1273462	-1146055
1	5	1246276	1182573	1309979
1	6	1138233	1074530	1201936
1	7	979001	915298	1042704
1	8	603498	539795	667201
2	3	-316228	-379932	-252525
2	4	-1009984	-1073688	-946281
2	5	1446050	1382347	1509753
2	6	1338007	1274304	1401710
2	7	1178775	1115071	1242478
2	8	803272	739569	866975
3	4	-693756	-757459	-630053
3	5	1762278	1698575	1825982
3	6	1654235	1590532	1717938
3	7	1495003	1431300	1558706
3	8	1119500	1055797	1183203
4	5	2456034	2392331	2519738
4	6	2347991	2284288	2411695
4	7	2188759	2125056	2252462
4	8	1813256	1749553	1876960
5	6	-108043	-171746	-44340
5	7	-267275	-330979	-203572
5	8	-642778	-706481	-579075
6	7	-159232	-222935	-95529
6	8	-534735	-598438	-471032
7	8	-375503	-439206	-311799

Figure 19. Pairwise Tukey Comparison for Fuel Burn.

Figure 17, 18 and 19 shows that none of the comparison include 0 in the interval, so all of them are statistically significant.

d) Preliminary Conclusions

As a conclusion, the three-factor complete factorial design linear model is statistically significant at 0.1 level of significance. However, the Ground Traffic Rules is not significant at 0.1 level of significance, so it has not effect in the Air Traffic KPIs. After, some empirical interviews the Air Traffic Controllers were changing the Ground Traffic Rules to make a Union pressure, so the Ground Traffic Rule produce an effect in the performance of the operation when the Air Traffic Rules use 7 to 10 minutes, which is way beyond the range of this particular experiment.

The main objective of the simulation model is to minimize the Air Traffic KPIs, so the interaction plots shows that the level 1 of Air Traffic Rules is the best to minimize the number of Sequence Actions, the Flight Time, the Track Mile Distance and the Fuel Burn, but it is not the case of Number of Conflicts. In other words, 3NM if is used as a standard for Air Traffic Rule can reduce almost all the KPIs, except the Number of Conflicts, which is better reduced at 7NM. Which is interesting for this air route redesign used for the simulation is that over 7NM the tendency for Number of Conflicts increase, while in the actual layout from empirical interviews over as increase the Air Traffic Rule decrease the Number of Conflicts.

Based on the Tukey Pairwise Comparison, the analysis when the number of Conflicts is the response variable appears to have 10 over 36 comparisons as not statistically significant. Which include 4 of the 8 comparisons using the level 3 of Air Traffic Rules. The level 3 of Air Traffic Rules is who minimize the number of conflicts based on the interaction plots.

2) Second DACE

Also, the RAID lab from the University of Texas at Arlington propose to Panama Academia a Conference to understand the Air operation and the opportunities to apply Airspace Optimization and Collaborative Decision Making. This conference was possible thanks to USA Embassy in Panama, the AAC, COPA, FAA, TOCUMEN, UTP, ICAO, IATA and other companies and international institutions.

Therefore, we collect a second set of data from TAAM that included 6 Key performance indicators. Those inputs and several meetings with COPA, Tocumen International Airport and The Autoridad de Aeronautica Civil (AAC) has the following outcomes to analyze:

a. The first experiment in the software name TAAM had 2 factor that were statistical significant at all the levels. Four of the Key Performance Indicators behave similarly, but one was the contrary of the others 4. In other words, safety is proportionally inverse to Fuel burn consumption.

b. The second experiment included delays, but did not include more than one itinerary, as result the first 5 Key Performance indicators were not statistically significant and just delays was. As a matter of fact, this study did not use Ground Traffic Rules, because was statistically insignificant the first time, and did not use more than one itinerary. The Factors were growth and Air Traffic Rules.

   The results of this Experiment were different than the first one, with not level of significant at alpha 0.1 for Air Traffic Rules per each KPI.

   Consequently, there are two possible options, since the first time the consideration was a model based on Point Merge, which is more sensible to the Air Traffic Rules.

The other option is the lack of two itineraries for the study. Although, the research can take that set of data as a control data in term of forecast for future studies.

3) Data Exploration and Classification

The data exploration includes some previous data from December 2015 that can give us an idea of the performance of the daily or hourly operation.

For example, figure 20 shows the data exploration of the arrivals and departures of December 2015.

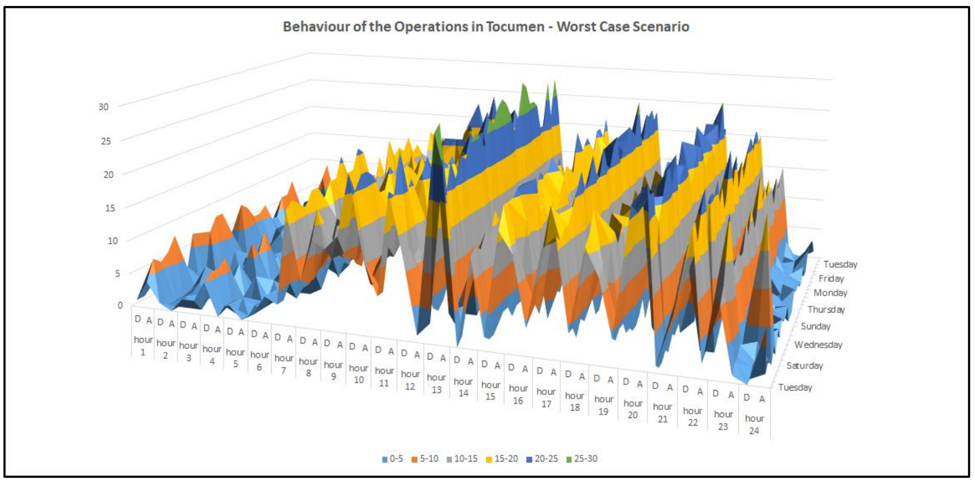


Figure 20. Behavior of the Operation in Tocumen – Worst Case Scenario- Dec 2015.

December is one of the most difficult months just in terms of volume, but that in particular was a headache for the Operation, since they lost the only control tower at the airport for an issue with bug infestation. (5) This data set can be an example of worst case scenario, that include high season, rain season, and emergency respond, since some of the flights were not able to land in Panama, but Costa Rica.

In addition, some other data exploration can help to understand the behaviour of that particular month. For example, the plots can present the behaviour of the arrivals and departures per hour and per day, or the sum of the operations per day. Figure 21 shows how is the pattern of the whole operation per hour. Figure 22 shows that there is not much difference between the sum of the arrivals and departures per day. Figure 23 shows that when the arrivals are high in one particular hour the departures are low and vice-versa, which means that from hour 7 to hour 24 the system is close to the maximum capacity and the maximum number of operations is fixed.

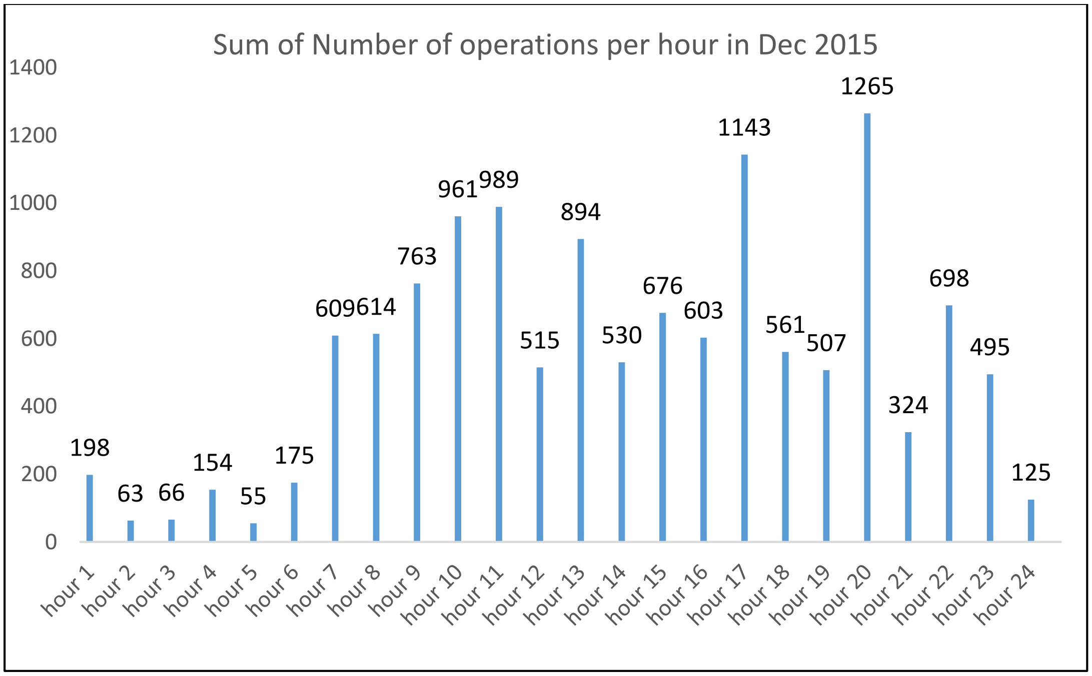


Figure 21. Sum of Number of Operations during December 2015.

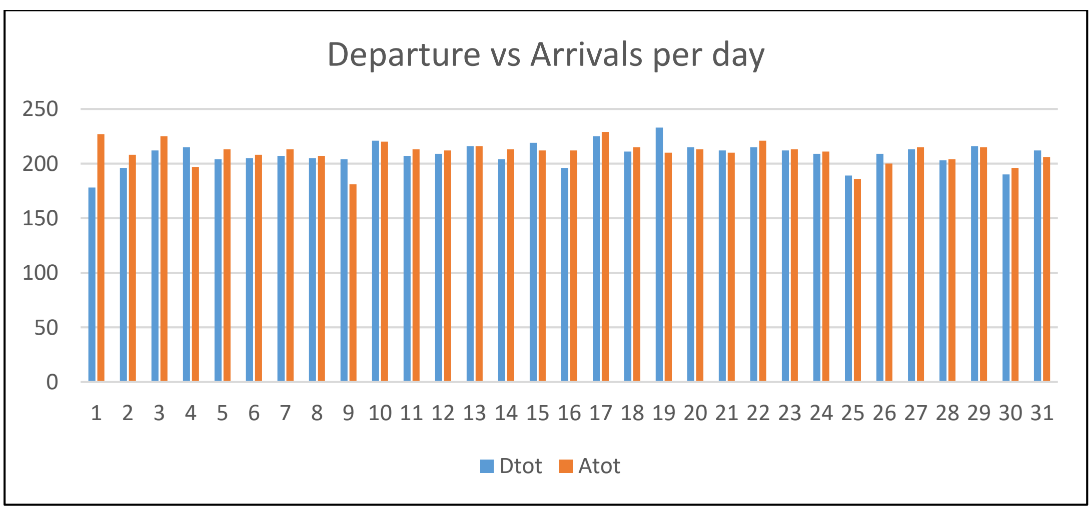


Figure 22. Contrast between Number of Departures and Number of arrivals per day.

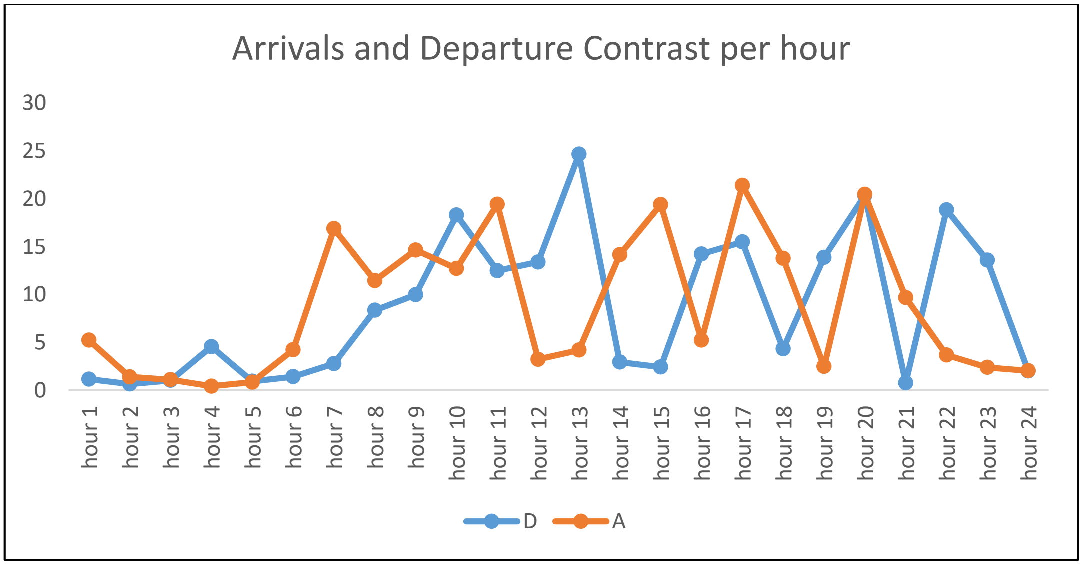


Figure 23. Contrast between Number of Arrivals and Number of Departures per hour.

VI. RESEARCH GOALS, OBJECTIVES OR AIMS

A. Research questions

The main concern that we found in the Airspace Optimization Conference and Workshop in 2017 was the level of congestion in the air, since the last year Copa Airlines described a comparison analysis between Tocumen International Airport and other airports in Latin America with better management of the demand, in number of operations per day, than Panama.

Since, the air traffic is being managed with fixed air routes without any redesign for more than 30 years, the AAC approach to solve the problem was to redesign the air routes. However, it came to my attention the possibility to improve the actual system without changing the air routes, process that can take 4 years in studies and implementation (2). In other words, there is something in the airspace management that can be improved to reduce the air traffic congestion in a short period of time. Therefore, this logic of improvement of the system without changing the design of the air routes through reducing congestion leads me to my research question.

Can Air Traffic Congestion in Panama be reduced through minimizing the impact of historical congestion variables?

B. Objectives

- ***Specific Objective #1:*** *Determine which variables are most relevant to minimize Air Traffic Congestion.*
- ***Specific Objective #2:*** *Determine the significance of the variables and their impact on the Air Traffic Congestion.*
- ***Specific Objective #3:*** *Identify the cost effectiveness of the variables on Air Traffic Congestion.*

C. Hypothesis

In order to address the specific objectives, I did investigate the following hypothesis.

Null Hypothesis 1: Equation 17 is statistically significant at 0.05 alpha value.

Alternative Hypothesis 1: Otherwise.

Null Hypothesis 2: The Congestion Factor has an exponential effect over the minimum Nautical Miles per model.

Alternative Hypothesis 2: Otherwise

Null Hypothesis 3: One objective function can be identified from Rastrigin, Rosenbrock, Levy and Sphere equations.

Alternative Hypothesis 3: Otherwise

Null Hypothesis 4: A Genetic Algorithm with Game Theory is the better approach to solve the multi objective integer optimization model.

Alternative Hypothesis 4: Otherwise

Null Hypothesis 5: All Factors studied in the META Model are statistically significant at alpha of 0.05.

Alternative Hypothesis 5: Otherwise

After some challenges in the data collection, we decide to adapt the hypothesis 4 and 5.

Null Hypothesis 4: Unified Optimization Method Applied to Vehicle Routing (82, 83) is the preferred model to minimize value of Nautical Miles.

Alternative Hypothesis 4: Otherwise

Null Hypothesis 5: Genetic Algorithms is the preferred Method for Multi-objective optimization between minimizing cost for fuel burn in contrast to maximize capacity (79)

Alternative Hypothesis 5: Otherwise

The specific objective 1 is addressed with the hypothesis one and partially the hypothesis two, the specific objective 2 is addressed for the hypothesis 3 and revised hypothesis 4, and the specific objective 3 is addressed with the revised hypothesis 5.

In order to test my hypothesis, I seek to create a model, which will be explained at the Design step of the Methodology, which pursues to minimize the air traffic congestion in Panama. Therefore, I will test the feasibility of these variables in scientific manner.

VII. METHODOLOGY

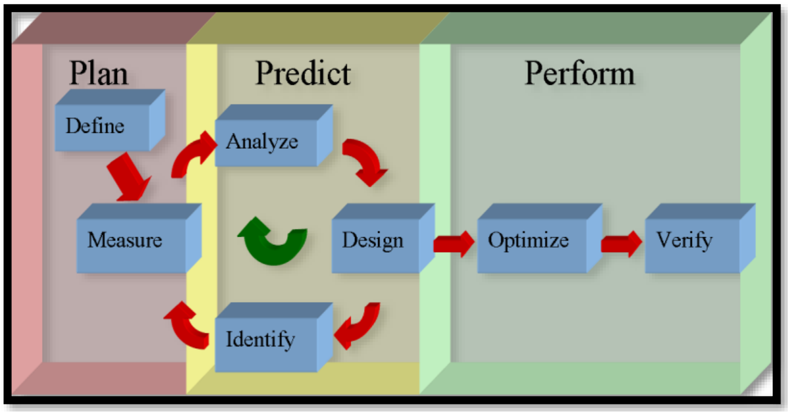


Figure 24. Design for Six Sigma-Research

A Motorola engineer named Bill Smith credits the term “Six Sigma”. The label “Sig Sigma” originates from statistical terminology. The meaning of Six Sigma is to reduce defects, increase company productivity, and improving company profitability. (8). Many practitioners utilize the Six Sigma methodology DMAIC or Define, Measure, Analyze, Improve and Control. In the planning phase, problems must be defined and measured before a set of solutions can be evaluated. In the Predict phase, researchers analyze the data measured, design and implement new process or tools to solve the problem, and then analyze the new processes and tools to determine how well the solution solved the problem (8). In the case of RAID labs, a methodology known as DFSS-R or “Design for Six Sigma-Research” is applied which is presented in figure 24.

A. Phase 1: Plan

1) Define: Methodology Definition

a) Data Management and Data Analytics

The plan for Data Management includes the objective 1 and 2 to understand the data and select which data is suitable to run any experiment, simulation or optimization. Also, the previous DACE studies can give a proper guide in the capacity optimization. In addition, there is a data base from 2007 to 2016 that was evaluated to understand the pattern in the arrivals and departures behavior.

b) Data Classification

Since, the operation of the Tocumen International Airport is based on six-bank-hub, there are two ways to classify the data. The first one is just taking the time frame per day of the six bank hub form 6:05 a.m. to 10:08 p.m. and the second one is taking the 24 hours of operations (52). This is an important consideration because the congestion in the arrivals is mainly in the time frame of the six-bank hub. However, the classification can be based on season, months, days and even per hour.

c) Multi Objective Optimization

We took in consideration the data from 2016 DACE model to evaluate the feasibility of the multi-objective optimization. The challenge is to maximize capacity and reduce 5 KPIs studied in the first DACE 2016 model. We evaluated the methodologies of multi-objective optimization, and we found that the evaluation of more than 3 objectives functions is not real. We tried to select just two Key Performance Indicators that in our case will be Fuel Burn Consumption and number of Conflicts, since the number of conflicts is the safety indicator. The safety will be always the first priority in aviation, so it is better to focus in minimize the number of conflicts alone. However, since the economic aspect is really important to balance the operation we will optimize two objective functions. The first one considering the Number of Conflicts and the second one minimizing the fuel burn consumption, which has impact on the safety, because if the aircraft run out of fuel it considers it a high risk situation.

However, one of the fundamental aspects of Air Traffic Flow and Capacity Management is to optimize capacity, there is a further consideration based on the DACE 2017 model, which is the direct relation between demands in contrast to each Key performance indicator.

d) DACE META MODEL

In many engineering optimization problem, evaluating the objective function is a challenge since the objective function is unknown. There are some computer simulators which are capable of simulating the response. However, the running time of these simulators are considerable. The problem design, like airspace optimization, is dealing with this challenge mostly. There are several design parameters that affect the objective function which should be optimized.

Designing a computer model that efficiently estimates the black box function in order to find the optimum solution is a response to this challenge. However, finding the optimum modelling parameters is the main focus of this study; number of initial observations, initial points design method, different objective function evaluation can be considered to get the best estimate of the objective function.

The model has the following steps: Initialization (Designing the initial data points), building a tree on top of initial data set, fit a model on terminal nodes' observations, find the confidence band over fitted model, choose the optimum new sample regarding the bands and rules of the tree for each terminal node, refit the model with the new data set.

The objective with the initialization is 1) to set different design of experiments methods (Sobol, Orthogonal Array, Latin Hypercube design) 2) to test different objective functions; Rastrigin, Rosenbrock, Levy and Sphere per eack key performance indicator. 3) to test different multiobjective optimization; Goal attainment, Minimax, Multiobjective genetic algorithm.

After the initialization step, building a tree and fitting a model on terminal nodes come next. In this study we try fit a Multiple Linear Regression (MLR) on terminal nodes.

Also, the project will include the optimum sample analysis. In this step, we will study adding one by one the optimum samples obtained by each terminal node.

Thus, the planning of the DACE model did not find a suitable approach, after we select the data from the AAC historical dataset, since we set the optimization model based on the actual operations constraints to be a Mixed Integer Linear Programming.

2) Measure

The Measurements was planned to evaluate if the historical data distribution cannot fit the Weibull, Normal, Logistic, Kernel, or Poisson Distribution, with a Maximum Likelihood estimate bellow 1. Also, identify a suitable objective function for the black box model, finding the minimum value per each KPI. Try to find the minimum value of Air Traffic Rules between 3NM and 7NM that can keep the KPIs in balance.

However, the actual measurement was the identification of the metrics that affect the congestion, the dependent variable and the independent variables. Also, the identification of the metrics for the optimization objective function and the constraints. Furthermore, the metrics for the transformation of the sum of the flight mileage to cost of fuel burn, took different measures and transformation of metrics.

B. Phase 2: predict

1) Analyze:

The analysis of the air traffic operation includes the Design and Analysis of Computer Experiments (DACE) to understand the previous simulation model and the relationship between factors and Key Performance Indicators. Also, the data mining analysis of the historical data from AAC. The other phase is the stochastic and probability analysis of the data from AAC and other inputs from the interviews, workshops, meetings and the empirical observation. Furthermore, the process diagram from figure 3 shows that the fourth phase is to define different approaches to minimize the objective function with multiple objectives. The last phase includes the DACE META MODEL and the control step that includes the reliability, statistical and economic analysis. However, as we stand before, we did not perform the DACE META MODEL, but we did a Statistical META MODEL that made a mixture between Statistical Process Control, Optimization and Engineering Economics using Operational Research tools for Uncertainty.

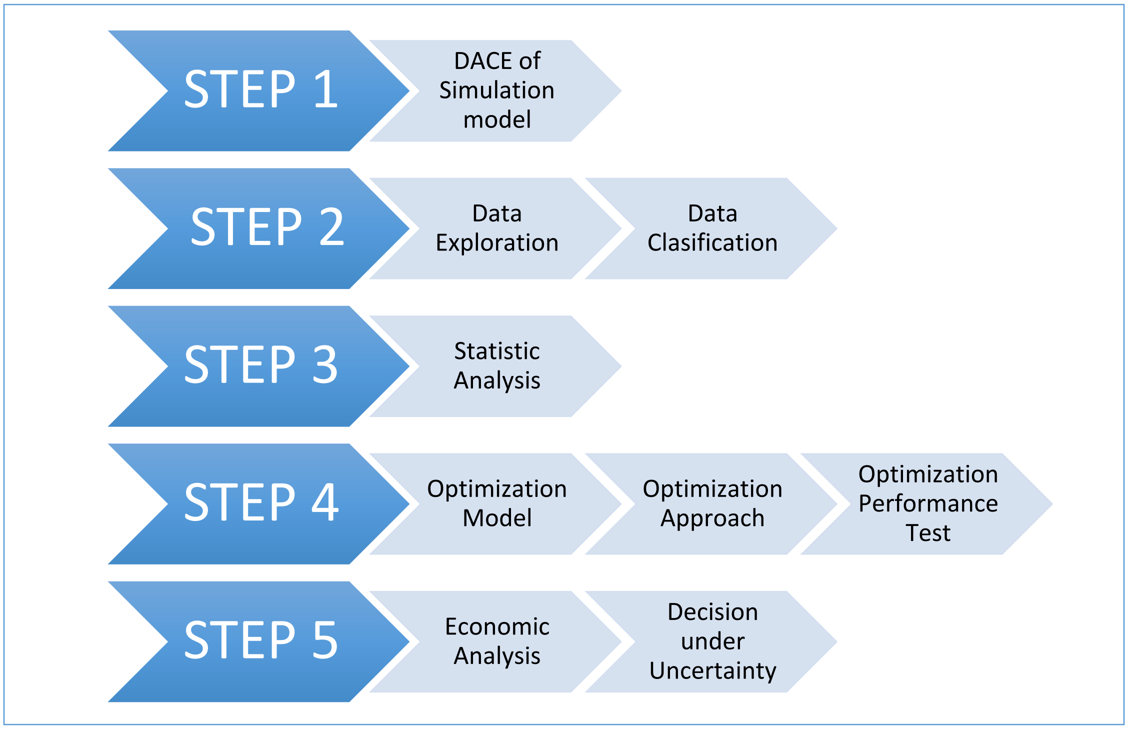


Figure 25. Analysis Process

2) *Design:*

a) Model Description

Some organizations and studies are based in the redesign of the Air routes, technology implementation and the interaction of the air traffic flow with the airports.

However, this study is focused on the necesisty to support the air traffic controler decision-making in real time. How the air traffic controllers can reduce uncertainty while making decisions, so the study has two levels

The first level is ideal air traffic flow, base on a simulation model where everything is smooth and we can even apply standard air traffic rules for every aircraft.

The second level of the study deals with the data availability about the traffic performanc, in the event that the simulation does not present realistic output. The factors include types of aircrafts, arrival distribution per route, time on the air space of Panama before landing.

Also, the study presents a model that include different stochastic and probability aproaches to understand the historical data, and translate that information for future possiblitlities, these will serve as an input for different optimization approaches by taking training and test data to ahieve a reliable source or options for the air traffic controllers.

The model starts with some level of uncertainty about the future of the air traffic flow. However, there are some fixed information, like the itinerary planned by each airline. The experience of each air traffic controller is a key factor to respond in real time to the issues about congestion in the arrivals. In contrast, the model does not include the experience of the air traffic controller as a mathematical factor, the study does include some of the inputs from interviews, meetings, workshops, and empirical observation with air traffic controllers, pilots, analyst, executives and real time decision-makers.

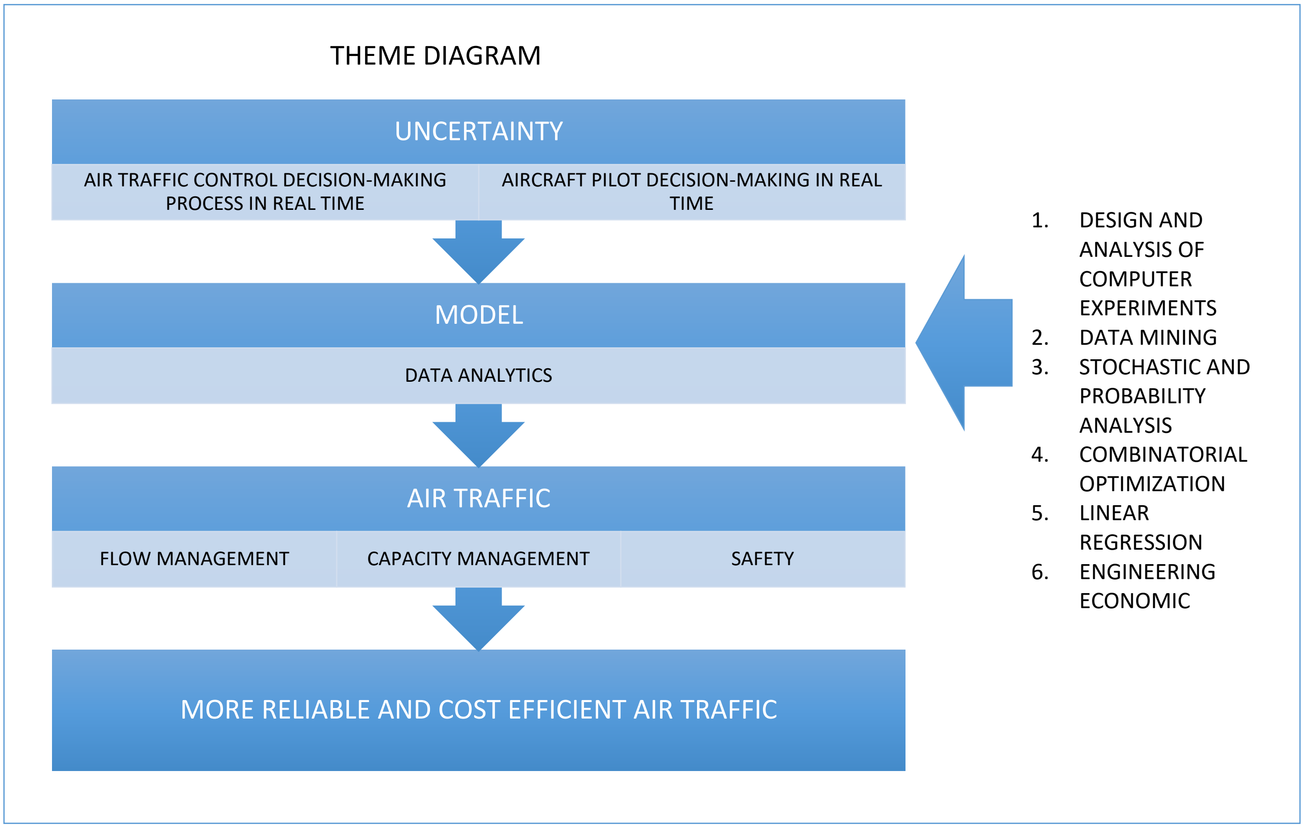


Figure 26. Theme Diagram

3) Identify:

The impact analysis in this case will be the test of each hypothesis. Also, Identifying the cost associated to the fuel burn and the cost per operation without the fuel consumption, for the final optimization model.

C. Phase 3: perform

1) Optimize:

The minimization functions can help reducing the risk of the landing and departure operation of the Tocumen International Airport. However, the other objective is to maximize the number of operations.

The proposed optimization technique is a linear programming to minimize the cost associated to the fuel burn obtained per flight time in contrast with the number of operations that the system wants to maximize between 40 operations per hour to 60 operations per hour. Based on the data set of AAC from December 2015, the average number of operations per hour at the Tocumen International Airport was 18, the minimum was 1 and the maximum was 43. This information is based on a whole day and not the six-bank hub operation.

Equation 7 Tentative Formulation for cost minimization

$$Minimize\colon Z = C_1 X_1 - C_2\, X_2$$

s.t.

$$a \le C_1\, X_1 \le b$$

$$c \le C_2 X_2 \le d$$

$$X_1, X_2 \le 0$$

Where Z = Total cost per flight

"C_1" is the cost per gallon

"a" is the minimum of fuel burn in gallons per flight obtained from the multi-objective optimization

"b" is the maximum of fuel burn in gallons per flight on record

"X_1" is the number of gallons per flight

"C_2" is the rest of the cost associated per flight per operation does includes the fuel cost.

"c" is the minimum of operations that the airport wants to handle per hour

"d" is the maximum number of operations that airport can handle per hour

"X_2" is the number of operations per hour

However, the optimization formulation was not used as it is in equation 7, instead we use the transformation of the sum of flight mileage to cost of fuel burn used.

2) Verify:

The data exploration and the classification will help in further statistical analysis to understand or even test with new data the model with DACE. Also, there is a comparison between the CART with the R-square value versus the p-value. Therefore, the proposal is to check each factor and if has a statistically significant impact over the percentage.

VIII. RESULTS

- The specific objective 1 is to determine which variables are most relevant to minimize Air Traffic Congestion. Consequently, we took in consideration the information from the empirical interviews, meetings, model comparison, the literature review and study of the data available to start using some tools from Statistical Process Control, such as Cause and Effect Diagram and Linear Regression to test the significance of those factors over the congestion addressing at the same time the specific objective 2.

A. DATA COLLECTION AND DATA EXPLORATION

During the first study at Panama, we collected data about the 5 key performance indicators that Copa were using to measure the performance of the airspace over the operation. However, that data was limited to 24 treatments of 5 key performance indicators, so it was too short to evaluate historical data. Also, the information was limited, since they did not share the itineraries used. So, we took the AAC data base that they share to us in 2017 that includes some operations from 2015, 2016 and 2017. The dataset from AAC has all the itinerary information and the specific air routes the aircraft was using during each operation. Consequently, we started the evaluation of the data to understand how the congestion can be measure, and what other information can we obtain that can impact the congestion.

The data sample contains information of 181 days, or 6 months that include February and March 2015 operations, June, July and August 2016 operations, and January 2017 operations.

Type I error test:

Ho: The sample is statistically significant at 0.01 alpha value.

Ha: The sample is not statistically significant at 0.01 alpha value.

When we calculate the alpha value for is far less than 0.01 alpha value.

The tools used for the data exploration were Tableau, Microsoft Excel and MATLAB. The tool used for the statistical analysis was SAS. The tool used for the Optimization was MATLAB, and the too used for the Economic Analysis was Microsoft Excel.

The original dataset from AAC (Panamanian Government Agency) is confidential, so we did not add that data and the processed data in the appendix. Consequently, the original dataset from AAC, the data used for the Multiple Linear Regression Models, the data used for the Optimization and the data used for Engineering Economic analysis is saved at a digital repository at the Dissertation Supervisor office.

B. ISHIKAWA ANALYSIS AND LINEAR MODELS

1) Approach

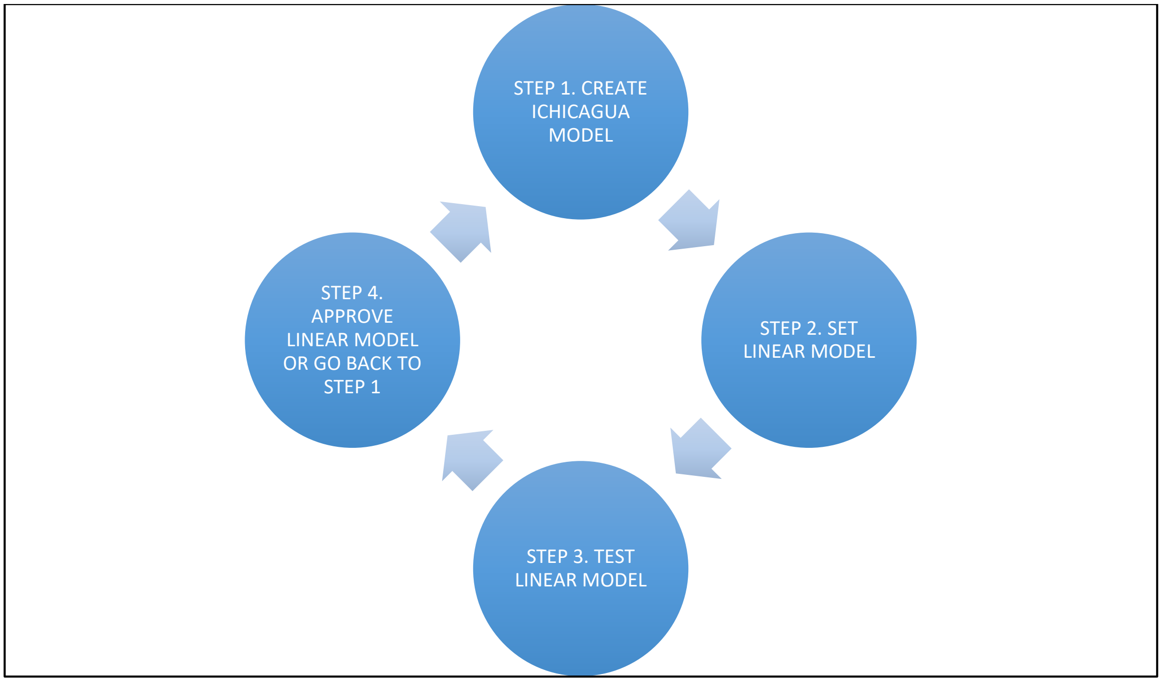


Figure 27. Approach for Ishikawa model with Linear model testing

The approach to analyse the linear models include four steps. The first step of the approach is the design of a Cause and Effect diagram model based on the empirical interviews, empirical observation, data collected, and literature review. The second step is the formulation of the linear model of the Cause and Effect diagram. The third step is to test the linear model based on linear regression methodology. The fourth step is to approve or not the linear model based on Design of Experiment model selection with F value testing.

C. *FIRST ISHICAWA MODEL*

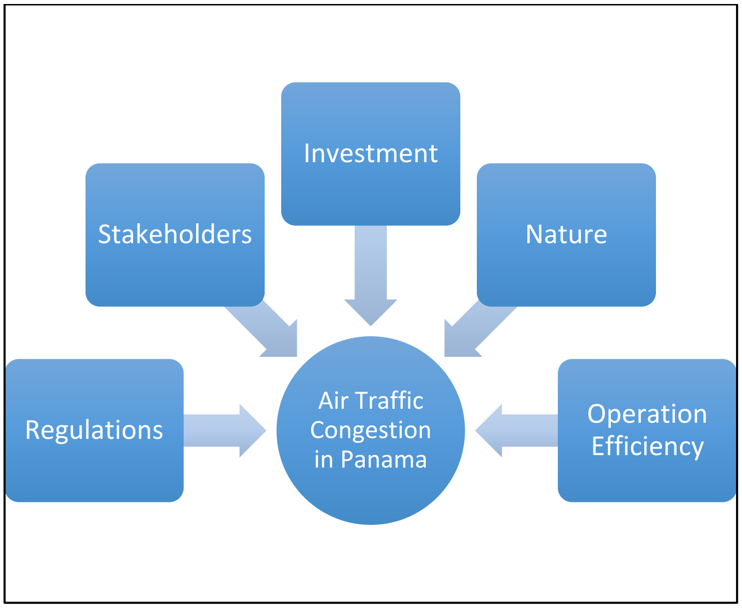


Figure 28. Cause and Effect Diagram

The first Cause and effect diagram shows the whole system with the consideration of the five main factors that affect the operation in Panama from empirical interview, data evaluation and the Air Optimization Conference. Also, this model includes sub equations that describe the sub branches from each main factor.

Equation 8. First Ishikawa Model.

$$Y = AX_1 + BX_2 + CX_3 + DX_4 + FX_5 + \varepsilon$$

Table 7. Model Notation

Variable	**Name**	**Metrics**
X_1	Regulations	No of Operations affected by a violation of the regulation
X_2	Stakeholders	No of Operations affected by a error or bad decision from the stakeholders
X_3	Investment	No of Operations affected by the investment
X_4	Nature	No of Operations affected negatively by the wind, lack of visibility or rain
X_5	Operation Efficiency	No of Operations affected inefficiency
ε	Error	
Y	Air Traffic Congestion in Panama	Rate of No of Operations with more than the minimum flight miles per route per day

Equation 9. Sub-Equation for Regulations

$$X_1 = IGR + IAR + NGR + NAR$$

Table 8. Equation 8 explanation

Variable	**Name**	**Metrics**
X_1	Regulations	No of Operations affected by a violation of the regulation
IGR	International Ground Rules (Airport)	No violations to this regulation
IAR	International Air Rules (Airspace)	No of violations to this regulation
NGR	National Ground Rules (Airport)	No of violations to this regulation
NAR	National Air Rules (Airspace)	No of violations to this regulation

The table 8 shows the sub-variables that affect the effect of the regulations of the operations, such as international rules for ground operations, international rules for airspace operation, national regulations for the airport operation and the regulations that affect the airspace operation.

The organizations that affect the international regulations in the case of Panama are IATA, ICAO and sometimes the FAA. For example, if an aircraft does not follow the procedure of landing it can be send to a penalty box, where the aircraft must wait until there is a space open in the line for landing.

The Panamanian Government rules over the national regulations, that includes penalties over the airlines for lack of cooperation or penalties over the airlines for not following a procedure. Also, the Panamanian Government rules over the permissions and taxes for use of the airspace or for landing in Panama.

Equation 10. Sub-Equation for Stakeholders.

$$X_2 = Airlines + AAC + Unions + Airport$$

Table 9. Equation 9 explanation

Variable	**Name**	**Metrics**
X_2	Stakeholders	No of Operations affected by a error or bad decision from the stakeholders
Airlines	All the airlines that use Tocumen International Airport, and Copa Airlines represent the 80% of the operation	Number of hours that the operation was affected by an error or bad decision from the airlines
AAC	Civil Aviation Authority	Number of hours that the operation was affected by an error or bad decision from the AAC
Unions	• Air Traffic Control Union • Pilots Union • Flight Attendance Union • Mechanic Union Airport Employees Union	Number of hours that the operation was affected by an error or bad decision from the Number of hours that the operation was affected by an error or bad decision from the airlines
Airport	In this particular case Tocumen International Airport.	Number of hours that the operation was affected by an error or bad decision from the Airport

Table 9 shows the sub factors that affect operational impact of the decisions of the stakeholders, such as, the airlines decisions, the Civil Aviation Authority's decisions, the Unions decisions and the Airport decisions. For example, a strike from the pilots over Copa Airlines can reduce the efficiency of the operation, the issues with the computational system in Copa Airlines in 2016 that cancel several flights, the energy black outs of June

2017 or the decision of the Air Traffic Controllers of increasing the Ground Traffic Rules from 1 to 2 minutes to 7 minutes lack between landing and departure.

Equation 11. Sub-Equation of Investment.

$$X_3 = Airport\ Investment + Airspace\ Investment$$

Table 10. Equation 10 explanation

Variable	**Name**	**Metrics**
X_3	Investment	No of Operations affected by the investment
Airport Investment	• New gates • New Taxi ways • New runways	Number of hours that the operation was affected for this investments
Airspace Investment	• Airspace Optimization • New Technology	Number of hours that the operation was affected for this investments

Table 10 shows the equation 11 explanation about the two kind of investment in the aviation industry at Panama. The first type of investment is the Ground or Airport investment, that include in the last 5 years, expansion of the Airport, new taxi ways, maintenance of the runways and the evaluation of the expansion to a third terminal with a third runway. The second type of investment is the evaluation of the airspace optimization to change the actual routes and to add more equipment that the AAC is in need, such as radars, radios, and servers (2).

Equation 12. Sub-Equation of the Nature Impact on the operation.

$$X_4 = Wind + Visibility + Birds$$

Table 11. Equation 11 explanation

Variable	**Name**	**Metrics**
X_4	Impact of the Nature over the congestion	No of Operations affected by the wind or bad weather
Wind	The wind intensity and direction can affect the performance of the aircrafts and the direction of the use of the runways.	Average wind speed per day
Visibility	Depending of the rain the visibility can be low, medium or clear	Average visibility per day
Rain	Precipitation of water	Average precipitation per day

Table 11 shows the explanation of the equation 12, that includes the weather factors that impact over the operation, such as wind, visibility and rain precipitation.

Equation 13. Sub-Equation of Operation efficiency.

$$X5 = Landing\ Delays + Departure\ Delays + Flight\ Miles$$

Table 12. Equation 12 explanation

Variable	**Name**	**Metrics**
X5	Operation Efficiency	No of Operations affected by inefficiency
Landing service rate	Average landing per hour each day	Operations per hour
Departure Service rate	Average landing per hour each day	Operations per hour
Flight Miles	The sum of the Nautical Miles that the aircraft take in the Airspace of Panama.	Sum of Nautical Miles per day

Table 12 explain the sub factors that affect the operation efficiency from the Panama stand point, such as landing service rate, departure service rate and flight miles. In the literature, there exist other operational efficiency factors, but we did not include those, since the evaluation is based on the critical factors that the collaboration between stakeholders detected back in the "Airspace Optimization Conference and Workshop at 2017". (2)

After some consideration of the dataset available to address the real impact of each factor over the extra mileage. Thus, we decline to follow this specific model and we decide to find the way to address a model in more accordance with the data availability.

D. SECOND ISHIKAWA MODEL

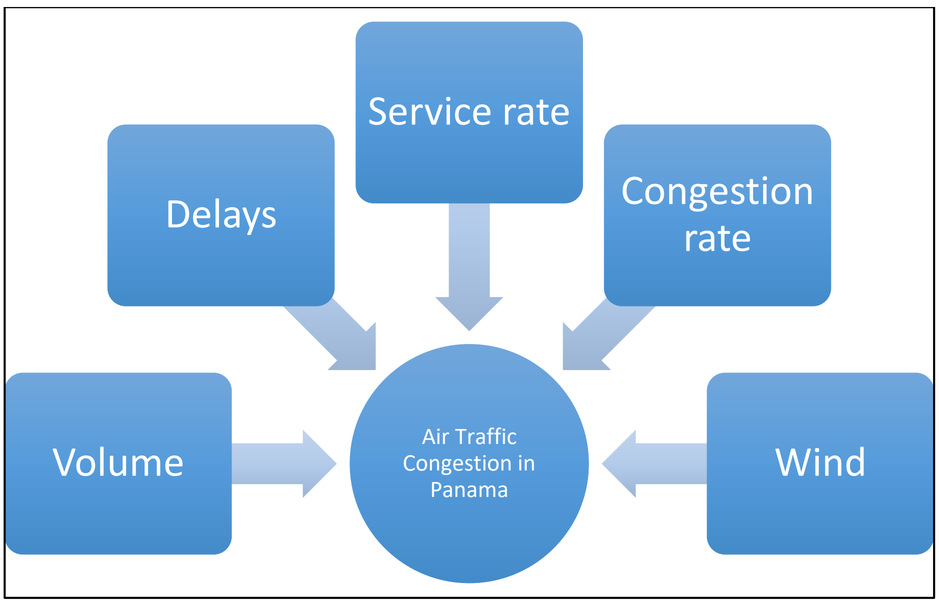


Figure 29. Second Cause and Effect Diagram

Figure 29 shows the description of the second Ishikawa model based strictly on the data available at the moment. The first assumption to construct the model is that the Airspace of Panama is a System. The second assumption to construct the model is that the extra mileage is a type of waste, like Lean Method manifest. The third assumption is that the extra mileage increases the level of utilization of the system that produce congestion in the landing and departure. The fourth assumption is that even though we are trying to describe at some level which factors can affect the extra mileage is still uncertain which factors can affect the extra mileage in the future in daily basis operation.

Consequently, we propose to measure the congestion based on the extra mileage produced by the operation per day. We are going to name this factor of congestion as "Congestion Uncertainty Factor" (CUF), since we are also trying to measure the unknown that produce an effect of extra mileage on the daily operation. The CUF is going to be measure based on the extra mileage rate. There is a minimum distance that an aircraft can achieve during an operation per route. The question of the extra mileage is what will happen if all the operations can address the historical minimum mileage per route per day.

The equation of the CUF is as follows:

Equation 14. Congestion Uncertainty Factor

$$CUF = \frac{\sum_{j=1}^{m}(minMR_j * \sum_{i=1}^{n} O_{ij})}{\sum_{k=1}^{m} M_k * O_k}$$

Where:

CUF is the Congestion Uncertainty Factor or rate

$minMR_j$ is the minimum number miles per route j

m is the number of routes

n is the number of operations at route j

O_{ij} is the number of operations i in a route j

M_k is the number of miles at the operation k

O_k is the number of operations k

k is the number of operations per day

Therefore, the Ishikawa model tries to address which factors can be responsible of the extra mileage that take each aircraft to flight on a specific route. In other words, the second Ishikawa model tries to describe, from the available data, which factors can possibly affect the production of extra mileage per route in a daily basis.

Equation 15. Multiple Regression Model for the Second Ichicagua Model

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \boldsymbol{\varepsilon}$$

Table 13. Second Ishikawa Model Notation

Variable	**Name**
X1	Number of Operations per day
X2	Number of delays
X3	Service rate
X4	Congestion rate
X5	Number of operations affected by high turbulence
ε	Random Error
Y	Extra Nautical Miles rate

Table 13 is the Multiple Regression model description, from mathematical notation is important to mention that there are some assumptions to respect on this model, such as normality assumption and constant variance.

The independent variables are the number of operations per day, the number of delays, the service rate, the number of operations in the 6 bank hub operation or congestion rate, and the number of operations affected by high turbulence in wind.

E. *FIRST LINEAR MODEL*

The Multiple Linear Regression mathematical model is as follows:

Equation 16. MLR

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon_i$$

for i = 1,…, a

Where ε_i are iid N(0,σ^2)

Y_i = is the ith response of the 5 variables presented

x_1=Independent Variable of Number of Operations per day

x_2= Independent Variable of Number of Delays

x_3= Independent Variable of Service rate

x_4= Independent Variable of Congestion Rate

x_5= Independent Variable of Number of operations affected by high turbulence

β_n= It is the least Square estimate from 0 to 5.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.95298	0.19060	18.75	<.0001
Error	132	1.34149	0.01016		
Corrected Total	137	2.29446			

Root MSE	0.10081	R-Square	0.4153
Dependent Mean	0.10784	Adj R-Sq	0.3932
Coeff Var	93.48300		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.15635	0.30004	3.85	0.0002
x1	1	-0.01685	0.01234	-1.36	0.1746
x2	1	0.01227	0.00143	8.57	<.0001
x3	1	0.32493	0.29730	1.09	0.2764
x4	1	-0.35424	0.20619	-1.72	0.0881
x5	1	0.00379	0.00210	1.80	0.0739

Figure 30. Analysis of Variance of the Multiple Linear Regression model

Ho: The MLR model is statistically significant at 0.01 alpha value.

Hi: The MLR model is not statistically significant at 0.01 alpha value.

This figure 30 shows the Analysis of variance of the model, so it seems that there is low variability, since the p-value is lower than 0.01 alpha value. So, we fail to reject the H_o, and the MLR model is statistically significant.

However, the R-square and the adjusted R-Square seems low. Also, the p-values of each independent variable shows that x_1 and x_2 are not statistically significant at 0.01, or 0.05 or 0.1 alpha value, which means that are candidates of further evaluation.

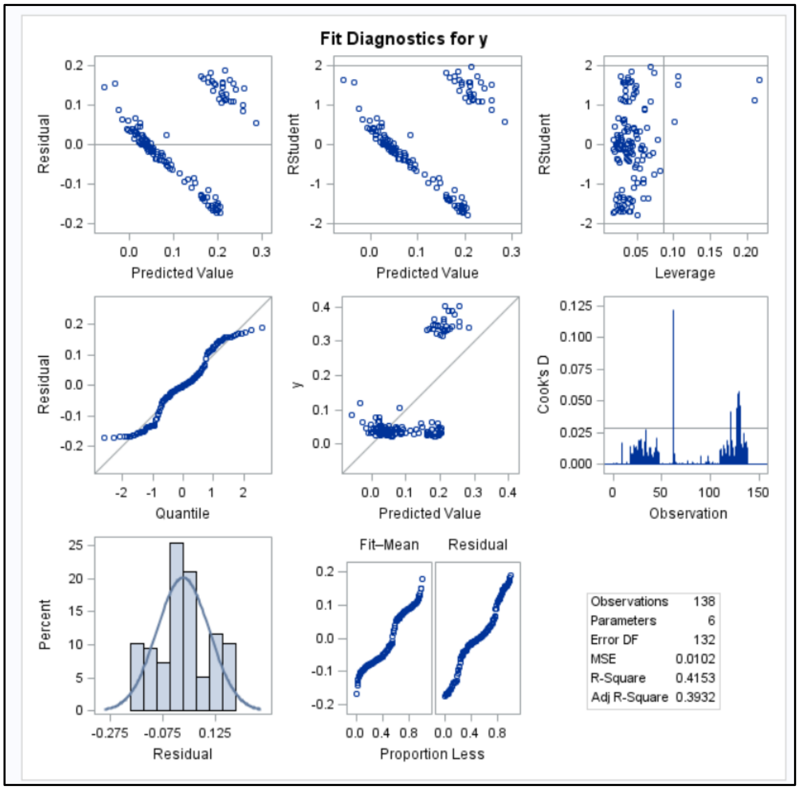


Figure 31. Analysis of Variance of the Multiple Linear Regression model

After the diagnostic of the y, we can see that the residual vs predicted value graph presented does not shows constant variance, which is one of the assumptions of the MLR models. Also, the Normal Probability plot does shows an ok normality, not perfect, but good enough.

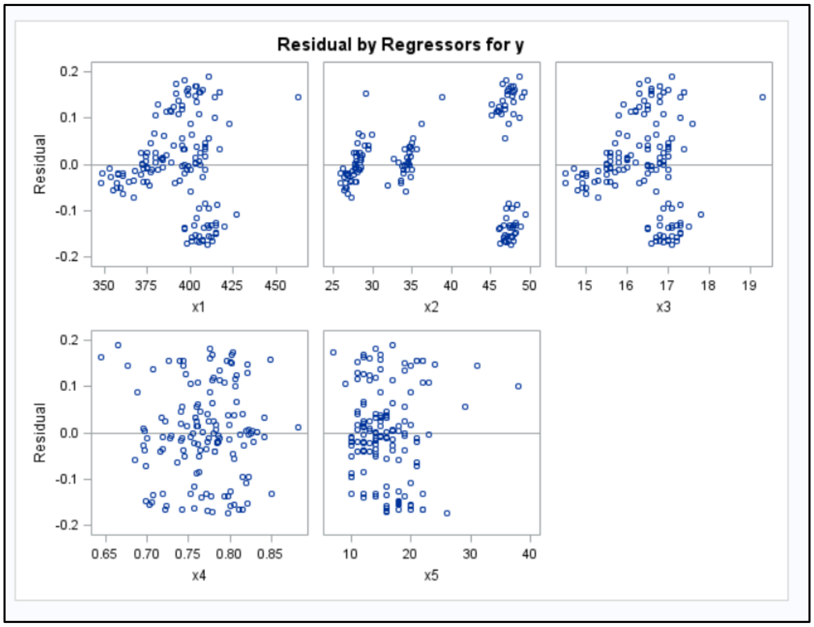


Figure 32. Analysis of Variance of the Multiple Linear Regression model

The plot of residuals versus each independent variable shows funnel image for x_1, x_2 and x_3. In other words, the residual vs x_1, x_2 and x_3 does not show constant variance. In contrast, the plots of residuals vs x_4 and x_5 it shows that the points are highly scatter that means that there no violation of the assumption of constant variance.

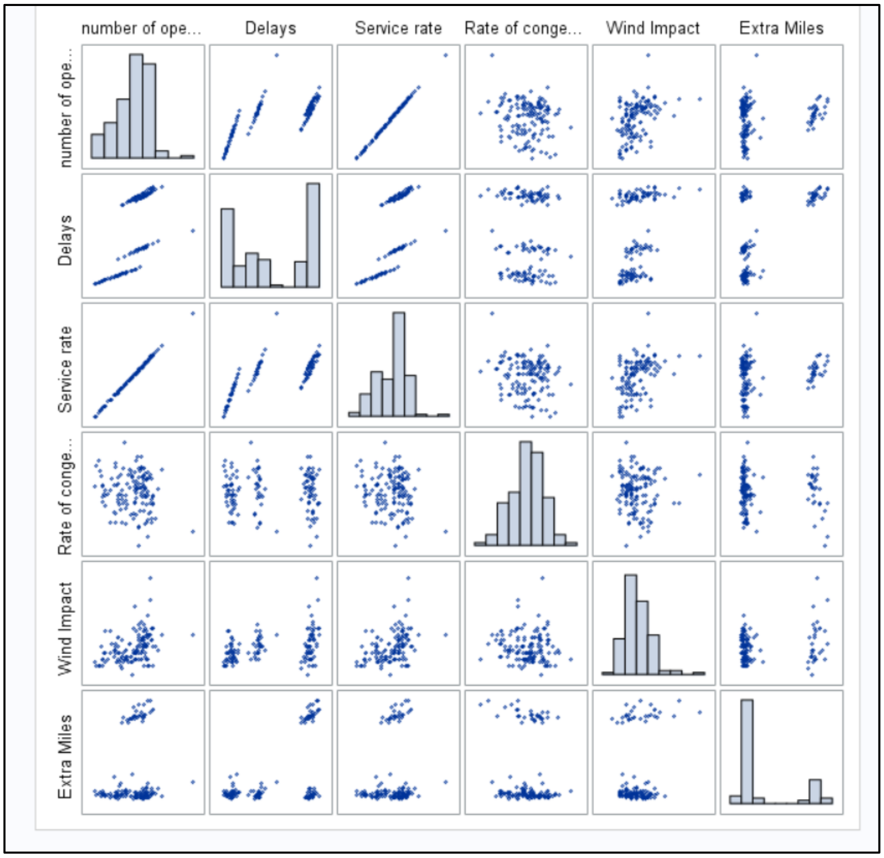


Figure 33. Correlation Matrix

Figure 19 shows the order of the variables, the first 5 are the independent variables and the last one is the dependent variable, extra miles. The extra mileage vs each independent variable it does not shows a strong linear relationship between variables, but shows two group of scatter points that each of them can be a separate linear model.

Thus, even with the MLR model with good p-value as a whole model there is not, high linear correlation between independent variables and response variable. Therefore, we prefer to move to construct another model that can include categorical variables.

F. THIRD ISHIKAWA MODEL

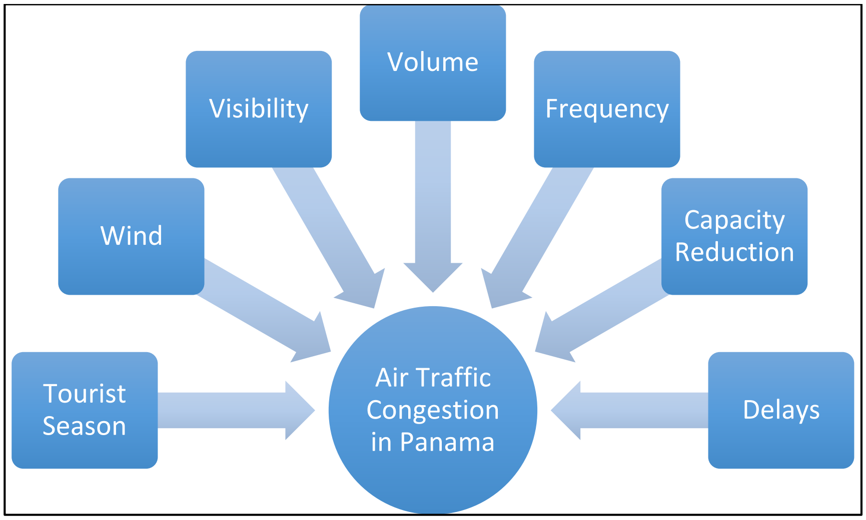


Figure 34. Third Cause and Effect Diagram

The figure 34 shows the description of the third Ishikawa that include one categorical variable and other 6 numerical variables. The categorical variable is the same categorical variable we took in consideration in the first DACE model that we did based on the simulation, which is the volume season or tourist season. The seasons categorize as high season and low season. The other 6 numerical variables are the average wind speed per day, the average visibility per day in feet, the volume of operations, the frequency of the operations per hour, the capacity reduction based on a maximum capacity of 60 operations per hour and the average number of delays per day(85).

Equation 17. Third Ishikawa Model Formulation

$$Y = X1 + X2 + X3 + X4 + X5 + X6 + X7 + \varepsilon$$

Table 14. Third Ishikawa a Model Notation

Variable	Name	Metrics
X1	Weather Season	Categorical Variable based on weather season, in Panama they have raining season and summer.
X2	Wind	Wind speed in knots
X3	Visibility	Number of miles of visibility
X4	Volume	No of Operations per hour per day
X5	Frequency	Average Frequency in a hour of each operation
X6	Capacity Reduction	Rate of operations over the maximum capacity of the airport per day
X7	Delays	Number of delayed operations per day
ε	Error	
Y	Extra Mileage Rate	The sum of the extra mileage per route per day per operation over the total number of miles per day.

Also, table 14 shows the basic linear model that is the mathematical expression of the cause and effect diagram. However, there will be another expression in the ANOVA linear model for the statistical analysis.

Furthermore, there are five assumptions to take in consideration for the construction of this Ishikawa model. The first assumption to construct the model is that the Airspace of Panama is a System. The second assumption to construct the model is that the extra mileage that an aircraft takes to flight on a route is a type of waste, like Lean Method manifest. The third assumption is that the extra mileage increases the level of utilization of the system that produce congestion. The fourth assumption is factors that affect the production of the extra mileage that takes a flight to assess route is still uncertain. The fifth assumption is that the Ishikawa model is an approximation to address the uncertainty of the factors that affect the extra mileage in the daily operation.

Consequently, we propose to measure the congestion based on the sum of extra mileage produced by the operation per day divided over the total sum of mileage produced by the operation in a day. As we state in the second Ishikawa model, we are going to name the factor of congestion as Congestion Uncertainty Factor (CUF), since we are also trying to measure the unknown that produce an effect over extra mileage on the daily operation. The CUF is going to be measured based on the extra mileage rate. There is a minimum distance that an aircraft can achieve during an operation per route. The question of the extra mileage is what happen if all the operations can address the historical minimum mileage per route per day. So the equation of 13 from the last Ishikawa model states the formula for CUF.

Therefore, the third Ishikawa model try to address which factors can be responsible of the extra mileage that takes each aircraft to flight on a specific route in a daily basis

G. SECOND LINEAR MODEL

The mathematical expression of the ANOVA linear model is presented as equation 2. For our case we did not present the model like the linear regression with qualitative and quantitative variables, since we define the model as an ANOVA model, like is the DOE models.

Equation 18. ANOVA linear model

$$Y_{ij} = \mu_{..} + x_{1i} + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + (x_1 * x_2)_i + (x_1 * x_3)_i + (x_1 * x_4)_i + (x_1 * x_5)_i + (x_1 * x_6)_i + (x_1 * x_7)_i + \varepsilon_{ij}$$

for i = 1,…, a, j=1, …., b, k=1, …., c, and t=1, ….., r.

Where ε_{ij} are iid $N(0,\sigma^2)$

Y_{ij} = j-th response observed for trt between i and the other 6 variables.

$\mu_{..}$ = is the overall mean.

x_{1i} = i-th independent variable of tourist seasons

x_2= is the independent variable of wind speed

x_3 = is the independent variable of visibility

x_4 = is the independent variable of volume

x_5 = is the independent variable of frequency

x_6 = is the independent variable of capacity reduction

x_7 = is the independent variable of delays

$(x_1 * x_2)_i$ = is the interaction effect between the categorical variable of season and the wind speed.

$(x_1 * x_3)_i$ = is the interaction effect between the categorical variable of season and the visibility

$(x_1 * x_4)_i$ = is the interaction effect between the categorical variable of season and the volume

$(x_1 * x_5)_i$ = is the interaction effect between the categorical variable of season and the frequency

$(x_1 * x_6)_i$ = is the interaction effect between the categorical variable of season and the capacity reduction

$(x_1 * x_7)_i$ = is the interaction effect between the categorical variable of season and the delays.

1) Analysis of Variance

The SAS System

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	2.41370988	0.18566999	701.20	<.0001
Error	167	0.04421958	0.00026479		
Corrected Total	180	2.45792946			

R-Square	Coeff Var	Root MSE	y Mean
0.982009	17.72642	0.016272	0.091797

Figure 35. ANOVA of the Second Linear Model

Ho: The ANOVA model is statistically significant at 0.01 alpha value.

Hi: The ANOVA model is not statistically significant at 0.01 alpha value.

H_o of this section is the first hypothesis of the Research, which is as follows:

Null Hypothesis 1: Equation 1 is statistically significant at 0.05 alpha value.

Alternative Hypothesis 1: Otherwise.

Consequently, the model shows, at figure 23, that the p value is less than 0.01 alpha value, so we fail to reject Ho and the ANOVA model is statistically significant at 0.01 level. So, we can say that we avoid type I error.

In addition, the R square value is higher than the first linear model, but we did add more variables, and the R square always increase when we add more variables.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x1	1	2.40087013	2.40087013	9067.14	<.0001
x2	1	0.00000780	0.00000780	0.03	0.8639
x3	1	0.00071065	0.00071065	2.68	0.1033
x4	1	0.00044995	0.00044995	1.70	0.1942
x5	1	0.00020710	0.00020710	0.78	0.3778
x6	1	0.00001424	0.00001424	0.05	0.8169
x7	1	0.00244105	0.00244105	9.22	0.0028
x2*x1	1	0.00388590	0.00388590	14.68	0.0002
x3*x1	1	0.00029104	0.00029104	1.10	0.2960
x4*x1	1	0.00290506	0.00290506	10.97	0.0011
x5*x1	1	0.00103998	0.00103998	3.93	0.0491
x6*x1	1	0.00000346	0.00000346	0.01	0.9091
x7*x1	1	0.00088351	0.00088351	3.34	0.0695

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x1	1	0.00000086	0.00000086	0.00	0.9546
x2	1	0.00098639	0.00098639	3.73	0.0553
x3	1	0.00032186	0.00032186	1.22	0.2718
x4	1	0.00006053	0.00006053	0.23	0.6332
x5	1	0.00078976	0.00078976	2.98	0.0860
x6	1	0.00000038	0.00000038	0.00	0.9698
x7	1	0.00084777	0.00084777	3.20	0.0754
x2*x1	1	0.00207175	0.00207175	7.82	0.0058
x3*x1	1	0.00011282	0.00011282	0.43	0.5148
x4*x1	1	0.00006921	0.00006921	0.26	0.6099
x5*x1	1	0.00086775	0.00086775	3.28	0.0720
x6*x1	1	0.00000045	0.00000045	0.00	0.9671
x7*x1	1	0.00088351	0.00088351	3.34	0.0695

Figure 36. Type 1 and type 3 regression

Figure 36 shows that at the type 3 regression p value of x_1, x_3, x_4 and x_6 are not statistically significant at 0.1 alpha value.

2) ***Residual Analysis***

a) Constant Variance Assumption

Table 15. Residual Analysis plots (Yhat vs Residuals and Y vs Residuals)

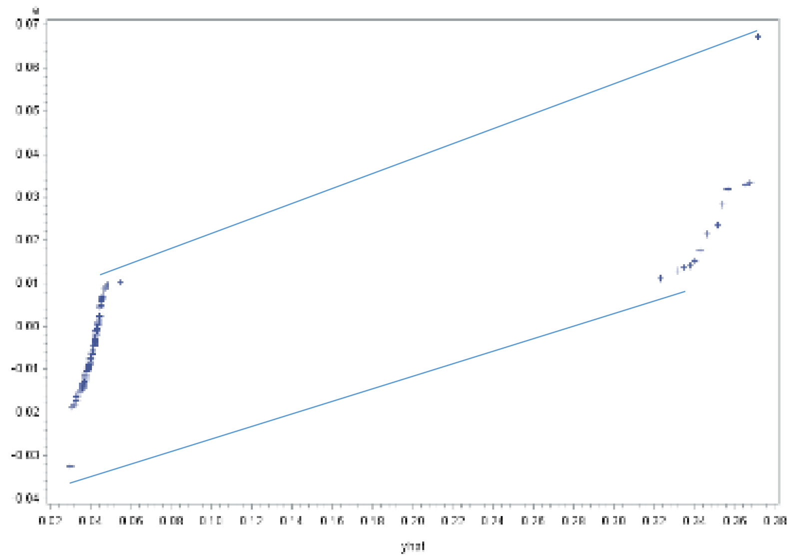
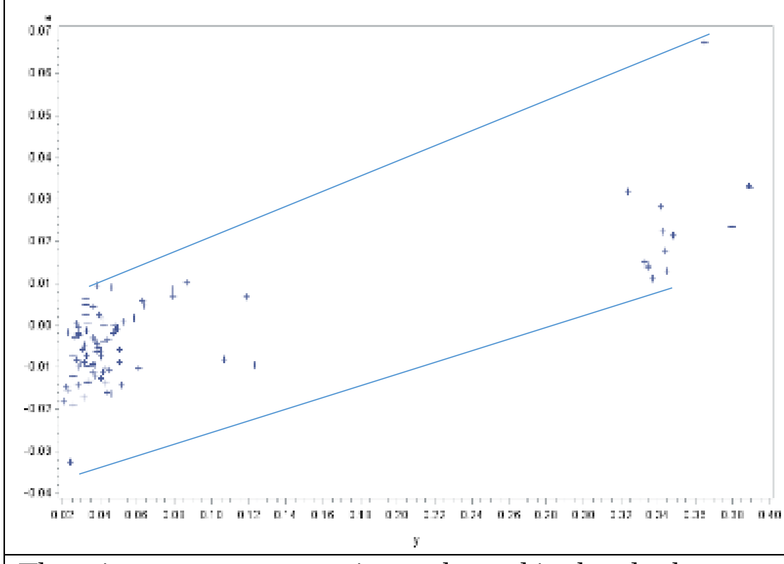
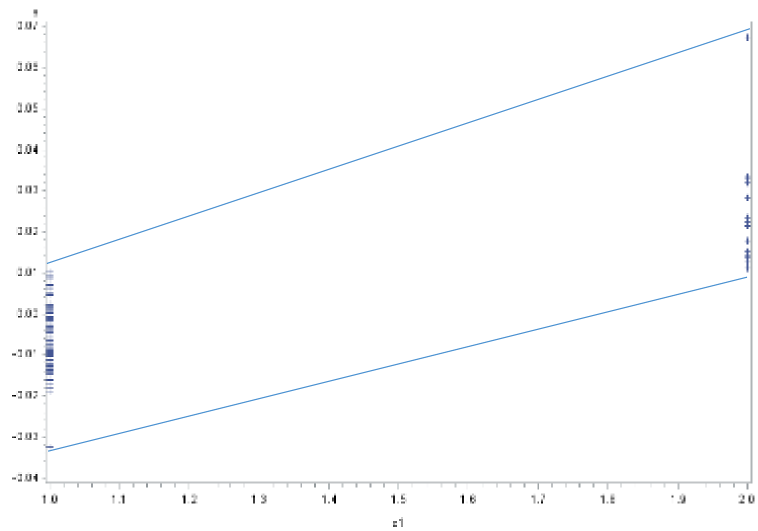
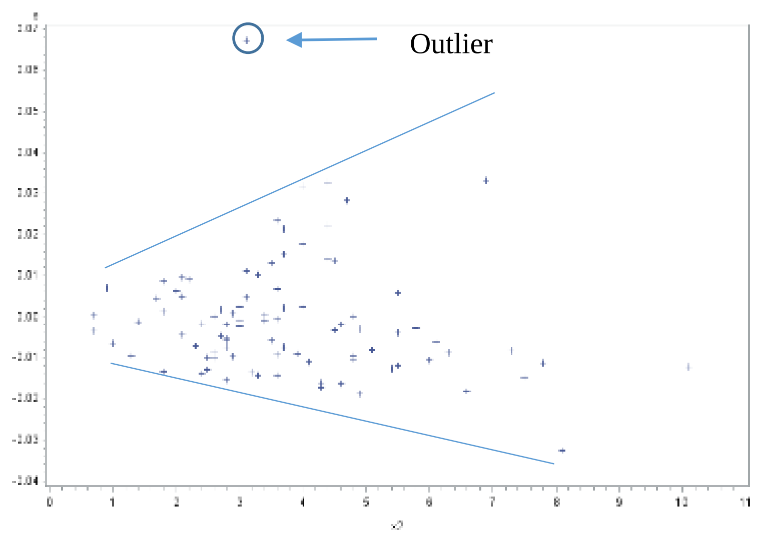

There is none constant variance showed in the plot between residuals and ŷ. It also reflects outliers.

There is none constant variance showed in the plot between residuals and y values. It also reflects outliers.

Table 16. Residual Analysis plots (x_1 vs Residuals and x_2 vs Residuals)

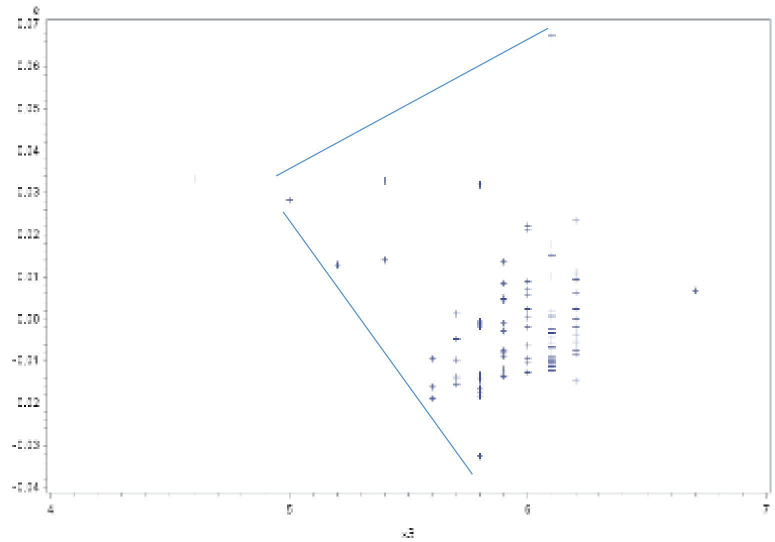


There is none constant variance showed in the plot between residuals and x1. It also reflects outliers.

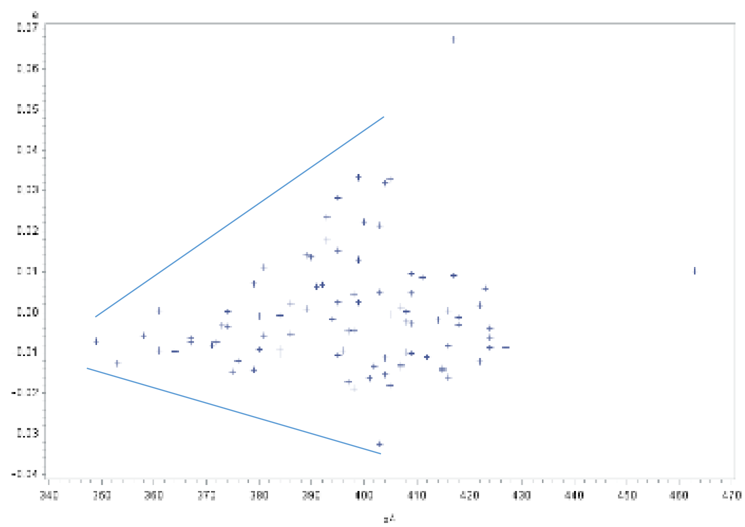


The residual vs x_2 shows none constant variance. I t also shows an outlier at the top of the plot.

Table 17. Residual Analysis plots (x_3 vs Residuals and x_4 vs Residuals)

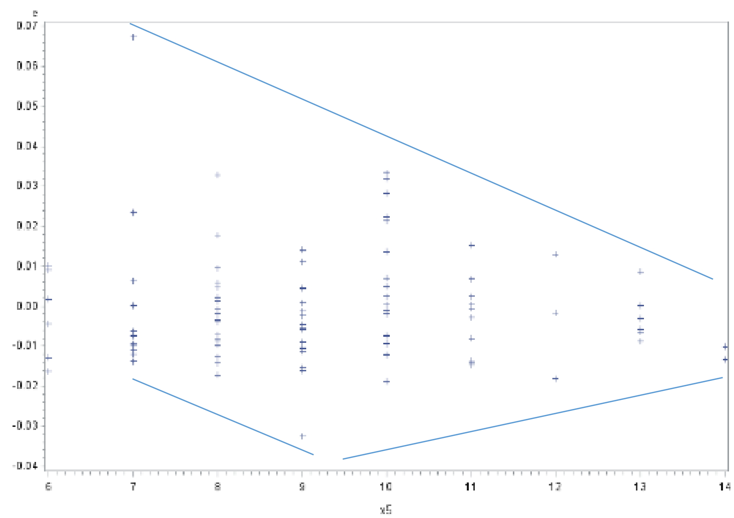


The plot of x_3 vs residuals does not show constant variance. The plot shows some outliers, but if we take them, the plot can be more scatter than the previous plots.

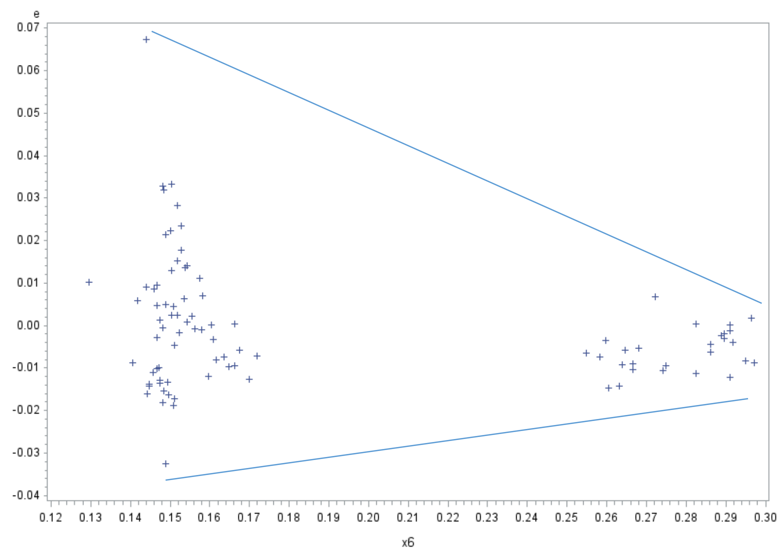


The plot of x_4 vs residuals does not show constant variance. Almost all the plots at this point shows a funnel shape. The plot shows some outliers, far at the top of the plot and at the far right.

Table 18. Residual Analysis plots (x_5 vs Residuals and x_6 vs Residuals)

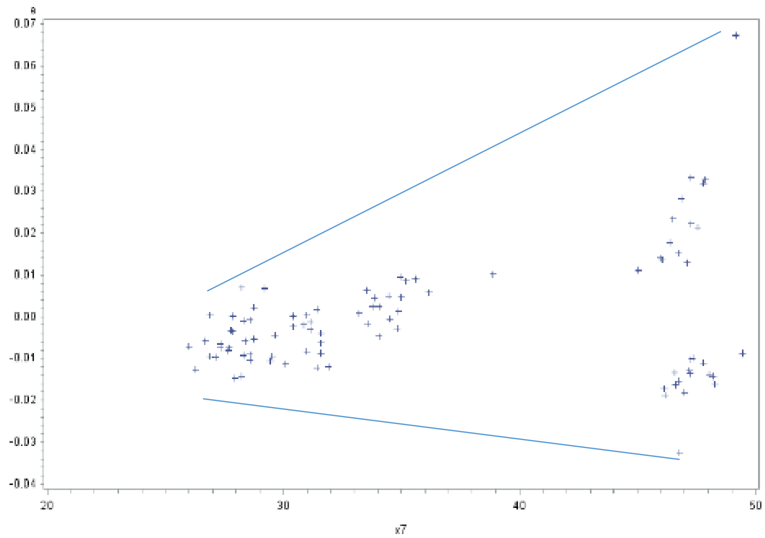


The plot of x_5 vs residuals does not show constant variance. Almost all the plots at this point shows a funnel shape. The plot shows some outliers at the top and the bottom if take out those two outliers, we can have a better spread in the plot.



The plot of x_6 vs residuals does not show constant variance. Almost all the plots at this point shows a funnel shape. The plot shows some outliers at the top and the bottom if take out those two outliers, we can have a better spread in the plot.

Table 19. Residual Analysis plots (x_5 vs Residuals and x_6 vs Residuals)



The plot of x_7 vs residuals does not show constant variance. Almost all the plots at this point shows a funnel shape. The plot shows some outliers at the top and the bottom if take out those two outliers, we can have a better spread in the plot as well as the last two plots.

b) *Normal Distribution Assumption*

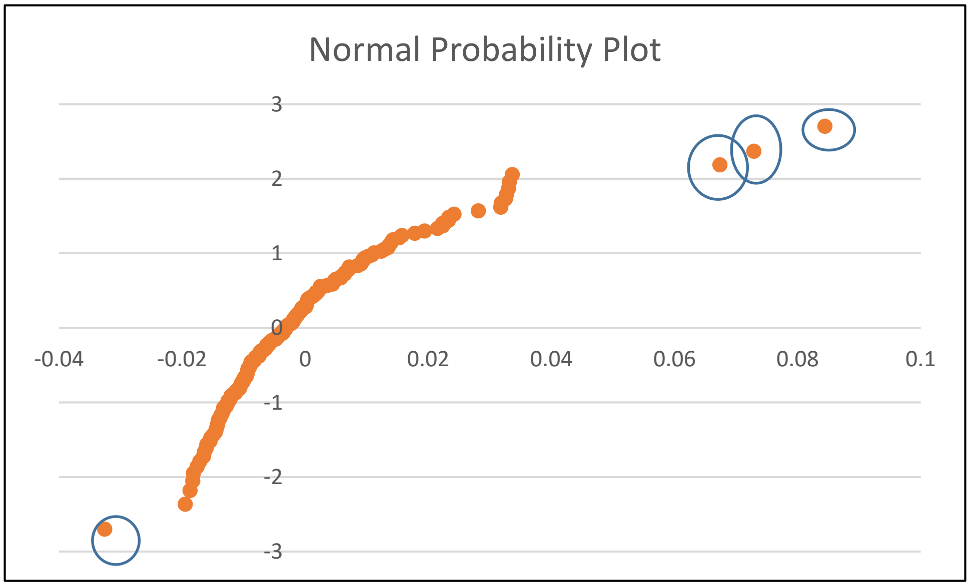


Figure 37. The Normal Probability Plot

The Normal probability plot in figure 37 shows that the normality is ok, especially if we take those 4 outliers out of the graph.

c) Correlation Analysis

The SAS System

The CORR Procedure

2 Variables:	e enrm

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
e	90	-0.0007230	0.01431	-0.06507	-0.03255	0.06740
enrm	90	1.16667	0.37477	105.00000	1.00000	2.00000

Pearson Correlation Coefficients, N = 90
Prob > \|r\| under H0: Rho=0

	e	enrm
e	1.00000	0.76754 <.0001
enrm	0.76754 <.0001	1.00000

Figure 38. Correlation Procedure.

The correlation analysis shows from figure 38 that the correlation is above 75% which is good for a model.

After the lack of a perfect normality and the lack of constant variance we conclude that the model may need a transformation of the y in order to address the assumptions of a linear model.

d) Model transformation

(1) Log y transformation.

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	43.46463257	3.34343327	18.65	<.0001
Error	167	29.93868235	0.17927355		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.592134	-38.70539	0.423407	-1.093923

Figure 39. ANOVA of the transformation of the y to Log y

H_0: First model is negligible

H_1: First model is not negligible

The F value of the first model is 701.2 The F value of the converted model with logy is 18.65. If value of the new model with logy is greater than the first model, then we reject H_o. In this case we fail to reject H_o.

(2) Inverse of y transformation

The SAS System

The GLM Procedure

Dependent Variable: yinv

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	17411.07417	1339.31340	20.19	<.0001
Error	167	11078.46535	66.33812		
Corrected Total	180	28489.53952			

R-Square	Coeff Var	Root MSE	yinv Mean
0.611139	34.36062	8.144821	23.70394

Figure 40. ANOVA of the transformation of the y to 1/y

H_0: logy model is negligible

H_1: logy model is not negligible

The F value of the converted model with logy is 18.65. The F value of the converted model 1/y is 20.19. If F value of the inverse of y model is greater than the first model, then we reject Ho. The F value of the y inverse model is greater than the first model, then we reject Ho, so we keep the log y model.

(3) Ln y transformation

The SAS System

The GLM Procedure

Dependent Variable: lny

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	125.1140907	9.6241608	99.97	<.0001
Error	167	16.0769721	0.0962693		
Corrected Total	180	141.1910628			

R-Square	Coeff Var	Root MSE	lny Mean
0.886133	-10.68646	0.310273	-2.903422

Figure 41. ANOVA of the transformation of the y to lny

H_0: logy model is negligible

H_1: logy model is not negligible

The F value of the converted model with logy is 18.65. The F value of the ln y model is 99.97. If value of the new model with ln y is greater than the first model, then we reject Ho. The F value of the ln y model is greater than the log y model, then we reject Ho, so we accept we keep log y.

(4) Square Root of y Transformation

The SAS System

The GLM Procedure

Dependent Variable: sqrooty

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	3.90071129	0.30005471	264.94	<.0001
Error	167	0.18913689	0.00113256		
Corrected Total	180	4.08984818			

R-Square	Coeff Var	Root MSE	sqrooty Mean
0.953755	12.79303	0.033653	0.263061

Figure 42. ANOVA of the transformation of the y to $y^{1/2}$

H_0: logy model is negligible

H_1: logy model is not negligible

The F value of the converted model with logy is 18.65. The F value of the square root of y model is 264.94. If value of the new model with square root of y model is greater than the log y model, we reject Ho. The F value of the square root of y model is greater than the log y model, we reject Ho, so we keep log y model.

e) Model Selection by Additives

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	43.46463257	3.34343327	18.65	<.0001
Error	167	29.93868235	0.17927355		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.592134	-38.70539	0.423407	-1.093923

Figure 43. ANOVA of the transformation of the y to log y

As we stated before, the selected model was the model with transformation of log y, even with a lower R-square than the first model offers to fix the normality assumption and the constant variance assumption for the model. Therefore, we need to test the model based on the interactions between the categorical variable of seasons and the numerical variables.

(1) Test of x7*x1 interaction

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.42786048	3.61898837	20.28	<.0001
Error	168	29.97545443	0.17842532		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.591634	-38.61372	0.422404	-1.093923

*Figure 44. ANOVA of the model without x_7*x_1 interaction*

H$_0$: Full model is negligible

H$_1$: Full model is not negligible

The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H$_0$. Based on the new ANOVA, in figure 31, the F value of the new model lies 20.28, and the F value of the old model was 18.65. Therefore, we reject H$_0$ and we keep the full model.

(2) Test of x6*x1 interaction

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.27234398	3.60602867	20.11	<.0001
Error	168	30.13097093	0.17935102		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.589515	-38.71376	0.423499	-1.093923

*Figure 45. ANOVA of the model without x_6*x_1 interaction*

H$_0$: Full model is negligible

H$_1$: Full model is not negligible

The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H$_0$. Based on the new ANOVA table, in figure 45, the F value of the new model is 20.11, and the F value of the old model was 18.65. Therefore, we reject H$_0$ and we keep the full model.

(3) Test of x5*x1 interaction

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.38722088	3.61560174	20.24	<.0001
Error	168	30.01609403	0.17866723		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.591080	-38.63989	0.422690	-1.093923

*Figure 46. ANOVA of the model without x_5*x_1 interaction*

H$_0$: Full model is negligible

H$_1$: Full model is not negligible

The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H$_0$. Based on the new ANOVA the F value of the new model is 20.24, and the F value of the old model was 18.65. Therefore, we reject H$_0$ and we keep the full model.

(4) Test of x4*x1 interaction

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.34170654	3.61180888	20.18	<.0001
Error	168	30.06160837	0.17893815		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.590460	-38.66917	0.423011	-1.093923

Figure 47. ANOVA of the model without x_4*x_1 interaction

H_0: Full model is negligible

H_1: Full model is not negligible

The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H_0. Based on the new ANOVA the F value of the new model is 20.18, and the F value of the old model was 18.65. Therefore, we reject H_0 and we keep the full model.

(5) Test of x3*x1 interaction

The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.46150580	3.62179215	20.32	<.0001
Error	168	29.94180912	0.17822505		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.592092	-38.59204	0.422167	-1.093923

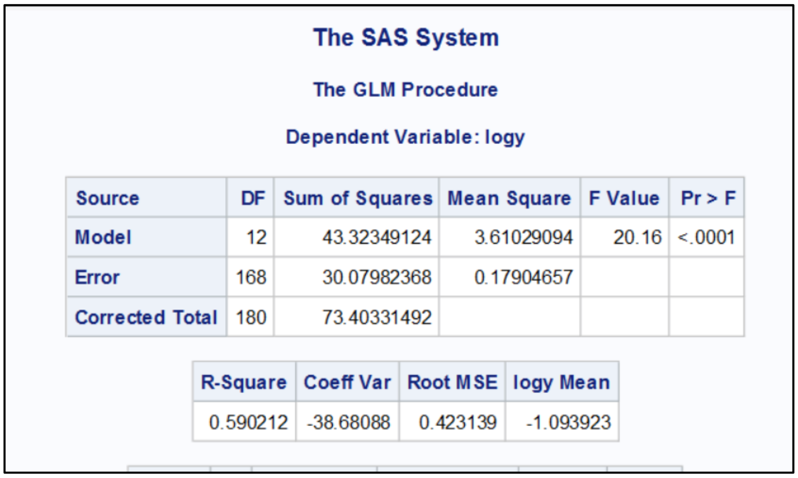
Figure 48. ANOVA of the model without x_3*x_1 interaction

H_0: Full model is negligible

H_1: Full model is not negligible

The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H_0. Based on the new ANOVA the F value of the new model is 20.32, and the F value of the old model was 18.65. Therefore, we reject H_0 and we keep the full model.

(6) Test of x2*x1 interaction



The SAS System

The GLM Procedure

Dependent Variable: logy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	43.32349124	3.61029094	20.16	<.0001
Error	168	30.07982368	0.17904657		
Corrected Total	180	73.40331492			

R-Square	Coeff Var	Root MSE	logy Mean
0.590212	-38.68088	0.423139	-1.093923

*Figure 49. ANOVA of the model without x_2*x_1 interaction*

H_0: Full model is negligible

H_1: Full model is not negligible

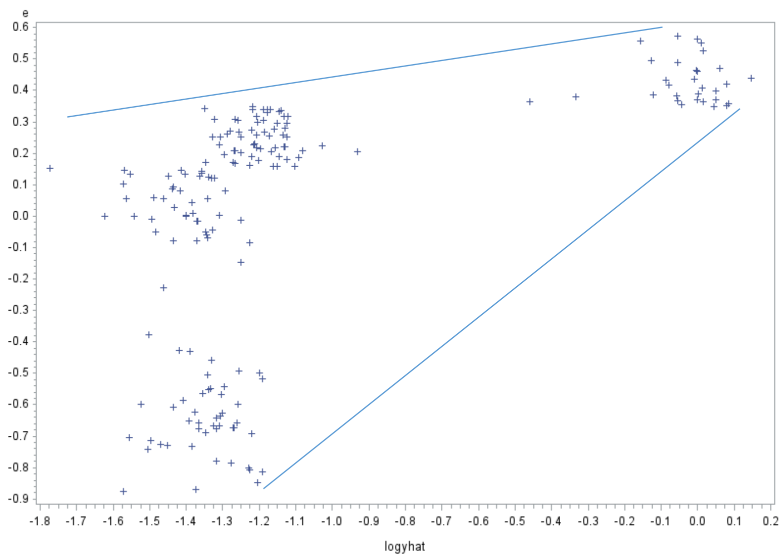
Based on the new ANOVA the F value of the new model is 20.16, and the F value of the old model was 18.65. Therefore, we reject H_0 and we keep the full model. The DoE procedure said that if the F value of the new model is greater than the F value of the old model we reject H_0.

3) Analysis of the Linear model assumptions for the new model logy.

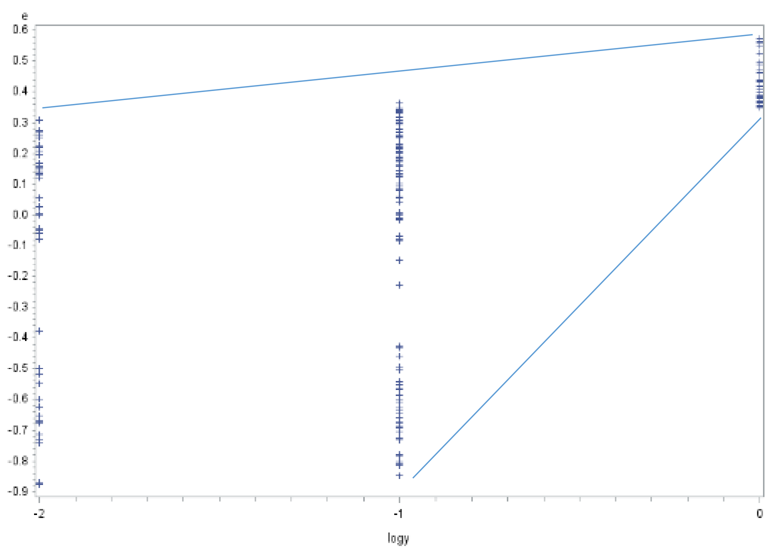
After the selection of the transformation of the y to log y, there is some analysis to perform. This section covers the evaluation of the constant variance assumption, the normality assumption and the correlation between y and the independent variables. Thus, the following sections will explain the results of those analysis.

a) ***Constant Variance Assumption***

Table 20. Residual analysis plots (log $\hat{y}$ vs residuals and logy vs residuals)



The variance spread of the logyhat vs residuals shows, that there is still some lack of constant variance between two groups of the sample. In other words, it seems that the small portion represent the variability of the high season and the down spread shows the variability of the low season, which is the majority of the interactions and sample.



Between -2 and – 1 the variance is constant, but it changes when it comes to 0. There is not constant variance in the model, but is better than the previous model without transformation.

Table 21 Residual analysis plots (x_1 vs residuals and x_2 vs residuals)

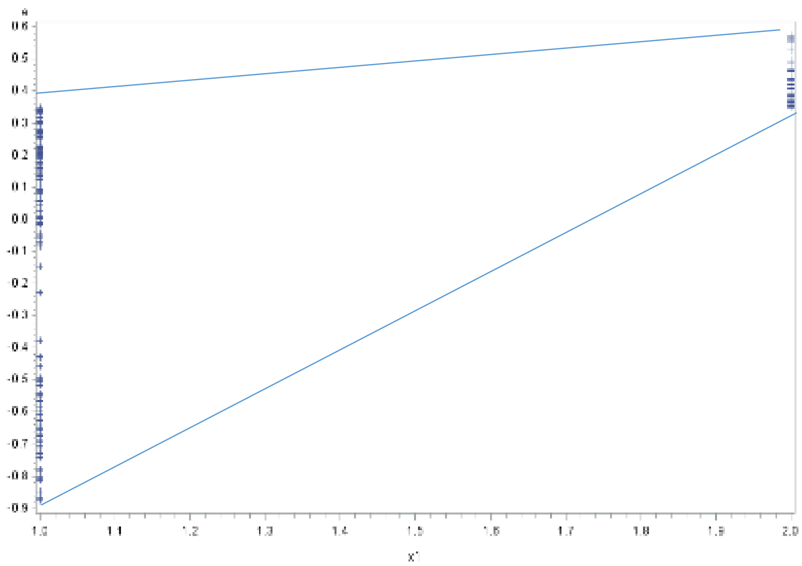
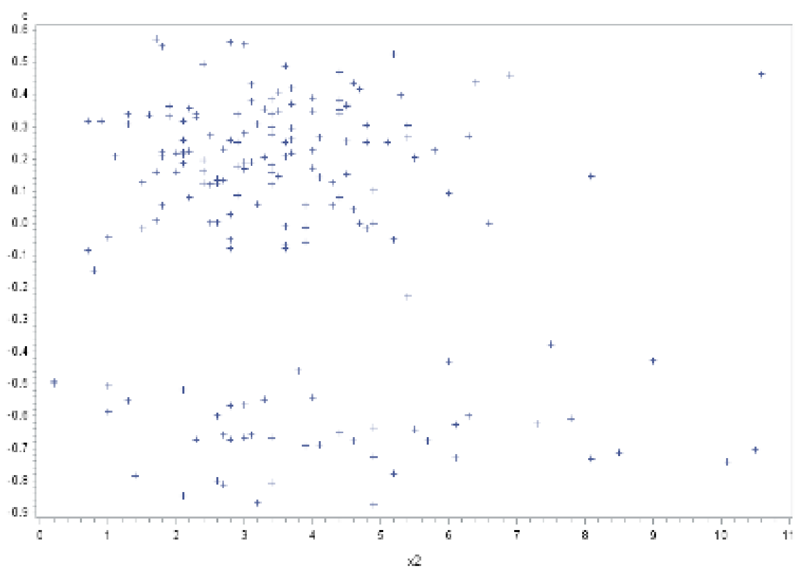
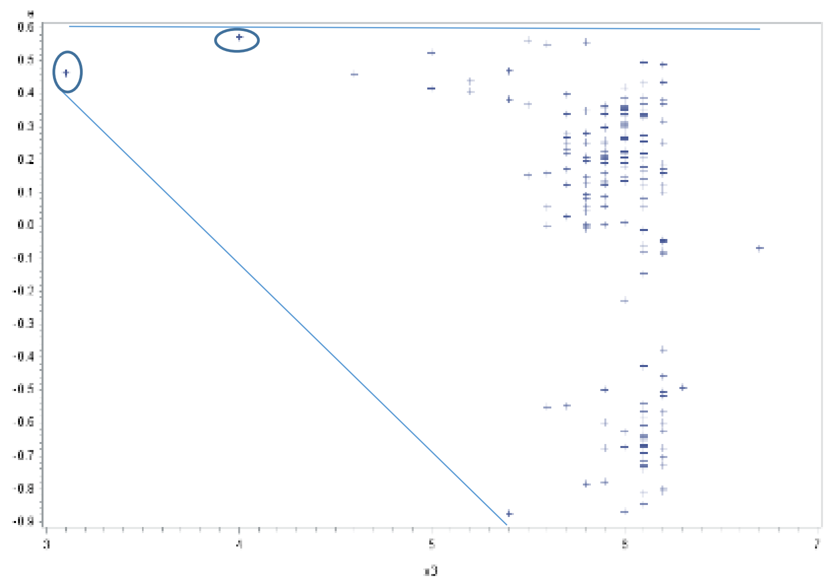
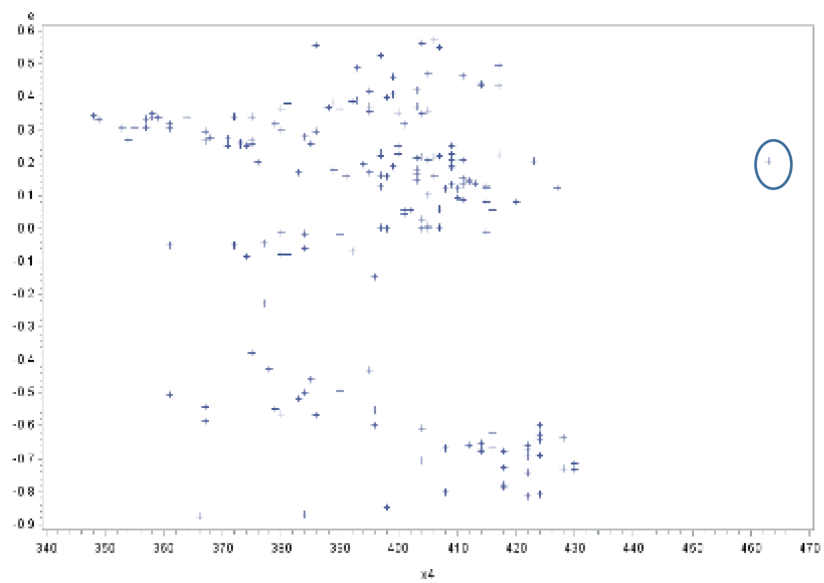

The two group of scatter points shows that the spread at each group is not similar, so x_1 vs residuals shows not constant variance.

The x_2 vs residuals plot shows a really good spread of the points that leads to the conclusion that the constant variance assumption is not violated by x_2.

Table 22. Residual analysis plots (x_3 vs residuals and x_4 vs residuals)

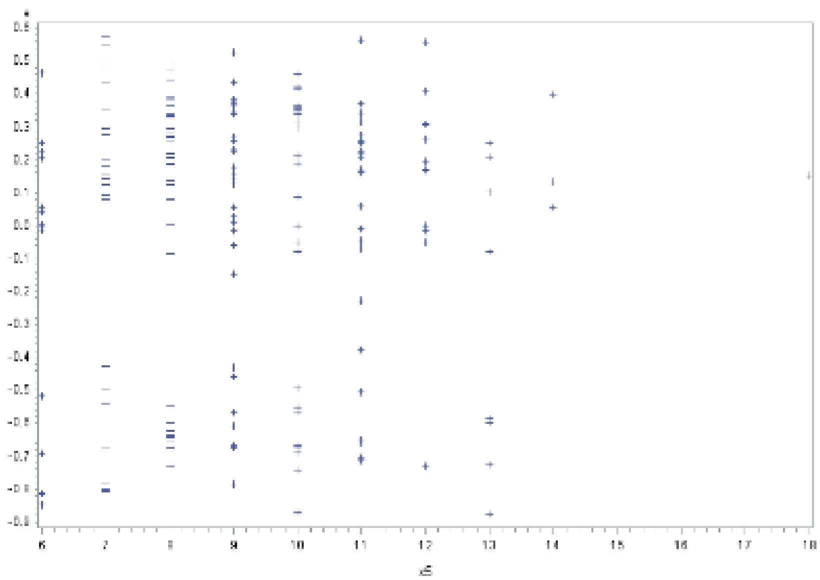


The plot of residuals vs x_3 shows that there are two outliers. If we take the two outliers out the plot the points can look better spread. At this point there is not constant variance.

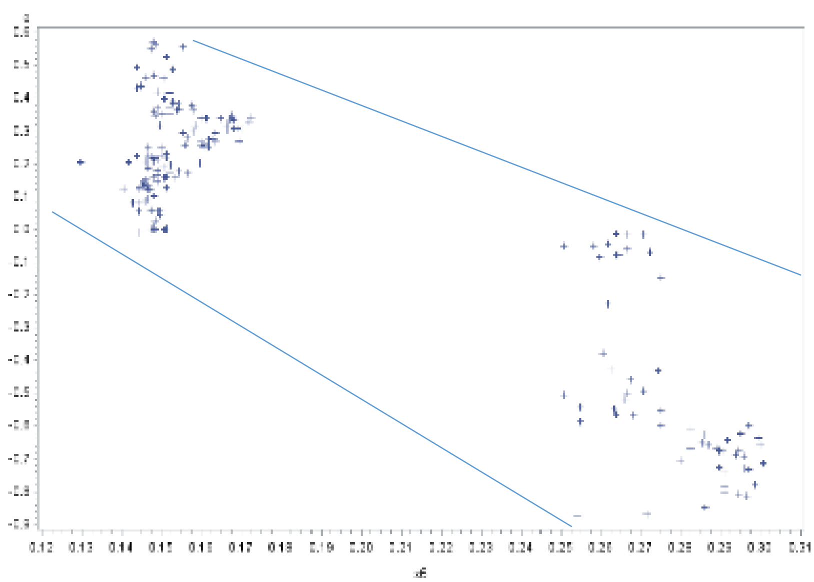


The plot of residuals vs x_4 shows a good scatter over the points, so the constant variance is not violated. It looks that there still an outlier from the plot.

Table 23. Residual analysis plots (x_5 vs residuals and x_6 vs residuals)

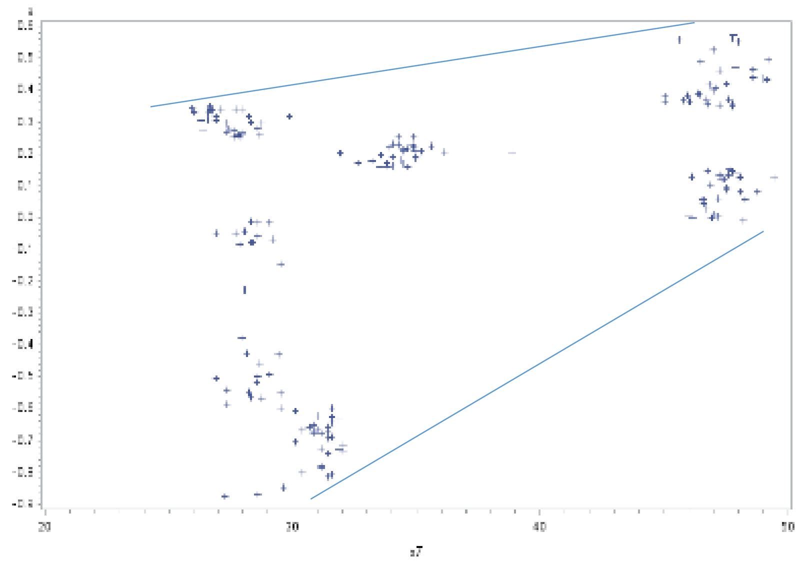


The plot of residuals vs x_5 the constant variance assumption is not violated, except for the last column at 14. Even though we think that the variance is ok.



The plot of residuals vs x_6 shows that the constant variance is violated, since the plot is not well scatter over the area.

Table 24. Residual analysis plots (residuals vs x_7)



The plot of residuals vs x7 shows a funnel shape that leads to the conclusion that the constant variance assumption is violated.

b) Normal Distribution Assumption

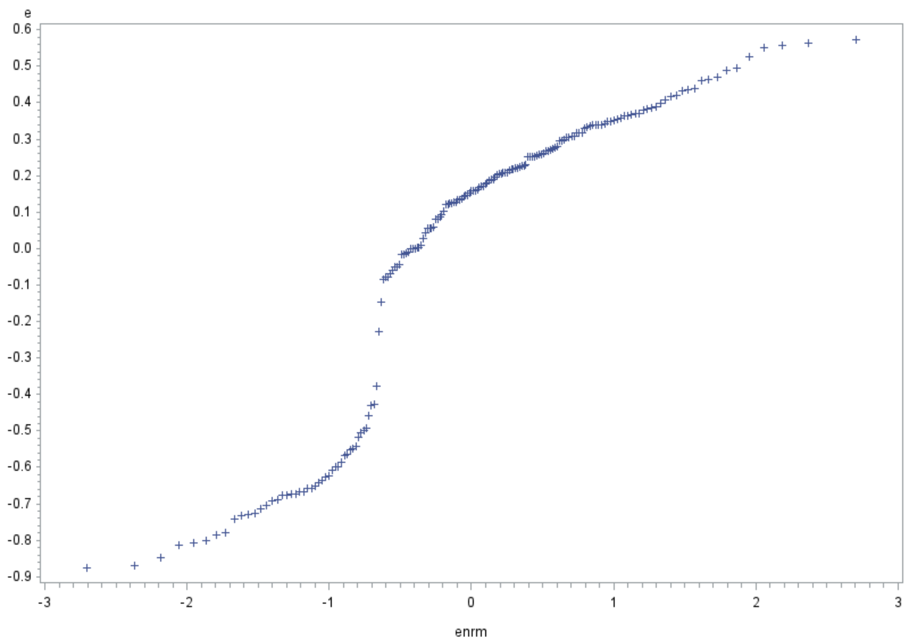


Figure 50. Normal Probability Plot

The normal probability plot from figure 50 shows that the normality is better than before, but it is not perfect, there is a long left tale, that seems to show skewness.

c) Correlation Analysis

The SAS System

The CORR Procedure

2 Variables:	e enrm

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
e	181	-5.5249E-8	0.40783	-0.0000100	-0.87521	0.57330
enrm	181	0	0.99329	0	-2.70180	2.70180

Pearson Correlation Coefficients, N = 181 Prob > |r| under H0: Rho=0		
	e	enrm
e	1.00000	0.92867 <.0001
enrm	0.92867 <.0001	1.00000

Figure 51. Normal Probability Plot

Figure 51 shows the correlation analysis, the correlation is 0.92867, which is very good. Thus, we conclude that the Normality assumption is not violated in from the log y model.

H. OPTIMIZATION AND SPC CHARTS

The results of the optimization are part of the analysis to address the hypothesis 3 which is trying to address the optimization formulation and hypothesis 4 which is trying to address the optimization model. The control charts are a testing methodology to measure the performance of the optimization. Thus, the results of the optimization and the control charts is trying to define the impact of the historical variables over the congestion in the operation.

1) Optimization

The null hypothesis 4 presents the testing of the Unified Optimization Method applied to Vehicle routing in the following way.

Null Hypothesis 4: Unified Optimization Method Applied to Vehicle Routing is the preferred model to minimum value of Nautical Miles.

Alternative Hypothesis 4: Otherwise

We reject the Null hypothesis, since we did not utilize genetic algorithms as this methodology does for searching method. However, we did use the Mix Integer Linear Programming approach and Heuristics as this method suggests. The difference is that we use Branch and Bound as a method of searching instead of genetic algorithms.

The objective function is the minimization of the sum of miles per day. The decision variable is the number of operations and the cost is the number of miles.

The first group of the constraints is for route selection per destination. In other words, instead of using directly the travel salesman problem we use the constraint of maximum capacity of operations per route based on the radial separation between aircrafts, for this study, we use from 3NM to 7NM. As an explanation, route is the internal route at the

Airspace of Panama and we are going to use destination as the main two airports of connection for the travel.

The second group of constraints is to set the minimum operation per day per destination. It helps to set the minimum number of operations to optimize as a control of the optimization, so the optimum value cannot go to 0 miles.

The formulation mathematical formulation is as follows:

Equation 19. Mixed Integer Linear Programming to minimize Flight Miles.

$$z_{LPN}, z_{IPN} = \min \sum_{i}^{N} \sum_{j}^{M} c_i x_{ij}$$

Subject to:

$$a_{11}x_{11} + a_{12}x_{12} + \cdots + a_{1M}x_{1M} \leq b_1$$
$$a_{11}x_{21} + a_{12}x_{22} + \cdots + a_{1M}x_{2M} \leq b_2$$
$$\ldots$$
$$a_{N1}x_{N1} + a_{N2}x_{N2} + \cdots + a_{NM}x_{NM} \leq b_N$$

$$a_{11}x_{11} + a_{21}x_{21} + \cdots + a_{N1}x_{N1} \geq b_{N+1}$$
$$a_{12}x_{12} + a_{22}x_{22} + \cdots + a_{N2}x_{N2} \geq b_{N+2}$$
$$\ldots$$
$$a_{1M}x_{1M} + a_{2M}x_{2M} + \cdots + a_{NM}x_{NM} \leq b_M$$

$$All\ x_{ij} \leq 0 \text{ \& Integers}$$

This formulation was coded in MATLAB to solve per day, the sample of 181 days per arrivals and departures. The iterations were based on the NM radio distance per aircraft.

It is important to mention that for this analysis we use the solver for Mix Integer Linear Programming. The solver includes some settings, so we used specifically the set listed at the table 25.

Table 25. Mixed-Integer Linear Programming solver settings in MATLAB

Nonnegative real.
Branch and Bound as a method of searching, we specifically use the fractional component with maximum pseudo cost.
Constraint Tolerance of 1e-9 through 1-3.
The normal cut generation
Group of 10 for the cuts
The basic Heuristic algorithm.
The maximum node of 50 and strictly positive integer that bounds the number of nodes.
Real from 1-6 through 1e-3 for integer tolerance
Strictly postie integer for the simplex algorithm.
Nonnegative real where reduced costs must exceed LP optimality.
The use of pre-processing for the solution to the relaxed linear program
Strictly positive integer that is the maximum number of nodes explores in its branch-and-bound process.
Strictly positive integer for feasible points
Best projection of node exploration.
The objective cut off of real greater than infinity.
The objective improvement threshold of nonnegative real.
The output specifying one or more functions that an optimization function calls at events.
The relative gap tolerance of real from 0 through 1.
The root LP Algorithm of Dual simplex.
Nonnegative integer that is the maximum number of simplex algorithm iterations to solve the initial programming problem.

a) Results of the Optimization

- After 5 Iterations based on Nautical Miles we kept just 3 of them, since the model did not found a feasible solution for the fourth and fifth iteration. The optimization ran 905 times for departures and 618 for arrivals. Also, the optimization was divided by days of the month based on the sample and it was divided based on the type of operation, departure or arrivals. In the case of the optimization for arrivals, the routes were 1,119 and the destinations were 265. In the case of the optimization for departures, the routes were 775 and 231 destinations

Figure 52 shows the results of the interaction for arrivals and figure 53 shows the results of the interaction for departures.

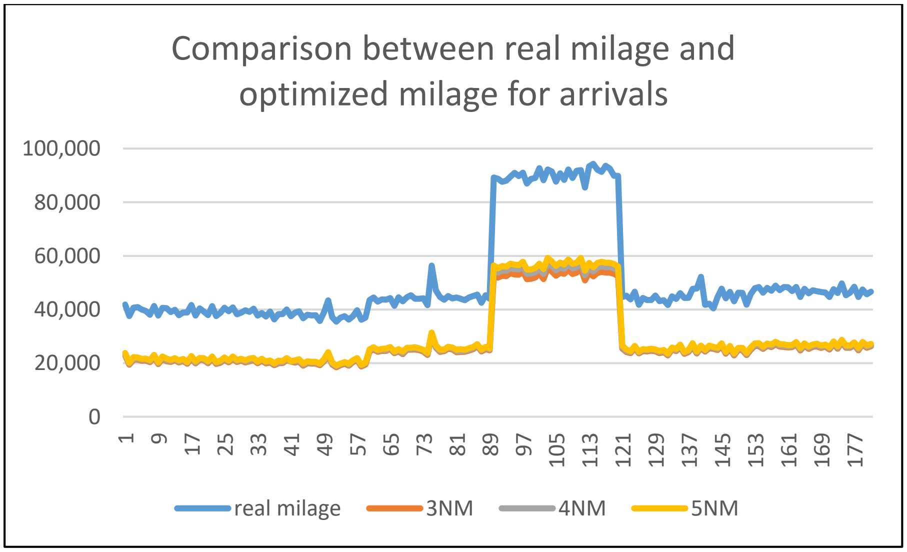


Figure 52. Comparison between optimization iterations and the reported daily mileage for arrivals

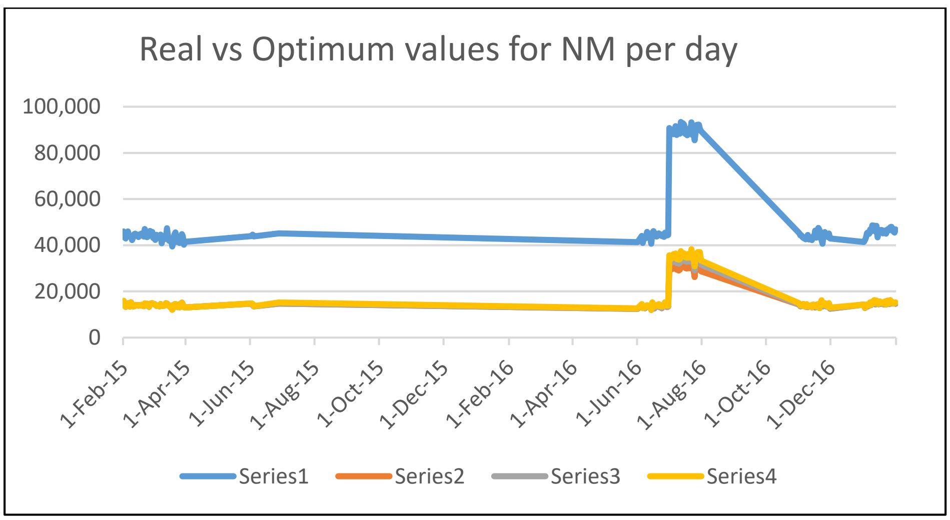


Figure 53. Comparison between optimization iterations and the reported daily mileage for departures

There are several observations based on the analysis of the data. One of them is that if we want to keep the system in control we will need to split the dataset based on months, as the behaviour of the data shows some the picks to evaluate as well. Also, the graph shows that the optimization follows similarly pattern that the original mileage, which the first line of the plot from the top to the bottom. The last observation based on figures 39 and 40 is that the optimized mileage shows similar lines through the days, except for the pick area.

2) *Sensitivity Analysis*

a) *SPC Charts*

As a measure of evaluation, we are using SPC charts as sensibility analysis per month per optimization per type of operation. The types of operations are arrivals and departures. The SPC charts used was the moving average plots for single sample line.

Table 26 . Arrivals SPC at 3NM distance between aircraft per route for February 2015

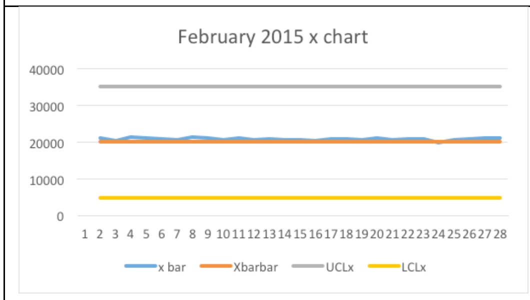
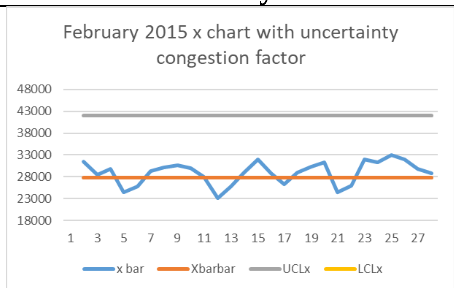
Regular SPC Charts	SPC Charts with the Congestion Uncertainty Factor
	
The performance of the operations is practically linear and in control.	The plot with CUF affect the performance increasing the variability.

Table 27. Arrivals SPC at 3NM distance between aircraft per route for March 2015

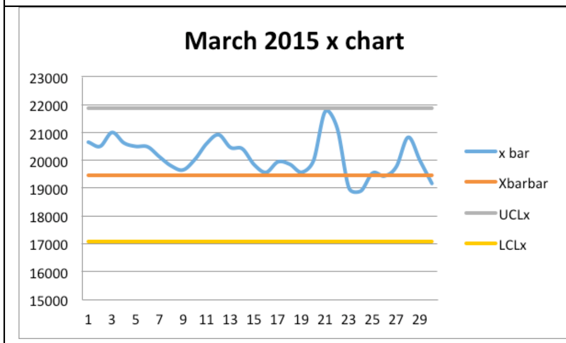
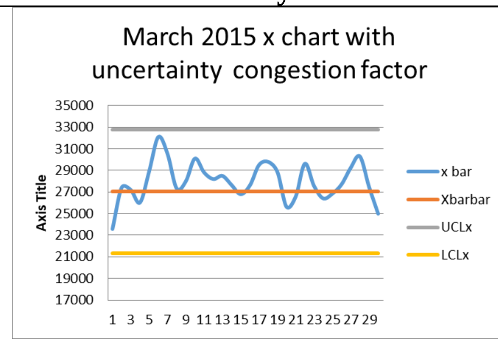
<table>
<tr><th>Regular SPC Charts</th><th>SPC Charts with the Congestion Uncertainty Factor</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>The SPC plot of March 2015 shows that the system is in control and the majority of the points are over the centre line.</td><td>The SPC chart shows that the system is still in control and it is closer of the centre line with the inclusion of the CUF</td></tr>
</table>

Table 28. Arrivals SPC at 3NM distance between aircraft per route for June 2016

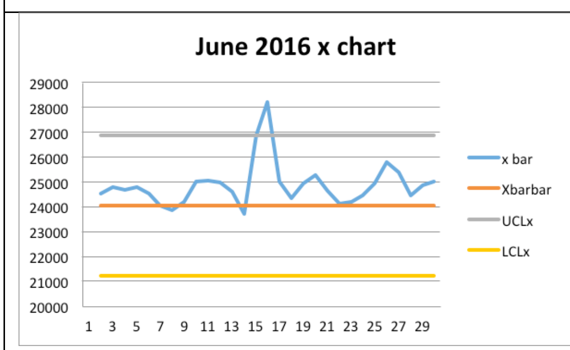
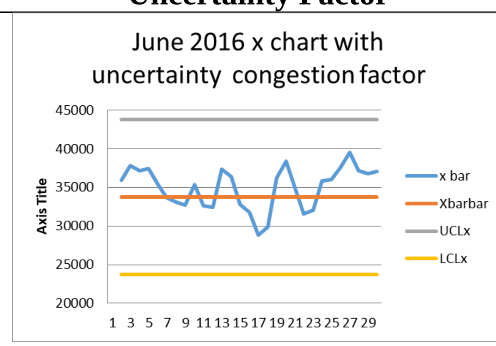
<table>
<tr><th>Regular SPC Charts</th><th>SPC Charts with the Congestion Uncertainty Factor</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>The SPC chart of June 2016 shows that there are two points out of control.</td><td>The SPC chart of June 2016 shows that the inclusion of UCF kept the system in control.</td></tr>
</table>

Table 29. Arrivals SPC at 3NM distance between aircraft per route for July 2016

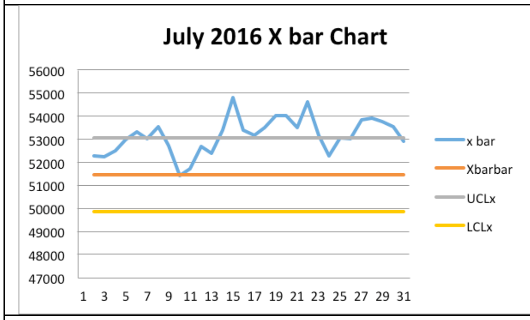
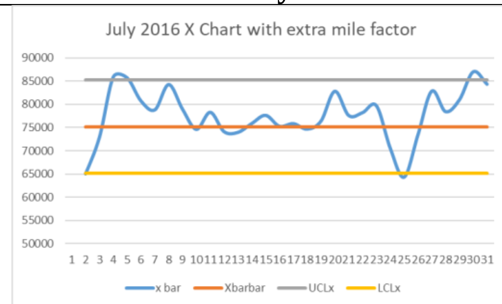
<table>
<tr><th>Regular SPC Charts</th><th>SPC Charts with the Congestion Uncertainty Factor</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>The SPC chart of July 2016 shows that the optimization is mostly out of control</td><td>The SPC chart of July 2016 with CUF shows that the optimization is in control except for two points.</td></tr>
</table>

Table 30. Arrivals SPC at 3NM distance between aircraft per route for November 2016

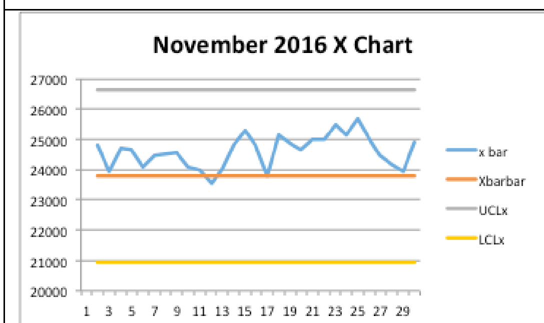
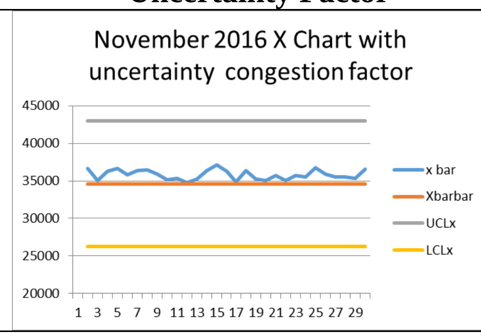
<table>
<tr><th>Regular SPC Charts</th><th>SPC Charts with the Congestion Uncertainty Factor</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>The x chart of November 2016 shows that the optimization is in control and that the majority of the points are over the mean.</td><td>The SPC chart of November 2016 with CUF decrease the variability of the system and</td></tr>
</table>

Table 31. Arrivals SPC at 3NM distance between aircraft per route for November 2016

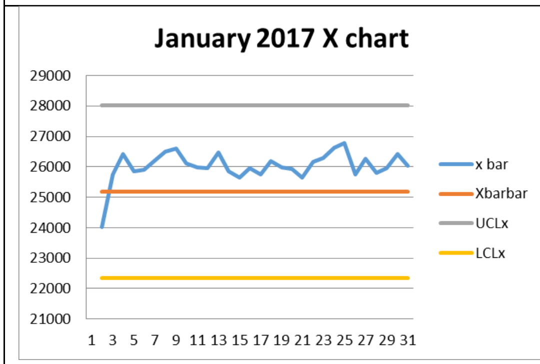
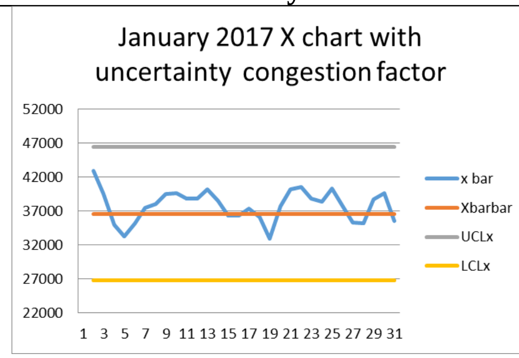
Regular SPC Charts	SPC Charts with the Congestion Uncertainty Factor
	
The SPC chart of January 2017 shows that the optimization of this month is in control and that the system is almost all above the mean value.	The SPC chart of January 2017 with CUF shows that the system is in control and the effect of the CUF over the performance is that the line is closer to the mean.

Table 32. Departure SPC at 3NM distance between aircraft per route for February 2015

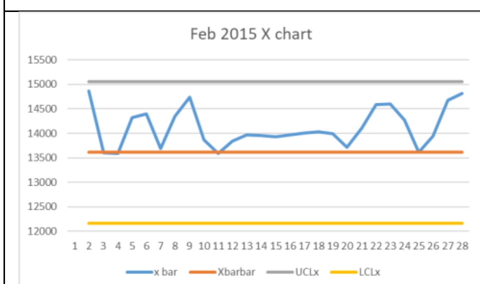
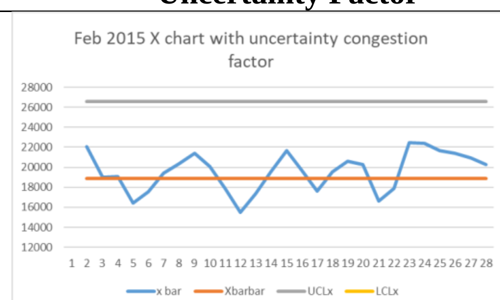
Regular SPC Charts	SPC Charts with Congestion Uncertainty Factor
	
The SPC chart of departures for February 2015 shows that the optimization is in control during the days.	The SPC chart with CUF keep the system in control and near to the centre line.

Table 33. Departure SPC at 3NM distance between aircraft per route for March 2015

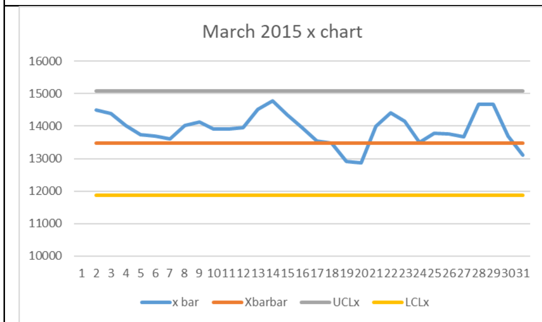
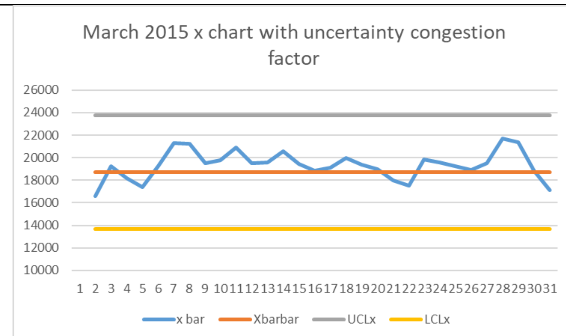
Regular SPC Charts	**SPC Charts with Congestion Uncertainty Factor**
	
The SPC chart of departures for March 2015 shows that the optimization is in control.	The SPC char with CUF keep the system in control and near to the centre line.

Table 34. Departure SPC at 3NM distance between aircraft per route for June 2016

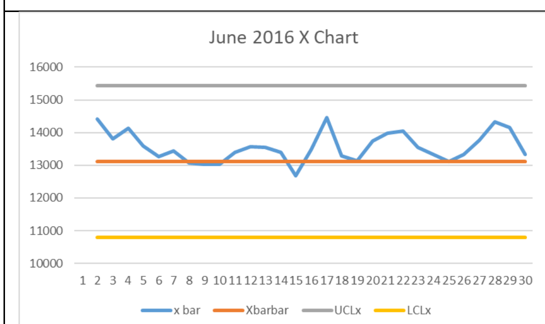
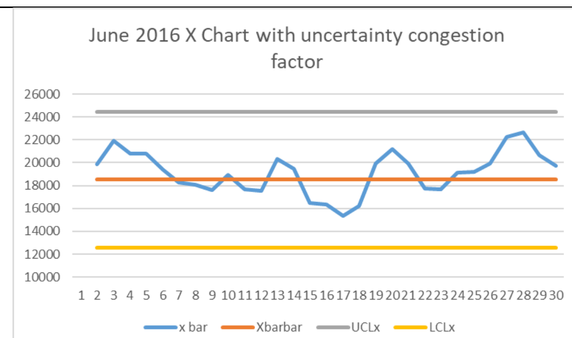
Regular SPC Charts	**SPC Charts with Congestion Uncertainty Factor**
	
The SPC chart of departures for June 2016 shows that the optimization is in control.	The SPC char with CUF keep the system in control and near to the centre line.

Table 35. Departure SPC at 3NM distance between aircraft per route for July 2016

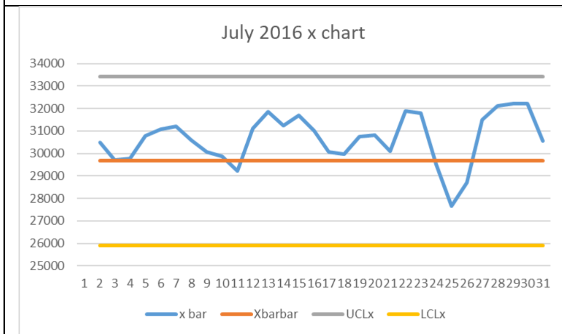
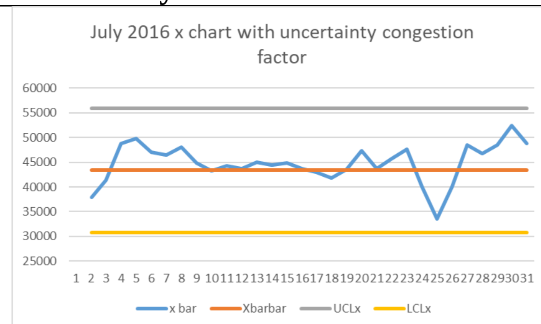
Regular SPC Charts	**SPC Charts with Congestion Uncertainty Factor**
	
The SPC chart of departures for July 2016 shows that the optimization is in control.	The SPC chart with CUF keep the system in control and near to the centre line.

Table 36. Departure SPC at 3NM distance between aircraft per route for November 2016

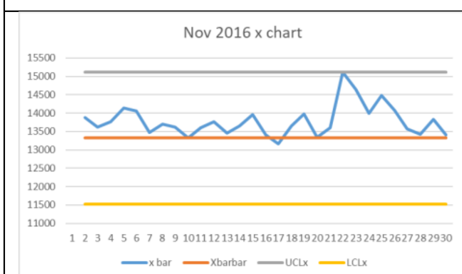
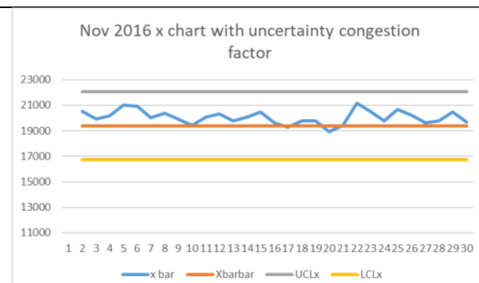
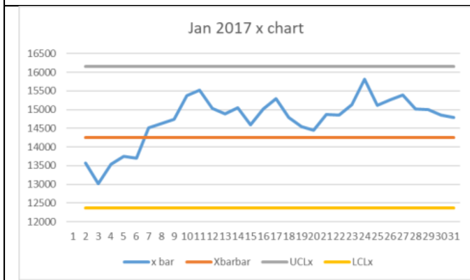
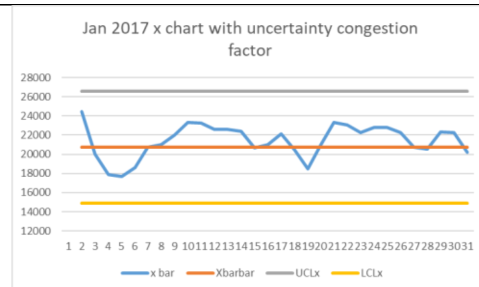
Regular SPC Charts	**SPC Charts with Congestion Uncertainty Factor**
	
The SPC chart of departures for November 2016 shows that the optimization is in control.	The SPC chart with CUF keep the system in control and near to the centre line.

Table 37. Departure SPC at 3NM distance between aircraft per route for November 2016

Regular SPC Charts	**SPC Charts with Congestion Uncertainty Factor**
	
The SPC chart of departures for January 2017 shows that the optimization is in control.	The SPC chart with CUF keep the system in control and near to the centre line.

b) Summary of the Sensitivity Analysis

The objective of the summary of the sensitivity analysis is to see the difference between the optimization without the CUF and how the impact of uncertainty increases the sum of the mileage, but keep the system below the historical average of the sum of miles per day.

Table 38. Sensitivity Analysis per month of the Arrivals

Sensitivity Analysis of the Arrivals at the Tocumen International Airport						
month	**Real Mean Mileage**	**Mean Optimum Milage at 3NM**	**Difference Real vs Optimum**	**Mean Optimum Milage with uncertainty factor**	**Difference Real vs Optimum with uncertainty**	**Reduction %**
Feb-15	39,537	20,849	**18,688**	29,042	**10,495**	**27%**
Mar-15	38,139	20,125	**18,015**	34,993	**3,146**	**8%**
Jun-16	44,520	24,875	**19,645**	34,993	**9,526**	**21%**
Jul-16	90,294	53,142	**37,152**	77,593	**12,701**	**14%**
Nov-16	44,717	24,623	**20,094**	35,859	**8,858**	**20%**
Jan-17	46,817	25,984	**20,833**	37,852	**8,965**	**19%**

Table 39. Sensitivity Analysis per month of the Departures

Sensitivity Analysis of the Departures at the Tocumen International Airport						
month	**Real Mean Mileage**	**Mean Optimum Milage at 3NM**	**Difference Real vs Optimum**	**Mean Optimum Milage with uncertainty factor**	**Difference Real vs Optimum with uncertainty**	**Reduction %**
Feb-15	44,447	14,158	**30,289**	19,685	**24,762**	**56%**
Mar-15	43,098	13,920	**29,179**	19,231	**23,867**	**55%**
Jun-16	44,011	13,582	**30,429**	19,141	**24,870**	**57%**
Jul-16	90,093	30,627	**59,466**	44,753	**45,341**	**50%**
Nov-16	43,940	13,771	**30,169**	20,053	**23,887**	**54%**
Jan-17	46,080	31,352	**14,728**	21,480	**24,600**	**53%**

In addition, the table 38 shows that the reduction of the optimization vs the real mean mileage per month of the arrivals is lower than the optimization of the departures at table 39. Consequently, there are several reason of this difference of the discrepancy in the mileage reduction between arrivals and departures. The first reason is that the airways used for departures are not the same as the airways of the arrivals. The second reason is that the 99% of the times the runways are set to land form the West-South-West and there are 3 International airports in that direction. The third reason is that there are more operations landing at the same time. The fourth reason is that there are 1119 air routes in contrast to 775 airways for departure.

I. ECONOMIC ANALYSIS AND FORECASTING

The specific objective 3 is fulfilled at this section, which is Identify the cost effectiveness of the variables on Air Traffic Congestion. The revised hypothesis 5 about Genetic Algorithms as the preferred Method for Multi-objective optimization between minimizing cost for fuel burn in contrast to maximize capacity (79) is also related to this section. However, after the analysis of the optimization there is need to do a multi-objective optimization, since the capability of the airspace is sufficient to cover the demand, but it requires adjustment to improve the efficiency. In other words, it is not relevant to do another optimization for the capacity in the airspace, when the issue with capacity is in the ground. Thus, the economic analysis is a transformation of the optimum value of the sum of miles to cost per fuel burn.

1) Economic Forecasting Analysis based on Scenarios

Dr. Jorge Beinstein presented at 2017 a workshop about three scenarios that can affect the Economy in Panama. This workshop took place in the Logistics and Transportation Strategic Planning for 2030 at the presidency of Panama. The first scenario is the dominant one summarized at table 18., the second is the alternative one summarized at table 19, and the last scenario is the one with low probability which is summarized at table 20.

Table 40. Summary of Economic Scenario 1

Economic Areas	Description
Global Economy	Economy Growth with low variability and with a moderate rate. Also, China GDP is going to be greater than the GDP of USA at 2020.
Finance System	The International banking will be stable and the replications of the crisis of 2008 totally stabilized. Also, there will be negotiations about the currencies between the US dollar and the yuan.
International Commerce	The expansion will be moderate, the commercial globalization will have the same rate of growth and the inflation will be similar to now.
Evolution of Geo Politics	Multi-polarity in control, the centralization from Asia will grow, the control of the power from the west will remains, but with less intensity, and less open wars.
International Maritime Transportation	Low variability in the growth rate, more volume in enterprises in these area, predictable seasons and predictable cost of transportation.
Latin America and the Caribbean	Low growing rate at upper trend, more regional trade agreements between the region, more integration of the region in cooperations, commercial expansion at low variability in upper trend, and stable commodities prices.
Economy of Panama	4% to 6% annual growth rate of the GDP for the following 13 years and slow decentralization of the type of incomes.

Table 41. Summary of Economic Scenario 2

Economic Areas	Description
Global Economy	Economic recession, low economic growth, and high variability in economic growth rate and more uncertainty.
Finance System	Des-acceleration in the international loans to the countries, currency crisis and down trends in the mass of global finance.
International Commerce	Commerce des-globalization, close of international trade agreements, and unstable prices of the goods.
Evolution of Geo Politics	Continuous power struggle between economic powers, war risks and gray zones in the negotiations.
International Maritime Transportation	Lack of growing, increment in companies trying to address the same services and lack of accuracy in the forecasting.
Latin America and the Caribbean	Lack of growing with a down trend in the regional GDP and increment in the uncertainty in the negotiations in the region.
Economy of Panama	Fall in the Government investment, lack of private sector investment and des-acceleration of the GDP at down trend of 2%.

Table 42. Summary of Economic Scenario 3

Economic Areas	Description
Global Economy	Economy goes back to the upper trends before the 2008 crisis.
Finance System	Financial derivate grow to 1100 Billion of dollar at 2030
International Commerce	Commercial re-globalization
Evolution of Geo Politics	Multi-polarity in control and good terms between China and USA.
International Maritime Transportation	Strong growth of the international maritime commerce with a large volume of new medium companies, smooth season cycles, and new commercial routes without a great alteration of the traditional ones.
Latin America and the Caribbean	High growing rate, as before the crisis of 2008.
Economy of Panama	Exponential rate growth of the GDP.

2) Impact of the Flight Miles over the Jet Fuel

The Jet fuel has an upper linear trend in the last two years, which is stable for forecasting. The scenario method for Forecasting is based in Macro-Economics evaluations of the GDP. In our case we are taking in consideration those scenarios stablishing the relationship with the GDP and impact on the fuel rate that William Greiner presents in the Mariner wealth advisor's webpage. The relationship that Greiner address is per each 10% increase in the jet fuel there is negative impact of 0.2% on the GDP.

Table 43. Impact of the economic Scenarios over the Jet Fuel price

Economic Scenario 1	**Economic Scenario 2**	**Economic Scenario 3**
Linear GDP Acceleration	Linear GDP Des-Acceleration	Exponential GDP Growth
$ 2.43	$ 2.29	$ 2.48
$ 2.48	$ 2.34	$ 2.69
$ 2.53	$ 2.38	$ 2.91
$ 2.57	$ 2.42	$ 3.14
$ 2.62	$ 2.47	$ 3.40

The timeline used for the analysis of the table 43 is an average jet fuel price per year for the next 5 years. Figure 54 shows the same information as table 43 in a graphical way.

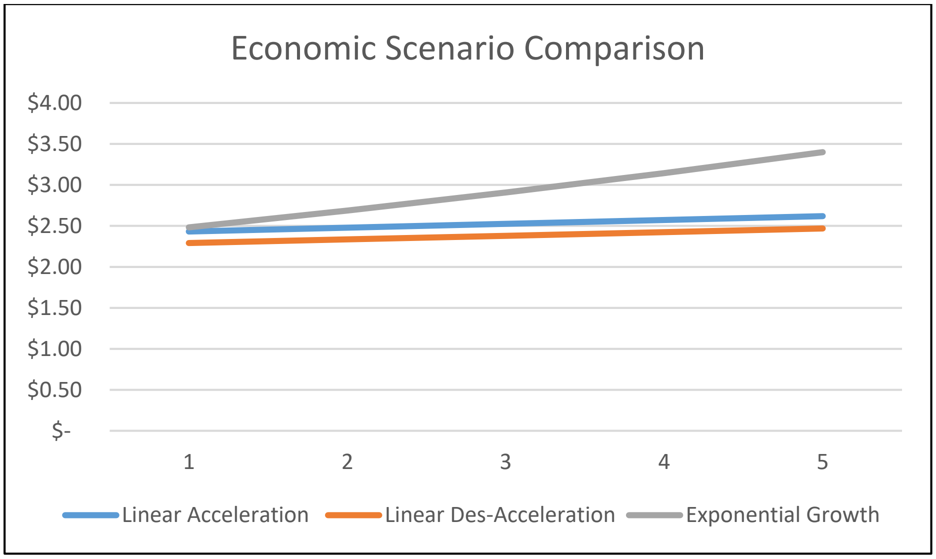


Figure 54. Jet fuel forecasting over the next 5 years

Also, figure 54 presents the mix in the air traffic based on aircraft, where the Boing 738 is representing 49% of the flights at the Tocumen International Airport with a fuel burn rate of 3.17kg/Km. However, the average consumption rate is 3.07 kg/Km with a standard deviation of 0.81.

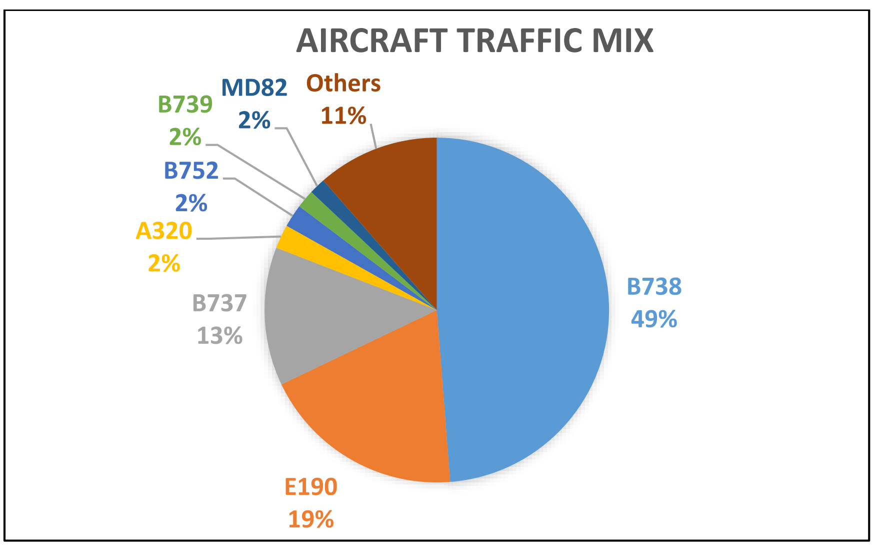


Figure 55. Air Traffic Mix by type of Aircraft

3) Summary of the Economic Analysis

Table 44. Summary of the Economic Analysis Scenario 1 / Average monthly fuel burn per year

Year	Basic in Historic WCS	Optimum based on Historic WCS	Difference Real vs Optimum	Optimum based on Historic WCS with uncertainty	Difference Real vs Optimum with uncertainty
1	315,179,532	146,364,603	168,814,929	213,767,166	168,814,929
2	321,256,679	149,186,738	172,069,941	217,888,926	172,069,941
3	327,333,825	152,008,873	175,324,952	222,010,686	175,324,952
4	333,410,971	154,831,008	178,579,964	226,132,446	178,579,964
5	339,488,118	157,653,142	181,834,975	230,254,206	181,834,975

Table 45. Summary of the Economic Analysis Scenario 2 / Average monthly fuel burn per year

Year	Basic in Historic WCS	Optimum based on Historic WCS	Difference Real vs Optimum	Optimum based on Historic WCS with uncertainty	Difference Real vs Optimum with uncertainty
1	296,996,098	137,920,492	159,075,606	201,434,445	95,561,653
2	302,722,639	140,579,811	162,142,829	205,318,411	97,404,228
3	308,449,181	143,239,130	165,210,051	209,202,377	99,246,804
4	314,175,723	145,898,449	168,277,274	213,086,343	101,089,380
5	319,902,265	148,557,769	171,344,496	216,970,309	102,931,956

Table 46. Summary of the Economic Analysis Scenario 3 / Average monthly fuel burn per year

Year	Basic in Historic WCS	Optimum based on Historic WCS	Difference Real vs Optimum	Optimum based on Historic WCS with uncertainty	Difference Real vs Optimum with uncertainty
1	321,797,256	149,437,774	172,359,482	218,255,567	103,541,689
2	348,284,513	161,738,055	186,546,458	236,220,267	112,064,246
3	376,817,057	174,988,136	201,828,920	255,572,161	121,244,895
4	407,546,552	189,258,449	218,288,103	276,414,115	131,132,437
5	440,635,602	204,624,503	236,011,099	298,856,411	141,779,191

Tables 44, 45 and 46 shows the summary of the economic analysis that includes arrivals and departures. The forecast considers the worst case scenario for the average mileage per month. The worst case scenario for the mileage was the month of July 2016 that registered the largest number of miles per day. Consequently, we are assuming that the monthly average per year is the monthly average of July 2016. This average is going to be the input for an Engineering Economic Analysis to determine he NPV for 5 years, considering each gradient per year.

Table 47. Net Present Value of each scenario per record.

Scenario	Basic in Historic WCS	Optimum based on Historic WCS	Optimum based on Historic WCS with uncertainty
1	$9,740,257,673.34	$4,528,416,880.88	$8,160,926,300.09
2	$11,577,728,494.173	$5,376,521,836.942	$6,091,335,168.934
3	$13,895,813,225.081	$6,558,833,289.929	$9,424,687,550.416

Table 47, shows the summary of the NPV per each scenario and each forecasting, based on historical Worst Case Scenario for the mileage, the Optimum WCS and the Optimum WCS with uncertainty. Our approach to this equivalent matrix is going to be the Hurwicz Criterion for uncertainty with alpha value equal to 0, which is the alpha value for the most pessimistic attitude. So, if $v(a_i, s_j)$ represents a loss in our case cost of fuel burn consumption, then the formula is as follows:

Equation 20. Hurwicz Criterion

$$min_{a_i}\left\{\alpha\, min_{s_j} v\,(a_i, s_j) + (1-\alpha) max_{s_j} v(a_i, s_j)\right\}$$

Therefore, the maximum values per column lays at the scenario 3 row, and the minimum value per row of the row is the optimum based on historic WCS. Thus, the Scenario 3 is the preferred scenario with the optimum value without the UFC factor, since it offers the lower cost or loss for fuel burn consumption.

J. Results summary

Table 48. Summary of the results

Specific Objective	Hypothesis	Conclusion
<u>Specific Objective #1:</u> *Determine which variables are most relevant to minimize Air Traffic Congestion.*	**Null Hypothesis 1:** Equation 17 is statistically significant at 0.05 alpha value. **Alternative Hypothesis 1:** Otherwise.	We fail to reject the Null Hypothesis, since the equation 1 was statistically significant at 0.05 level and the transformation of the y to log y was the alternative to accomplish the Multiple Regression Assumptions for ANOVA models.
<u>Specific Objective #1:</u> *Determine which variables are most relevant to minimize Air Traffic Congestion.*	**Null Hypothesis 2:** The Congestion Factor has an exponential effect over the minimum Nautical Miles per model. **Alternative Hypothesis 2:** Otherwise	We reject the null hypothesis, since the CUF has a logarithmic effect over the Nautical Miles.
<u>Specific Objective #2:</u> *Determine the significance of the variables and their impact on the Air Traffic Congestion.*	**Null Hypothesis 3:** One objective function can be identified from Rastrigin, Rosenbrock, Levy and Sphere equations. **Alternative Hypothesis 3:** Otherwise	We reject the Null Hypothesis since there was not a DACE model to test, we use the Worst case scenario to test capacity vs cost, since historically the capacity was challenged by the demand.
<u>Specific Objective #2:</u> *Determine the significance of the variables and their impact on the Air Traffic Congestion.*	**Null Hypothesis 4:** Unified Optimization Method Applied to Vehicle Routing is the preferred model to minimum value of Nautical Miles. **Alternative Hypothesis 4:** Otherwise	We reject the Null hypothesis, since the model utilized was a MILP with Branch and Bound and Heuristics.
<u>Specific Objective #3:</u> *Identify the cost effectiveness of the variables on Air Traffic Congestion.*	**Null Hypothesis 5:** Genetic Algorithms is the preferred Method for Multi-objective optimization between minimizing cost for fuel burn in contrast to maximize capacity (79) **Alternative Hypothesis 5:** Otherwise	We reject the Null Hypothesis, since we use the economic analysis by scenario and Engineering Economics instead of multi-objective optimization between cost minimization and maximization of the capacity.

IX. CONCLUSIONS

We conclude that the Air Traffic Congestion in Panama can be reduced through minimizing the impact of historical congestion variables such as the sum of the mileage based on airways assignation per route destination using the actual layout of airways. However, the model is not limited to the actual design of the airspace.

We strongly believe that we achieve the objective one which is to determine which variables are most relevant to minimize Air Traffic Congestion, in the case of Panama based on the three Ishikawa models. However, as we state before the model is open to understand more variables, such as air traffic mix, and the categorical variables from the first Ishikawa model. Also, we studied using Linear Regression the significance of the variables and their impact on the Air Traffic Congestion using the optimization analysis and the sensitivity analysis to evaluate the CUF impact over the MILP over 181 iterations at three levels. Finally, we Identify the cost effectiveness of the variables on Air Traffic Congestion using the theory of scenarios.

From the practical perspective the Air traffic controllers can take the model to select the preferred historical air route to minimize the mileage, since there is a lack of standardization in the practice to manage the air traffic between the air traffic controllers.

In addition, the use of the full model can decrease the existed gap between a future re-design of the airspace and the actual design of the airspace. As a matter of fact, the actual airspace can handle the more than 60 operations per hour and the bottleneck in the operation is still Tocumen International Airport Capacity. However, it is necessary an adequate assignation of air routes in order to increase the capacity and reduce risks on the operation

Therefore, the CUF can be utilized as a reference to measure congestion, since the literature review did not present that indicator specifically (2)

Furthermore, the Ishikawa models can be used for strategic purpose by the Collaboration Decision Making to study the congestion in the air traffic.

Finally, the economic analysis based on scenarios can serve to evaluate strategic planning in terms of air transportation.

X. FUTURE STUDIES

There are several small studies that we can conduct based on the previous research, the first one is being the further evaluation of the first Ishikawa collecting more data that can measure the impact of those variables and sub-variables affect the CUF. The second future research can be the analysis of other searching methods such as genetic algorithms for the optimization model. The third future study can be the evaluation of the different settings and alternatives of the solver of Mixed Integer Linear Optimization using the DACE approach. The fourth future study can be the change of formulation of the MILP to make a scheduling optimization model (72). The fifth future study can be the economic analysis with considering the different studies about rate of fuel consumption based on individual aircraft performance during the landing and takeoff (70,84). The sixth study can be considering the adaptation of the model for Artificial Intelligence to give the optimum route in real time.

XI. LIMITATIONS

The first limitation was the software's, since MATLAB was not able to import the millions of data points for data management at the beginning of the study, so we did split the data to filtering and clustering using Tableau.

The second limitation was the data availability, since the first effort of the study was to collect more data directly from the Simulation model, but it was not feasible for Copa Airlines during the 2018. The explanation that they gave was the lack of expertise of the new personnel performing analysis with the Software TAAM.

The third limitation was the data complexity, there was inaccuracy to challenge and double check, for example the assignment of flight cancellations, or errors in some indicators or records.

The fourth limitation was the change of plans for data collection from the simulation model, since I was not able to travel to Panama in 2018.

XII. PROJECT PLAN

The commitment at the proposal of the dissertation was to complete the analysis of the hypothesis and the objectives during the summer of 2018. The first objective took the majority of the time, since we decided to use the historical dataset instead that collect new data from the simulation model. The second specific objective took the month of July and part of August. In order to evaluate the model proposed we travel to Iowa State University to receive feedback from the Operations Research department and the Chair of IMSE department. Consequently, we did further studies in the MLR and we did adjust the MILP. The third objective was completed in August, which include a direct input from the Logistic Strategic Planning until 2030 that include air transportation.

Table 49. Gantt Chart

Phase of the Dissertation	**May**	**June**	**July**	**August**
• ***Specific Objective #1:*** *Determine which variables are most relevant to minimize Air Traffic Congestion.*				
• ***Specific Objective #2:*** *Determine the significance of the variables and their impact on the Air Traffic Congestion.*				
• ***Specific Objective #3:*** *Identify the cost effectiveness of the variables on Air Traffic Congestion.*				

XIII. CONTRIBUTION TO THE BODY OF KNOWLEDGE

The contributions to the body of knowledge includes the classes that was worked as a Graduate Teacher Assistant, such as: Advance Engineering Economics, Facilities Planning and Design, Engineering Probability, Operation Research, Quality Systems, Production and Inventory Control, Planning Control of Enterprise Systems, Management of Knowledge Technology

In addition, the Design for Six Sigma Research and the Multi-Objective Optimization from Dr. Erick Jones. Also, Deep Learning and Genetic Algorithms from IE 5345 Management of Knowledge and Technology of Dr. John Priest.

Furthermore, from the data analytics area, the Design and Analysis of Computer Experiment, knowledge of DOE and SAS from IE 6308 and IE 6309 the classes of DOE and Response Surface Methodology and DACE respectively from Dr. Victoria Chen. Data Mining, Data Exploration, and MATLAB knowledge from the Data Mining and Analytics class from Dr. Shouyi Wang. IE 6318. Linear Optimization from Introduction of Operations Research class of Dr. Bill Corley. IE 5301, Queueing Theory, Dynamic Programming with Dr. Corley as well, and Combinatorial Optimization with Dr. Jay Rosenberger as instructor.

XIV. APENDIX A

Table A. Research Journey Summary in a Time Line

Date	Steps
January 2015	Presentation of the Air Space Challenge to the Dean Office
June 2015	Pre Proposal of Airspace Optimization Conference
January 2016	First Data Collection in Panama
July 2016	First DACE model
June 2017	Airspace Optimization Conference in Panama
July 2017	Second Data collection
June 2018	Dissertation Proposal

XV. REFERENCES

[1] ALG Consulting Power Point Reports. 2015.

[2] Tocumen International Airport. Panama Air Space Optimization 2017. Panama, Panama.

[3] http://panamatribune.com/business/aviation-safety-audit-result-alarms-panama/).

[4] Air Traffic by the numbers, FAA, 2017

[5] http://www.telemetro.com/nacionales/AAC-aclara-incidente-control-insectos_0_868713598.html

[6] Bevilacqua, M., Ciarapica, F. E., Mazzuto, G., &, Postacchini, L. Air Traffic Management of an Airport using Discrete Event Simulation method. 2012 IEE Interantional Conference on Industrial Engineering and Engineering Management, pp. 1034-1038. ISSN. 2157-3611. DOI.10.1109/IEEM.2012.6837898. 2012.

[7] Cedeño, M., Suira, A. & Guerra de Castillo, Z. “Simulation model applied in passenger’s flow at Latin American Airport Hub”. Eleventh Latin American and Caribbean Conference for Engineering and Tehcnology.Cancun, Mexico. August 14-16, 2013.

[8] Jones, E. C., & Chung, C. A. (2008). RFID in Logistics: A Practical Introduction (2008 ed.). CRC Press.

[9] Chen, V., Tsui, K., Barton, R. &, Meckesheimer, M. “A review of design and modeling in computer experiments”. IIE Transactions, Apr 2006, Vol 38 Issue 4, pp273-291. DOI: 10.1080/07408170500232495.

[10] Current Market Outlook 2015-2034. Boeing, 2015. www.boeing.com/resources/boeingdotcom/commercial/about-our-market/assets/downloads/Boeing_Current_Market_Outlook_2015.pdf Accessed 1 June 2016.

[11] Subbu, R., Lizzi, J., Iyer, N., Jha, P., &, Suchkov, A. "MONACO – Multi-Objective National Airspace Collaborative Optimization". 2007 IEEE Aerospace Conference. Doi. 10.1109/AERO.2007.352955.

[12] Gilbo, Eugene. "Optimizing Airport Capacity Utilization in Air Traffic Flow Management Subject to Constratins at Arrival and Departure Fixes". IEEE Transactions on Control System Technology. Vol 5. No 5, September, 1997.

[13] Gelhausen, M., Berster, P. &, Wilken, D. "Do Airport capacity constraints have a serious impact on the future development of air traffic?". Journal of Air Transport Management. Vol.28. 2013, ppr5. 3-13. DOI. 10.1016/j.jairtraman.2012.12.004.

[14] Global Market forecast 2016-2035. Mapping Demand, 2016. http://www.airbus.com/company/market/global-market-forecast-2016-2035/. Accessed 10 May 2016.

[15] Hub of the Americas. Copa Airlines. 2016. https://www.copaair.com/en/web/us/hub-of-the-americas/. Accessed 15 June 2016.

[16] Ivanescu, D., Shaw, Ch. Zeghal, K., &, Hoffman, E. "Propagation of Airborne Sapcing Errors in Merging Traffic Streams". 7th USA/Europe Air Traffic Management R&D Seminar. Barcelona, Spain. July 2007.

[17] Ivanescu, D., Shaw, Ch, &, Tamvaclis, C. “Models of Air Traffic Merging Techniques: Evaluating Performance of Point Merge”. 9th AIAA Aviation Technology, Integragion and Operations Conference (ATIO). South Carolina, USA. 21-23 September 2009.

[18] Jones, E., Franca, R., Richards, C., &, Carlson, J. “Multi-objective stochastic supply chain modeling to evaluate tradeoffs between profit and quality”. International Journal of Production Economics. 2010. Vol 127, issue 2, pp. 292-299. DOI: 10.1016/j.ijpe.2009.09.005.

[19] Lee, H. &, Balakrishnan, H. “Fast-Time Simulations of Detroit Airport Operations for Evaluating Performance in the Presence of Uncertainties”. Massachusetts Institute of Techonology. Cambridge, MA. 2012. IEEE. 978-1-4673-1900-3. DOI: 10.1109/DASC.2012.6382349

[20] NextGen Works. Federal Aviation Administration, 2016. http://www.faa.gov/nextgen/. Accessed 2 May 2015.

[21] Meric, Ozlem. “Optimum Arrival Routes for Flight Efficiency”. Journal of Power and Energy Engineering. Vol. 3, 2015, pp. 449-452. dx.doi.org/10.4236/jpee.2015.34061

[22] Oglivie, Cesar. “Airspace re-design Project/PBN Concept -MPTO”. 2015. Copa Airlines.

[23] Point Merge: Improving and harmonising arrival operations with existing technology. EUROCONTROL, 2016. https://www.eurocontrol.int/services/point-merge-concept/. Accessed 12 May 2016.

[24] Simaiakis, I., &, Balakrishnan, H. “Design and Simulation of Airport Congestion Control Algorithms”. 2014 American Control Conference. June 4-6, 2014. Portland Oregon, USA. AACC 978-1-4799-3274-0.

[25] Tocumen International Airport. Technical Aspects, 2016. http://www.tocumenpanama.aero/datos-tecnicos/. Accessed 2 June 2016.

[26] Total Airspace and Airport Modeller (TAAM). “Product Profile 2015”. Jepessesen a Boeing Company.

[27] World Air Cargo Forecast 2014-2015. Boeing World Air Cargo Forecast Team 2014. www.boeing.com/commercial/cargo. Accessed 1 May 2015.

[28] https://en.wikipedia.org/wiki/Torrijos%E2%80%93Carter_Treaties

[29] Castillo, Juan. “FAA Visit to the Command Center Report”, 2016.

[30] Delahaye, D. &, Puechmorel, S. “Modeling and Optimization of Air Traffic”. John Wiley & Sons, Inc. NJ, USA. 2013. ISBN 978-1-84821-595-5.

[31] Caccavale, M., Iovanella, A., Lancia, C., Lulli, G., &, Scoppola, B. “A model of inbound air traffic: to Heathrow airport”. Journal of Air Transportation Management 34, 2014. Pgs. 116-122. Doi. Dx.doi.org/10.1016/j.jairtraman.2013.09.004

[32] Gwiggner, C., &, Nagaoka, S. “Data and queueing analysis of a Japanese air-traffic flow”. European Journal of Operation Research 235, 2014, pgs 265-275. Doi. Dx.doi.org/10.1016/j.ejor.2013.10.056.

[33] Lee, H., Simaiakis, I., &, Balakrishnan, H. “A Comparison of Aircraft Trajectory-based and aggregate queue-based control of Airport taxi processes”. MIT, 2010. 978-1-4244-6618-4.

[34] Pyrgiotis, N., Malone, K., &, Odoni, A. "Modelling delay propagation within an airport network". Transportation Research Part C 27, 2013, pgs. 60-75. Doi: 10.1016/j.trc.2011.05.017.

[35] Mota, M., Scala, P., &, Boosten, G. "Simulation-based Capacity Analysis for a Future Airport". 2014 Asia-Pacific Conference on Computer Aided System Engineering. Doi: 10.1109/APCASE.2014.6924479

[36] Cem Cetek, Ertan Cinar, Fulya Aybek, Aydan Cavcar, (2013) "Capacity and delay analysis for airport manoeuvring areas using simulation", Aircraft Engineering and Aerospace Technology: An International Journal, Vol. 86 Issue: 1, pp.43-55, https://doi.org/10.1108/AEAT-04-2012-0058

[37] Simaiakis, I., &, Balakrishnan, H. "Design and Simulation of Airport Congestion Control Algorithms". 2014 American Control Conference. June 4-6, 2014. Portland, Oregon, USA. Doi: 10.1109/ACC.2014.6859480.

[38] Lee, H., &, Balakrishnan, H. "Fast-Time Simulations of Detroit Airport Operations for Evaluating Performance in the Presence of Uncertainties". 2012. Doi: 10.1109/DASC.2012.6382349.

[39] Bennel, J., Mesgarpour, M., &, Potts, C. "Airport runway scheduling". 4OR-Q Journal Operation Research. 2011. 9:115-138. Doi: 10.1007/s10288-011-0172-x.

[40] Breiman, L., &, Friedman, J. "Predicting Multivariate Response in Multiple Linear Regression". Journal of the Royal Statistical Society: Series B (Statistical Methodology)/ Volume 59, Issue 1. Doi: doi.org. /10.1111/1467-9868.00054.

[41] Box, G., &, Hunter, J. "Multi-Factor Experimental Designs for Exploring Response Surfaces". The annals of Mathematical Statistics. Vol. 29. No. 1 pgs. 195-241. URL http://www.jstor.org/stable/2237033

[42] Shah, H., Montgomery, D., &, Carlyle, M. "Response Surface Modeling and Optimization in Multiresponse Experiments Using Seemingly Unrelated Regressions". Quality Engineering. Vol. 16, No 3, pgs. 387-397, 2004. Doi: 10.1081/QEN-120027941.

[43] Steinberg, D., &, Hunter, W. "Experimental Design: Review and Comment". Technometrics Voil. 26, No 2, May 1984. Pgs. 71-97. Doi: 10.1080/00401706.1984.10487928

[44] Box, G., &, Behnken, D. " Some New Three Level Designs for the Study of Quantitative Variables". Technometrics. Vol. 2. No 4. Nov. 1960. Pgs. 455-475. Doi: 10.2307/1266454

[45] Corker, K., Gore, B., Fleming, K., &, Lane, J. "Free Flight and the Context of Control: Experiments and Modelling to Determine the Impact of Distributed Air-Ground Air Traffic Management on Safety and Procedures". 3rd USA/Europe Air Traffic Management R &D Seminar. Napoli, 13-16 June 2000.

[46] Hoekstra, J., Van Gent, R., &, Ruigrok, R. "Designing for safety: the free flight air traffic management concept". Reliability Engineering & System Safety. Vol. 75 , Issue 2, February 2002, pgs. 215-232. https://doi.org/10.1016/S0951-8320(01)00096-5.

[47] Tumer, K., &, Agogino, A. " Distributed Ageng-Based Air Traffic Flow Management". AAMAS07. May 14-18, 2006. Honolulu, Hawaii, USA. ACM 1-59593-303-4/06/005.

[48] Majumdar, A., &, Ochieng, W., “Factors affecting air traffic controller workload: Multivariate analysis based on simulation modelling of controller workload”. Centre for Transport Studies. Department of Civil and Environmental Engineering. Imperial College of Science, Technology and Medicine. London, UK. https://doi.org/10.3141/1788-08

[49] Sacks, J., Welch, W., Mitchell, T., &, Wynn, H. “Design and Analysis of Computer Experiments”. Statistical Science. 1989, Vol. 4, No. 4. 409-435. doi:10.1214/ss/1177012413

[50] Deng, X., &, Lin, C. “Design and Analysis of Computer Experiments”. Handbook of Research on Applied Cybernetics and Systems Science. 10.4018/978-1-5225-2498-4.ch013

[51] What Is Simple Linear Regression? Pennsylvania State University, pdfs.semanticscholar.org/presentation/e391/418b2b395a5b40ffe87e1cb2359c4612f593.pdf.

[52] Multiple Linear Regression. Mathworks, www.mathworks.com/help/stats/regress.html.

[53] Variance Inflation Factor. Wikipedia, www.wikipedia.org/wiki/ Variance_inflation_factor

[54] Christopher, Dean. “Design of Experiments (DOE) Tutorial.” American Society for Quality, asq.org/learn-about-quality/data-collection-analysis-tools/overview/design-of-experiments-tutorial.html.

[55] Montgomery, Douglas C. Introduction to Statistical Quality Control. 5th ed., John Wiley & Sons, Inc., 2005.

[56] Grant, Trudi. The Randomized Complete Block Design (RCBD). 2010, pdfs.semanticscholar.org/presentation/e391/418b2b395a5b40ffe87e1cb2359c4612f593.pdf.

[57] Jenkins, Bruce. "Latin Hypercubes And All That: How DoE Works." Ora Research, 23 Feb. 2015, oraresearch.com/2015/02/latin-hypercubes-and-all-that/.

[58] "Computer Experiment." Wikipedia, Wikimedia Foundation, 28 Feb. 2018, en.wikipedia.org/wiki/Computer_experiment.

[59] Visintini, A. , Glover, W., &, Lygeros, J. "Monte Carlo Optimization for Conflict Resolution in Air Traffic Control". IEEE Transactions on Intelligent Transportation Systems, Vol 7, No 4. December 2006. Doi: 10.1109/TITS.2006.883108.

[60] Copa Airlines. Panama Air Space Optimization 2017. Panama, Panama.

[61] https://www.fraport.de/content/fraport/de/nachbarschaft-region/dialog/wissenswertes/flugbetrieb/was-ist-point-merge-.html

[62] Walpole, R., Myers, R., Myers, S., &, Ye, K. "Probability & Statistics for Engineers & Scientists" Prentice Hall. 9th Edition. 2012. ISBN 978-0-321-62911. Boston, MA, USA.

[63] Dean, A. & Voss, D. "Design and Analysis of Experiments". Springer, 1999. ISBN 0-387-98561-1. New York, USA.

[64] Lodi, A., Panconesi, A., &, Giovanni R. "Integer Programming and Combinatorial Optimization". 13th International Conference, IPCO 2008. Bertinoro, Italy. Springer. ISBN 0302-9743.

[65] Innis, T., &, Ball, M. " Estimating One-Parameter Airport Arrival Capacity Distributions for Air Traffic Flow Management". August 2002. Doi.org/10.2514/atcg.12.3.233.

[66] Beinstein, Jorge. "Workshop The Alternatives Scenarios". Panama Strategic Planning for 2030. 2017.

[67] World Bank. Data.worldbank.org/country/panama.

[68] MATLAB. www.mathworks.com/help/optim/ug/intlinprog.html.

[69] Wikipedia. "Fuel economy in aircraft". en.wikipedia.org/wiki/fuel_economy_in_aircraft.

[70] William B. Greiner. "Oil Prices and Economic Growth". June 2018 www.marinerwealthadvisors.com/insights/oil-prices-and-economic-growth.

[71] UTP headquarters of the Conference Panamanian Airspace Optimization Workshop. www.utp.ac.pa/utp-sede-de-la-conferencia-taller-de-optimizacion-aerospacial-panameno.

[72] Jackman, J., Guerra de Castillo, Z., & Olafsson, S. "Stochastic flow shop scheduling model for the Panama Canal. Journal of the Operational Research Society. Doi.1057./jors.2009.188.

[73] Kuo, T., Okudan-Kremer, G., Thi Phuong, N, &, Hsu. C.W. "Motivations and barriers for corporate social responsibility reporting: Evidence from the airline industry". Journal of Air Transport Management. Volume 57. 2016 doi.org/10.1016/j.jairtraman.2016.08.003.

[74] National Climatic Data Center, U. S. Department of Commerce. www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv&georegionabbv.

[75] Menon, P.K, ; &, Sweriduk, G.D. "New Approach for Modeling, Analysis and Control of Air Traffic Flow". Journal of Guidance, Control, and Dynamics. Vol. 27, No. 5 September-October 2004.

[76] Sudjianto, A., Du, X., & Chen, W. " Probability Sensitivity Analysis in Engineering Design using Uniform Sampling and Saddlepoint Approximation". Ideal.mech.northwestern.edu. SAE International 2005.

[77] Besterfield, Dale. "Quality Improvement". Ninth Edition. Pearson Editorial.ISBN-13:978-013-262441-1. New Jersey, USA.

[78] Montgomery, D., & Runger, G. "Applied Statistics and Probability for Engineers". Second Editio. John Wiley Editorial. 1999. ISBN 0-471-17027-5. New York, USA.

[79] Corley, H.W. &, Charoensri, S., "Some Pareto scalar equilibria for n-person prescriptive games" Natural Science. Vol 6 No. 13. 2014. Doi:10.4236/ns.2014.613098.

[80] Kung, Pin. "Multivariate Modeling for a Multiple Stage, Multiple Objective Green Building Framework". The University of Texas at Arlington. August 2012.

[81] Jones, Erick C. "Quality Management for Organizations Using Lean Six Sigma Techniques. CRC Press.2014. ISBN 978-1-4398-9782-9. Boca Raton, FL. USA.

[82] Shapiro, Jeremy. "Modeling the Supply Chain". Second Edition. 2007. Cengage Editorial. New Delhi, India. ISBN-10: 81-315-0156-6.

[83] Ozan, Turgut. “Applied Mathematical for Production and Engineering Management”. Prentice Hall. 1986. New Jersey USA. ISBN 0-8359-0026-6.

[84] Bifulco, G., Galante, F., Pariota, L., &, Spena, M. “A Linear Model for Estimation of Fuel Consumption and the Impact Evaluation of Advanced driving assistance systems”. Journal of Sustainability. Agust 2015. Doi: 10.3390/su71014326.

[85] European Commission, EUROCONTROL & FAA. “2015 Comparison of Air Traffic Management-Related Operational Performance: U.S./Europe”, August 2016. https://www.faa.gov/air_traffic/publications/media/us_eu_comparison_2015.pdf.

[86] Kutner, M., Nachtsheim, C., Neter, J., &, Li W. “Applied Statistical Models”. McGrawHill, Fifth Edition. New York, USA 2005. ISBN 007-112221-4.

[87] Hopp, W. &, Spearman, M. “Factory Physics”. Waveland Press Inc. Third Edition. Illinois, USA, 2008. ISBN 1-57766-739-5.