

Original Article

Flight Profile Analysis: Finding Correlations using Scatter Plot Method and Evaluation of Scatter Plot

Mohd Harridon^{1,*}, Mohd Rizal²

¹ Universiti Kuala Lumpur Malaysian Institute of Aviation Technology, Jalan Jenderam Hulu, Kampung Jenderam Hulu, 43900 Sepang, Selangor, Malaysia

² Weststar Aviation Services, 16100 Kota Baru, Kelantan, Malaysia

* Correspondence: mdharridon@unikl.edu.my

Received: 10 July 2019; Accepted: 18 October 2019; Published: 21 December 2019

Abstract: This paper analyzed flight profiles where parameters of the trajectory of the airplane were found to be correlated in some way. The Scatter Plot Method was used to find these correlations but somehow the degree of correlations was mixed. We proceeded by evaluating the usage of Scatter Plot where Scatter Plot seems to have a good prospect.

Keywords: flight profile, scatter plot, fixed wing

1. Introduction

After each flight, when there's a need or if its set a routine, the flight would be analyzed in order to identify anomalies or safety issues during flight. Companies now days, such as Weststar Aviation Services in Kota Baru, invested in Flight Data Monitoring Program where data of flights are recorded during flights and after each flight these data are analyzed at the ground station. In this paper, the authors analyzed data of a flight profile from a source which is not named (confidentiality clause). The flight profile is from a fixed wing aircraft. The analysis was actuated to search for correlations between parameters of flight. The authors are interested to know what transpired during the flight and whether the flight is clad with certain degree of anomalies. The authors used Scatter Plot Method, which is part of Big Data Analytics, to show the correlations and also gave elaborations upon the method.

2. Literature Review

It is a norm now days for aviation organizations to actuate post flight analysis in order to alleviate safety, increase performance of flights, reduce cost, and others. Chati and Balakrishnan had conducted a study which analyzed the performance of the engines of Airbus A330 using data collected by the Flight Data Recorder (FDR) of the aircraft [2]. They also find the correlation between the aircraft altitude and the performance of the engines. Campbell in his paper stated that FDR are routinely used to detect events during flights [3]. This detection is imperative in order to increase the safety of the aircrafts. Campbell mentioned that several parameters were recorded by the FDR and several correlations exist among those parameters. Furthermore, analysis of flight profile would detect anomalies that occur during flight. Li, Gariel and et al. in their paper stated that by analyzing flight data and detecting it anomalies, solutions can be derived which would alleviate the quality of flights [4]. Their research also seek correlations between recorded parameters of flight. Chang and Tan had actuated a research where they had performed post flight analysis of data recorded in the Quick Access Recorder (QAR). In their paper they explained that the post flight analysis was carried

out so that they could identify irregular deviations (anomalies) of flight control surface and detect correlations among several parameters [5]. Scatter Plot is a method that is used to find and identify correlations among several parameters. Friendly and Denis in their paper mentioned the usage of Scatter Plot to show relationships or correlations among variables and further iterated the graphical means of Scatter Plot where Scatter Plot has the advantage of being visually explanatory [6]. Goo in her paper described Scatter Plot as a good method to show relationship between two variables [7]. The paper further stated that Scatter Plot shows correlation and not causality. What are shown in Scatter Plot are mere association the paper asserted. Sarikaya and Gleicher had commented that Scatter Plot displays association between parameters and they proposed several means to enhance or evolve Scatter Plot where this enhancement would aid in the identification of patterns [8]. Haroz in his paper pointed out that Scatter Plot is used to indicate correlations between data sets. He dwell further into a customized Scatter Plot called Connected Scatterplot (CS) where he expressed his view that CS also has merit in presenting and displaying data albeit in a more structured way [9]. We can observe that review or analysis of flight data or profile is an important step in increasing safety and had been actuated by numerous parties and individuals. Quite number of parties and individuals use Scatter Plot as the chosen method to show correlations between two or more parameters.

3. Methodology

Several steps were taken in order to find the correlations between several flight parameters. The correlations told a significant picture or story upon the movement of flights. The methodology is shown in Figure 1. Peering Figure 1, the flight data that we used originated from a test flight where as mentioned earlier the source could not be revealed (confidentiality clause). We confined our research pertinent to only data of trajectory. This is an arbitrary decision and this aid us to concentrate on a data set which is optimum in size. The time frame of the data was also chosen via arbitrary mean where we took 10 minutes worth of data from a random starting point. The data that were chosen were Altitude, Pitch, Roll, and Yaw. During flight, the data were recorded every 4 seconds but there were parts of the flight where the data were not recorded for approximately 6 minutes due to technical glitch. We had also taken for 6 minutes void to form correlations between flight parameters. The correlations that we were interested were Altitude and Pitch, Altitude and Roll, and Altitude and Yaw. Altitude is the central parameter here. Several literature indicate changes of pitch would affect altitude, changes in roll would affect altitude if the roll is not properly executed or if there's a pre-plan to increase or decrease altitude during the roll, and changes in yaw would also affect altitude if the yaw is not properly executed or if there's a pre-plan to increase or decrease altitude during the yaw but the degrees of effect are different based upon several factors. Examples of these literature are papers composed by Singh [10], Flight Literacy [11], and Stengel [12]. Results obtained by the Scatter Plot method were then discussed based upon heuristics and literature review. Scatter Plot method was then reviewed based upon 3 criteria as shown in Table 1.

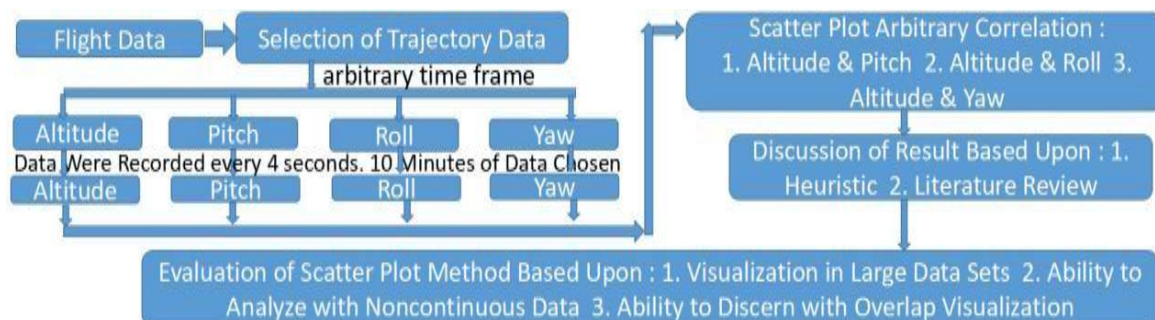


Figure 1. Method to find correlations between Flight Parameters and to evaluate Scatter Plot

Table 1. Evaluation of Scatter Plot

Evaluation of Scatter Plot	
1. Visualization in Large Data Sets	Good Visualization / Moderate / Not Good Visualization
2. Ability to Analyze with Noncontinuous Data	Good Ability / Moderate / Not Good Ability
3. Ability to Discern with Overlap Visualization	Good Ability / Moderate / Not Good Ability

4. Results and Discussion

The results are shown in the following figures. In Figure 2, the flight data worth of 10 minutes are plotted. Figure 2 shows the scatter plot for Altitude versus Time and also for Pitch versus Time. Two trend lines (the non-continuous lines) are shown in Figure 2 where the blue trend line represents the trend of pitch while the orange trend line represents the trend of altitude. The pitch trend line shows a decrease value as time goes by and this is similar to the altitude trend line. We can make a rough gesticulation that there's a correlation here since a decrease in pitch decreases the altitude and also both trend lines are in almost parallel state. We explored further and plotted the scatter plot shown in Figure 3. Figure 3 shows the scatter plot of altitude versus pitch. The red line is the trend line. Figure 3 shows that there is minute correlation where several dots are along the red line at certain segments but this is not significant. There is a majority of dots which lies away from the trend line. Peering at the data sheet, we found out that a decrease in pitch still invoke an increase in altitude which is reasonable (for example decreasing the pitch from 10 degrees to 7 degrees still invokes a climb or an increase in altitude since the nose is still pitched up). Thus, we concluded by looking at both scatter plots (Figures 2 and 3) and peering the data sheet, certain segments of the flight showed low degree of positive and negative correlations. In Figure 4, the flight data worth of 10 minutes is plotted. Figure 4 shows the scatter plot for Altitude versus Time and also for Roll versus Time.

Two trend lines (the non-continuous lines) are shown in Figure 4 where the blue trend line represents the trend of roll while the orange trend line represents the trend of altitude. The roll trend line shows a decrease value as time goes by and this is similar to the altitude trend line. We can make a rough gesticulation that there's a correlation here since a decrease in roll decreases the altitude and also both trend lines are in almost parallel state for a brief period. The intersection of both trend lines merely shows the difference in the decreasing rate of both parameters (roll and altitude). We explored further and plotted the scatter plot shown in Figure 5. Figure 5 shows the scatter plot of altitude versus roll. The red line is the trend line. Figure 5 shows that there is minute correlation where several dots are along the red line at certain segments but this is not significant. There are 2 clusters of dots at the value roll = 0 where these clusters are a little bit away from the trend line.

There are outliers which lie away from the trend line and their number is significant. A negative roll value indicates a turn to the left while a positive roll value indicate a turn to the right. The scatter plot in Figure 5 visually shows that a roll to the left or right puts the aircraft at equal range of altitude (left roll puts the aircraft at a range of altitude from 10000 meters till 27500 meters and right roll puts the aircraft at a range of altitude from 10000 meters till 22500 meters).

Peering at the data sheet, we found out that for several segments a decrease in roll value still increase the altitude of the aircraft. Perhaps these segments are minor in count thus this explained why both trend lines in Figure 4 showed decreasing trends. We concluded by looking at both scatter plots (Figures 4 and 5) and peering the data sheet, certain segments of the flight showed low degree of negative correlation. In Figure 6, the flight data worth of 10 minutes are plotted. Figure 6 shows the scatter plot for Altitude versus Time and also for Yaw versus Time. Two trend lines (the non-continuous lines) are shown in Figure 6 where the blue trend line represents the trend of yaw while the orange trend line represents the trend of altitude. The yaw trend line showed an increase

value as time goes by and this is the opposite to the altitude trend line which decreases as time goes by. We can make a rough gesticulation that through visual analysis of the scatter plot of Figure 6 there's a low degree of negative correlation which perhaps apply for some segments of the flight since Figure 6 also showed a portion where altitude increases as yaw increases. We explored further and plotted the scatter plot shown in Figure 7. Figure 7 shows the scatter plot of altitude versus yaw. The red line is the trend line.

Figure 7 shows that there is minute correlation where several dots are along the red line at certain segments but this is not significant. There is a majority of dots which lies away from the trend line and there is a lot of outliers. But these outliers followed the movement of the trend line where these dots moved almost in parallel with the trend line. Peering at the data sheet, we found out that in some segment an increase in yaw invoked an increase in altitude and in another segment a decrease in yaw invoked an increase in altitude as well.

By looking at both scatter plots (Figures 6 and 7) and peering the data sheet, certain segments of the flight showed a low degree of negative correlation and certain segments of the flight showed a low degree of positive correlation. In Figure 8, the flight data worth of 19.9 minutes are plotted. But the in the middle of the flight there was a 6 minutes void where data was not captured for the duration of 6 minutes. Figure 8 shows the scatter plot for Altitude versus Time and also for Pitch versus Time. Two trend lines (the non-continuous lines) are shown in Figure 8 where the blue trend line represents the trend of pitch while the orange trend line represents the trend of altitude. The pitch trend line shows a decrease value as time goes by while the altitude trend line shows an increase value as time goes by.

A rough gesticulation that there's a low degree of negative correlation here but we are concerned that this would be a misnomer as there's a 6 minutes gap of data which would lead to inaccuracies of the trend lines. Also conventional wisdom states that, in a majority of cases, when an aircraft pitches down, the altitude would stop to increase and began to decrease. We explored further and plotted the scatter plot shown in Figure 9.

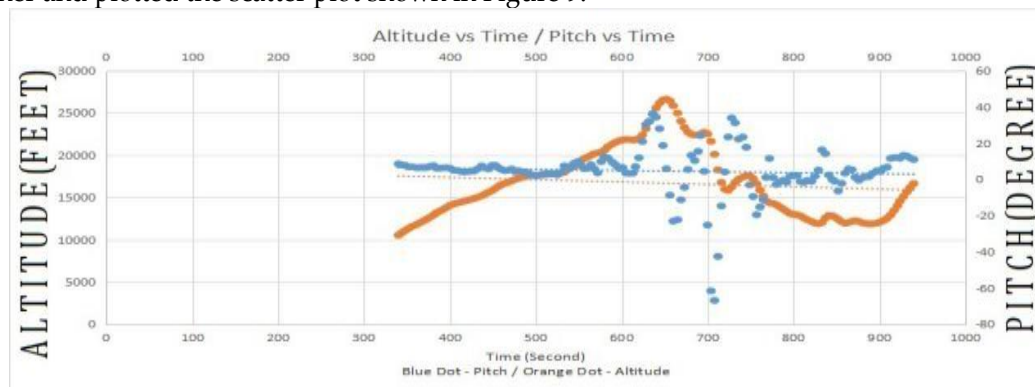


Figure 2. Altitude versus Time and Pitch versus Time

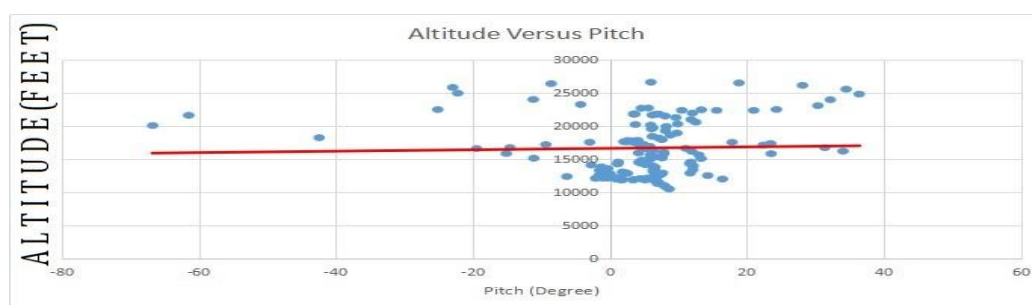


Figure 3. Altitude versus Pitch

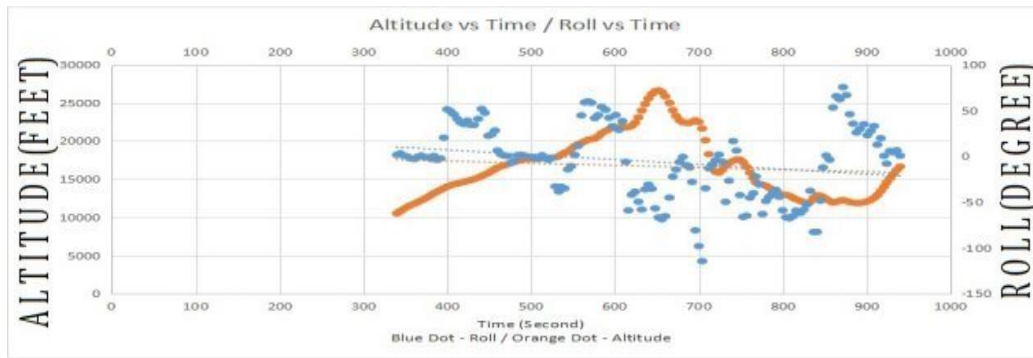


Figure 4. Altitude versus Time and Roll versus Time

Figure 9 shows the scatter plot of altitude versus pitch with the void data in the middle. The red line is the trend line. Figure 9 shows that there is minute correlation where several dots are along the red line at certain segments but this is not significant. There is a majority of dots which lies away from the trend line. A lot of the dots converged within an altitude range of 17500 meters till 20000 meters. Peering at the data sheet, we found out that for some segments an increase in pitch invoked an increase in altitude.

By considering both scatter plots (Figures 8 and 9) and peering the data sheet, certain segments of the flight showed low degree of positive and negative correlations. Since there's a 6 minutes void of data, we postulated that perhaps there's a misrepresentation of the real scenario and more exploration is needed to address this issue. We proceeded in evaluating the scatter plot method in order to gauge significance in identifying correlations. Our evaluation is shown in Table 2. We gave a grade of Moderate to all 3 criteria. For the first criteria which is "Visualization in Large Data Sets", we observed that when the dots get crowded, it's difficult to determine the behavior of the dots.

The trend line in scatter plot gave us assurance that the crowded dots do follow a certain trend but discrete details are unavailable. Based upon literature review, other parties or individuals had commented or raised concern upon this and a method called binning was introduced to overcome this issue. This was highlighted by Hao et al. where in their paper they segregated the data into different cells (bins) where each cell could be analyzed or visualized more discretely [13.] Stepner concurred and stated similarly in his presentation slides [14]. For the second criteria which is "Ability to Analyze with Noncontinuous Data", we observed that when there's a void in the middle of the data, as seen in Figures 8 and 9, there is perhaps a chance of visual misinterpretation of data.

A tool like scatter plot, even though it is a part of Big Data Analytics, has its short comings. Although the trend lines are supposed to be indicators of trends and able to extrapolate over seas of non-existing data, the inclusion of these non-existing data, if they somehow miraculously appear, would perhaps alter the trend lines. This thus supports our evaluation of scatter plot pertaining to these criteria. For the third criteria which is "Ability to Discern with Overlap Visualization", we observed that a lot of the dots had overlapped each other or had hidden other dots. The dots that were hidden are hard to detect visually and perhaps there are 6 or 7 or more layers of hidden data or dots. Scatter plot does not offer us the ability to comprehend these hidden layers, hence making the interpretation of data less accurate. This issue had been addressed by Mayorga and Gleicher in their paper where they had introduced an enhanced version of scatter plot which takes layers of data or dots and fashion them into contours and this aids the visualization process [15].

A scatter plot has advantages which are: 1. Offer graphical means to identify correlations 2. A quick and robust way to gain correlations 3. An easy approach to identify correlations 4. A good visualization tool that aids analysis of correlations 5. Has numerous enhancements to solve issues such as overlaps, large data sets, noncontinuous data, and others. There are however other methods which is mathematical in nature that could be used to identify correlations. Those methods are Karl Pearson's Coefficient of Correlation [16], Spearman's Rank Correlation Coefficient [16], and Methods of Least Squares [17].

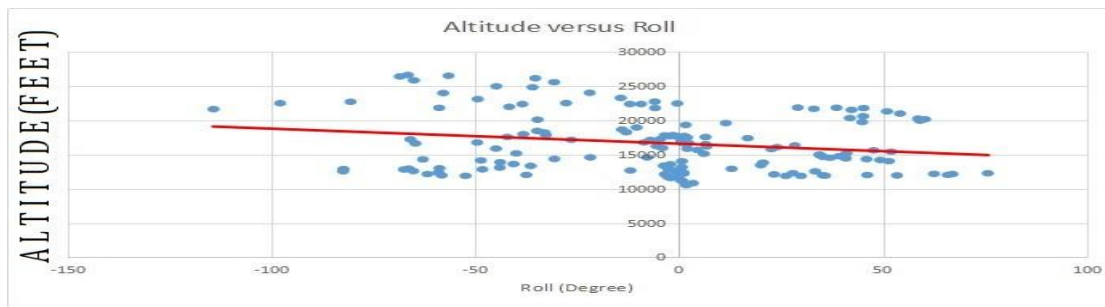


Figure 5. Altitude versus Roll

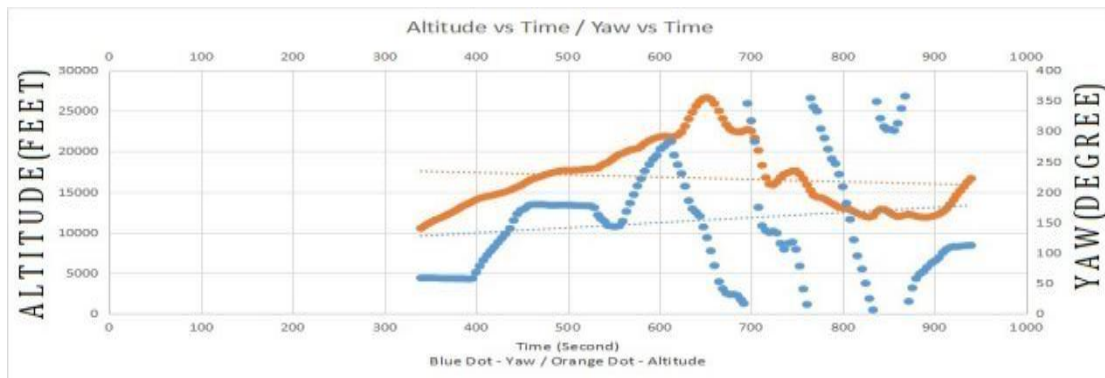


Figure 6. Altitude versus Time and Yaw versus Time

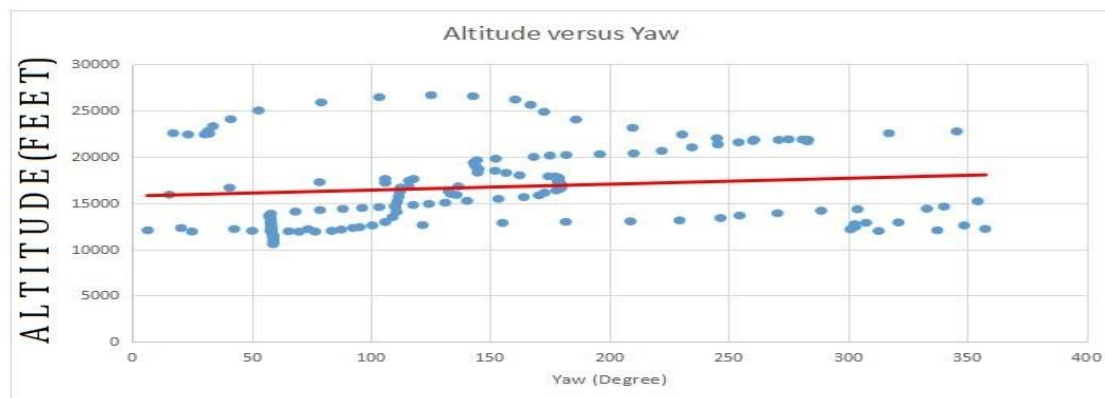


Figure 7. Altitude versus Yaw

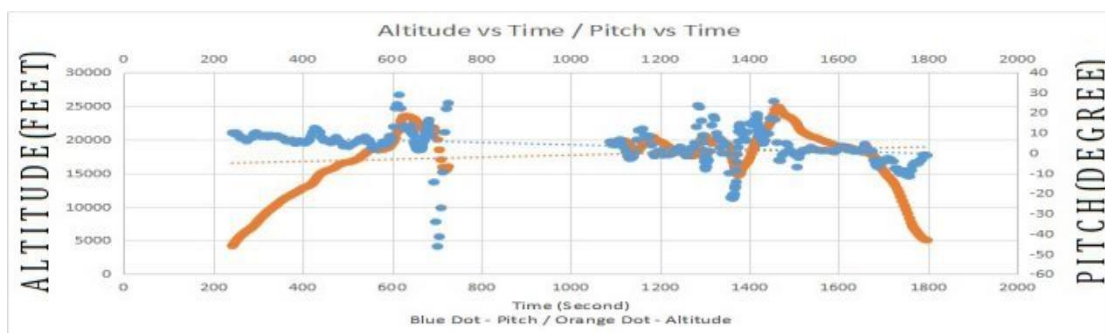


Figure 8. Altitude versus Time and Pitch versus Time and Void Data

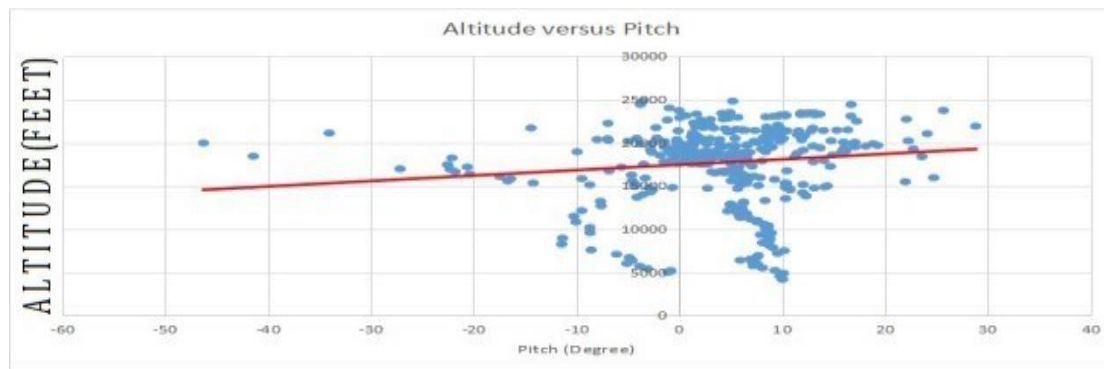


Figure 9. Altitude versus Pitch and Void Data

Table 2. Evaluation of Scatter Plot

Evaluation of Scatter Plot		
1.	Visualization in Large Data Sets	Moderate
	Ability to Analyze with Noncontinuous Data	Moderate
2.	Data	

5. Conclusions

Flight data of a flight test was analyzed using scatter plot method for the purpose of identifying correlations. Several parameters of the trajectory of flight showed several types of correlations but in a segmented fashion. The scatter plot method was indeed useful for visualization of correlations but drawbacks exist mainly when the data or dots of scatter plot are in large data sets, overlapping, and non-continuous. Several enhancements of scatter plot exist to overcome these drawbacks but other tools which are mathematical in nature are also available to identify correlations.

References

- [1] SAS Institute Incorporated, "Data Visualization Techniques : From Basics to Big Data with SAS Visual Analytics", White Paper, Retrieved from Web 23rd July 2018, North Carolina.
- [2] Chati, Y., Balakrishnan, H., "Aircraft Engine Performance Study Using Flight Data Recorder Archives", American Institute of Aeronautics and Astronautics Aviation Conference, Los Angeles, USA, 2013.
- [3] Campbell, N., "The Evolution of Flight Data Analysis", Australian Society of Air Safety Investigators (ASASI) Conference, Wellington, New Zealand, 2007.
- [4] Li, L., Gariel, M., et al., "Anomaly Detection in Onboard-Recorded Flight Data using Cluster Analysis", IEEE / AIAA 30th Digital Avionics Systems Conference, 16th - 20th October 2011, Seattle, Washington, USA.
- [5] Chang, R.C., Tan, S., "Post Flight Analysis based on QAR in FOQA program for Jet Transport Aircraft Part 1 : Angular Position Monitoring of Flight Control Surface", Journal of Aeronautics, Astronautics and Aviation, Series A 44 (1) : 9 -16, March 2012.
- [6] Friendly, M., Denis, D., "The Early Origins and Development of the Scatterplot", Journal of the History of the Behavioral Sciences, Vol. 41(2), 103-130 Spring 2005.

- [7] Goo, S.K., "The Art and Science of the Scatter Plot", Factank News in the Numbers, Pew Research Center, September 16, 2015, Washington DC, USA.
- [8] Sarikaya, A., Gleicher, M., "Scatterplots : Tasks, Data, and Designs", IEEE Transactions on Visualization and Computer Graphics, Volume 24, Number 1 - Jan 2018, NSF Awards IIS - 1162037.
- [9] Haroz, S., Kosara, R., Franconeri, S.L., "The Connected Scatterplot for Presenting Paired Time Series", August 21, 2016, Transactions on Visualization and Computer Graphics (TVCG), Vol. 22, No. 9, pp. 2174-2186, 2016.
- [10] Singh, A.K., Dahiya, R., "Design and Modeling of Controllers for Aircraft Pitch Control Movement", International Journal of Engineering and Computer Science, ISSN : 2319-7242, Volume 5, Issue 11, Nov. 2016, Page Number 19198 - 19211. [11] Flight Literacy, "Steep Turns", Cory Doggett, Odessa, Texas, USA, 2018.
- [12] Stengel, R.F., Berry, P.W., "Stability and Control of Maneuvering High-Performance Aircraft", Journal of Aircraft, Volume 14, No. 8, August 1977, pp. 787 - 794, American Institute of Aeronautics and Astronautics.
- [13] Hao, C.M., Dayal, U., Sharma, R.K., Keim, D.A., Janetzko, H., "Variable Binned Scatter Plots", Information Visualization, Vol. 9, No. 3, pp. 194 - 203, 2010.
- [14] Stepner, M., "BinScatter : Binned Scatterplots in Stata", Massachusetts Institute of Technology, Department of Economics, Cambridge, United States, August 1, 2014.
- [15] Mayorga, A., Gleicher, M., "Splatterplots : Overcoming Overdraw in Scatter Plots", IEEE Transactions on Visualization and Computer Graphics, Volume 19, Issue 9, September 2013, pages 1526 - 1538.
- [16] Mukaka, M.M., "Statistics Corner : A Guide to Appropriate Use of Correlation Coefficient in Medical Research", Malawi Medical Journal, 24(3) : 69 - 71 September 2012.
- [17] Miller, S.J., "The Method of Least Squares", Mathematics Department Brown University Providence, RI 02912.



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).