RESEARCHERID
THOMSON REUTERS

[Maina* 4(12): December, 2017] ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

ESTIMATING THE EXPECTED FLIGHT CANCELLATIONS DUE TO WEATHER CONDITIONS AT DALLAS FORT WORTH INTERNATIONAL AIRPORT (DFW)

Eugene Vida Maina, Ph.D*, Albert Forde, Ph.D
* Operations Systems Research Analyst, Dallas Fort Worth International Airport
Civil / Transportation Engineer, New Jersey Department of Transportation (NJDOT)

DOI: 10.5281/zenodo.1119489

KEYWORDS: Cancellations Flight operations Airport capacity Irregular operations (IROPS) Empirical Bayes (EB) Validation.

ABSTRACT
Prevailing weather conditions significantly influence flight cancellations by airlines, airports and the Federal Aviation Administration (FAA). As a result, Dallas Fort Worth International Airport's (DFW) Operations (OPS) Division is interested in estimating the anticipated departing flight cancellations to efficiently, and safely mitigate the negative effects caused. Accurately estimating the anticipated number of departing flight cancellations well in advance can assist airport operators to plan efficiently, and safely execute flight operations with minimal impact on passenger movement, airport capacity and throughput; consequently minimizing or completely reducing the chances of irregular operations (IROPS).This study focuses on weather related departure flight cancellations. Operations and weather data associated with at least one departure cancellation from 2010 to 2015 are used for mathematical modelling, analysis and validation. The Empirical Bayes (EB) model is used to evaluate the various weather factors that significantly affect flight departure operations. Using Pearson matrix correlation, five weather factors, each with 1047 observations, (6282 entities) are determined and used in the model. The use of EB model accounts for the regression-to-mean (RTM) phenomena not accounted for by the negative binomial (NB) distributions used in the evaluation process. Both qualitative and quantitative statistical validation approaches are used to justify the results found. Actual departure cancellation DFW data from 2016 is used against the estimated results for validation. Results indicate that the number of weather related departure flight cancellations increases when the dew point temperature, the visibility distance and the cloud ceiling height decrease. Results, however, indicate that the number of weather related departure flight cancellations increases with increasing precipitation and cloud cover index. With a strong and positive correlation of R2 = 0.9869 and the means of both the estimated and actual cancellations being statistically different, this study renders the variable selection process and mathematical approach appropriate and acceptable.

INTRODUCTION
One factor that greatly affects flight delays, but is not entirely understood is flight cancellations (Seelhorst, 2014). Cancellations – especially those related to weather conditions cannot be prevented or be avoided. However, if a satisfying prediction solution to the related impact(s) were available, it would significantly assist the various stakeholders within the aviation industry – in this case, the airports to estimate, plan and safely manage the resultant effects on flight operations and airfield operations such as runway inspections and repairs. For the purpose of this study, cancellations refer to the departing/outbound cancelled flights. In addition, accurately predicting the expected number of cancellations can assist the FAA run inter-airline compression to fill in the open slots created (Xiong, et al., 2013).

The scheduling of flights is based on an assumption that there will be no disruptions, and that the flights will depart and arrive on time (Lan, et at., 2006). However, due to flight delays or aircraft shortages caused by factors such as adverse weather conditions at given airports, airline carrier decisions (safety, mechanical or marketing), National Air System (NAS) and security reasons flight cancellations are usually adopted to minimize the potential economic losses (Mao, et al., 2015) and schedule disruptions (Gershkof, 2016). Mechanical problems and delays in the schedule of incoming flights may also contribute to flight cancellations (Jarrah et al., 1993).

As a result, passenger itineraries, airlines and the FAA schedules and airport operations are significantly interfered with. This can have a dominos effect since a flight that is cancelled at one airport might be the one supposed to provide an aircraft for a departure from the destination airport. On average, masFlight estimates that it costs an airline $5,770 for every cancelled flight segment (Ausick, 2014; Schoen, 2015). Flight delays and cancellations also affect passenger movements and costs. In 2015, 380,000 passengers had their flights cancelled resulting in a total cost of $180 million (Jones, 2015); that is $474 per passenger per cancellation.

To be able to effectively predict cancellations, the factors leading to cancellations have to be identified and understood. At DFW, 62% of all cancellations are due to adverse weather, which usually leads to IROPS. IROPS negatively affects the airport's capacity, throughput, safety, and scheduled runway inspections. In addition, cancellations in their own right are a major source of delay (Seelhorst, 2014). Seelhorst further acknowledges that there exists little work on the effect of flight cancellations hence the motivation of this study.

Focusing at DFW, this study determines and investigates the weather factors that significantly contribute towards cancellations using Pearson's correlation matrices, and then applies NB and EB models to historical data to calculate the number of expected cancellations given the pre-determined weather factors. This study then uses both qualitative and quantitative statistical methods to validate the results found. Upon knowing the anticipated or rather the estimated number of cancellations, DFW plans to use this information to optimally plan, manage and safely execute flight and airfield operations by minimizing or totally eliminating IROPS.

Although a single value is found from the mathematical process employed, this study understands that when the forecasted weather values are given, they are presented as a range of the minimum to the maximum value. For example, if it is temperature, it is expressed in degrees Fahrenheit or Celsius as a minimum of say 60 to a maximum of say 70 for a given day. As a result, the expected number of cancellations will naturally vary from a minimum to a maximum value. The second part of this study plans to address this task by optimizing of the developed model. This is important for airport operations planners because they will be able to perform sensitivity analysis and as a result optimize the available resources without negatively affecting operations and safety.

LİTERATURE SEARCH

Although flight cancellations is a common act that significantly impacts both airline operators and their customers, not much has been published compared to related topics such as flight delays and flight diversions. The minimal publication could not be due to lack of research, this is because airlines have developed models they use to determine which and how many flights to cancel given the prevailing or forecasted conditions. However, due to either or both policy and marketing issues, the approach(s) taken is not public. Therefore, the factors and methods for flight cancellation are well understood and largely depend on airlines. However, DFW's OPS division realizes that the airport's capacity, operational efficiency and safety are significantly impacted and therefore the decision making of flight cancellations should be conducted in a more global way (Mao, et al., 2015). For example, the arrival capacity at San Francisco International Airport (SFO) varies between 30 and 60 lights per hour depending on the weather conditions (Xiong, Jing, and Hansen, 2013), that is a reduction of up to 50 percent. As a result – from the airport`s point of view, there is a need to estimate the expected flight cancellations prior to the occurrence to mitigate these impacts.

To estimate the relative forecasted number of cancelled flights, direct mathematical modelling is required (Sridhar, et al., 2009). The use of both binomial regression and Poisson regression are usually recommended (Zou, 2004), since the data used are count data (Niaparast & Schwabe, 2013). These two predictive methodologies have been well researched, applied and are the most common (Elvik, 2008). Specifically, both the Poisson and NB models fit well with count data that are discrete, non-negative, and sporadic (Maina, et al., 2016). Naturally, the Poisson model is preferred (Poch and Mannering, 1996). Contrary to the findings of Zhao, et al., (2009), a limitation of the Poisson model is that it assumes that the variance is approximately or equal to the mean. This is not usually the case with count non-negative data, the variance tends to be significantly larger than the mean. When the variance is larger than the mean, the data is said to be over-dispersed. To overcome the over-dispersion problem, researchers have proposed either the NB framework or the two-state Markov switching models (Zou, et al., 2013).

While analyzing two or more predictors or dependent variables, both the Poisson and NB have a limitation. They do not account for RTM phenomena usually exhibited by over-dispersed data. RTM is caused by both intra-individual variance and measurement error (Shephard, 2003). The RTM phenomena and its influence on statistical studies is well researched and documented as demonstrated by Biggeri, et al. (2013), Harewood (2015) and Cockrell et al. (2016). The general ways of accounting for RTM include the use of EB models (Maina, et al., 2015) or fully Bayesian (FB) methods (Cao, et al., 2015). In addition, depending on the nature and sample size of the data, both Zero-inflated and OLS (ordinary least squares) regression models are appropriate. Zero-inflated models attempt to account for excess zeros with the assumption that two zeros (true zeros and excess zeros) exist in the data. Hence, estimate two equations concurrently, one for the count model and one or the excess zeros. OLS models are appropriate for log-transformed count outcome variables. However, the OLS model has a major limitation, there is data lost due to unidentified values generated by taking the log of zero and as a result, the findings are biased.

As recommended in engineering processes, and perhaps the most crucial step in model development and analysis, is model validation (Catalina, et al., 2008). Model validation is the empirical technique of assessing the credibility of a mathematical model. The hypothetical expectation of model validation is that the observed data should generally match the results obtained from the developed model. Model validation is a measure of ensuring that the data used fits well into the given model, hence accurately representing the studied population. One of the techniques in performing model validation involves the use of a dataset that is independent of that used in the model development. The separate data can be either a percentage of the original sample, or an additional set of data from either the same location or a separate but similar location or from a computer program via Monte Carlo methods (Washington & Wolf, 1997: Sargent, 2004: TRB online *(Last accessed in May, 2017)*). Another reason why it is important to validate, is the relationships or trends portrayed in a given sample may not be usable over a period or at different locations is that the chosen sample might not accurately represent the given population.

MATHEMATICAL MODELING

This study uses a predictive approach to evaluate and obtain results. Predictive modeling is a mathematical technique wherein the conventional goal is to find the relationship between an objective, a response variable and a set of given predictor variables, with the intention of estimating the future values of the predictors and substituting them into the mathematical relationship to predict the future target variable values.

Here, both the actual and the NB predicted flight cancellations are used to determine the expected future flight cancellations given the forecasted weather conditions. This sets up a before-and-after condition. Generally, two methods are appropriate under this condition, one simpler and one rather more complex. The simpler of the two is the comparison group method and the EB method the complex one – but the more robust of the two (Gross, et al., 2010). This study adopts the EB approach to increase the precision of flight cancellation evaluation by estimating a weighted combination of the actual and predicted cancellations as follows:

$$\boldsymbol{E_{fc}} = (\boldsymbol{w} * \boldsymbol{p}) + [(\boldsymbol{1} - \boldsymbol{w}) * \boldsymbol{A}] \qquad \textit{(Equation I)}$$

Where:

$\boldsymbol{E_{fc}}$ = number of expected flight cancellations per day

$\boldsymbol{w}$ = weight factor determined as follows:

$$\boldsymbol{w} = {}^{\boldsymbol{1}}\!/_{[\boldsymbol{1} + (\boldsymbol{p} * \boldsymbol{k})]} \qquad \textit{(Equation II)}$$

$\boldsymbol{k}$ = NB dispersion parameter

$\boldsymbol{A}$ = number of actual flight cancelations determined using extrapolation of historical data under prevailing precipitation conditions

$\boldsymbol{p}$ = predicted number of flight cancellations per day

As the over-dispersion parameter nears zero, that is less variation, the distribution also nears a Poisson distribution (Tegge et al. 2010). The linear model for the I^{th} cancellation with q parameters $X_{i1}, X_{i2}, X_{i3} \ldots X_{iq}$, and regression coefficients $\beta_0, \beta_1 \ldots \beta_q$ are takes on the following form:

RESEARCHERID
THOMSON REUTERS

[Maina* 4(12): December, 2017] ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

$$\log(\mu) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq} \qquad \textit{(Equation III)}$$

Equation III is determined using the NB distributions. Like the Poisson distribution model, the NB model describes the occurrence of a random and rare event. The main difference between these two log-linear models is that unlike the Poisson distribution where the assumption is that the variance is equal or approximately equal to the mean, the NB distribution compensates for situations where the variance $(\mu + \mu^2/k)$ is larger than the mean, a phenomenon referred to as over-dispersion. The NB model utilizes the following distribution function:

$$P(y_i) = \left(\Gamma\left(y_i + \frac{1}{k}\right) \Big/ y_i!\,\Gamma\left(\frac{1}{k}\right) \right) * \left(k\mu_i / 1 + k\mu_i \right)^{y_i} * \left(1 / 1 + k\mu_i \right)^{1/k} \qquad \textit{(Equation IV)}$$

Where:

$\boldsymbol{\Gamma}$ = is the gamma function
$\boldsymbol{\mu}$ = is the negative binomial distribution mean
$\boldsymbol{k}$ = is the dispersion parameter

DATA ANALYSİS

Data for 1,047 days of observations with at least one weather related cancellation from 2010 to 2015 at DFW airport is considered for analysis. Data recorded from 2016 are used to validate the mathematical model and the results of the analysis. The response/dependent variable is the number of cancelled departing flights due to prevailing weather conditions and the predictor/independent variables are temperature means (${}^0F$), dew point (${}^0F$), humidity (%), sea level pressure (*in*), visibility (*mi.*), wind speed (*mph*), precipitation (*in*), cloud cover, wind direction (*degrees*) and cloud ceiling height (*ft.*). Data for the cancelled flights are obtained from the BTS database. Temperature, dew point, humidity, sea level pressure, visibility, wind speed, precipitation and cloud cover data are obtained from the Underground Weather database, while the cloud ceiling height data are obtained from the Iowa Environmental Mesonet (IEM) database. Table 1 presents the descriptive statistics for each of these variables. Each variable has 1,047 actual corresponding observations and their distributions seem significantly reasonable. Except for sea level pressure and visibility that show some under-dispersion tendencies, the other predictors are over-dispersed, meaning that their variances are higher than their corresponding means. This suggests that the NB model is more appropriate.

Table 1: Descriptive Statistics

	N	Min.	Max.	Mean	Std. Dev.	Var.
Cancelled Flights	1047	1.0	686.0	21.4	59.8	3575.3
Temperature (°F)	1047	17.0	96.0	64.7	18.1	329.3
Dew Point (°F)	1047	3.0	74.0	50.2	17.0	289.3
Humidity (%)	1047	29.0	97.0	63.9	13.6	185.1
Sea Level Pressure (in)	1047	29.5	30.7	30.0	0.2	0.0
Visibility (mi.)	1047	1.0	10.0	9.1	1.6	2.6
Wind Speed (mph)	1047	0.9	125.8	11.3	6.2	38.9
Precipitation (in)	1047	0.0	5.2	0.2	0.5	0.2
Cloud Cover	1047	0.0	8.0	4.9	2.5	6.1
Wind Direction (degrees)	1047	1.0	360.0	175.9	98.6	9717.7
Cloud Ceiling Height (ft.)	1047	0.0	30000.0	7048.2	6479.1	4.2E+7

RESEARCHERID
THOMSON REUTERS
[Maina* 4(12): December, 2017] ISSN 2349-4506
Impact Factor: 2.785
Global Journal of Engineering Science and Research Management

Referring to Table 1 and like most predictive models, several predictors can be used for analysis. However, these being count observational data and more than two predictors, there is a high chance of problems (multi-collinearity) occurring (Dickey, 2012). Multi-collinearity is the inter-correlation between two or more independent predictors in a given regression model, this phenomenon introduces complications into the regression function. Independent predictors are more likely to be correlated when count observational data, such as those used in this study are collected. A model can be highly bias when two or more highly correlated predictors are included in the same model, or when a reliable predictor is correlated with a non-reliable predictor is omitted from the model.

When predictors in a given model are interrelated, they are truly independent or orthogonal. When two predictors are orthogonal, and their correlation coefficient (R^2) is zero, then their estimated coefficients will remain the same whether or not they are included in the same model. This study understands that usually predictors are orthogonal only because of controlled experiments. As a result, and as a measure to minimize the effects of multi-collinearity, this study uses the Pearson correlation matrix to determine which predictors are suitable to be in the same model as presented in Table 2.

To select which predictor variables to include in the model, three constraints have to be met:

a. the predictor must have a weak or no correlation with the response variable, in this case the number of cancelled flights;
b. the predictor must have a weak or no correlation with the rest of the predictors; and
c. the R^2 between the predictors have to be statistically significant at 95% confidence level.

The range of less or no correlation is between the coefficient of -2.99 and 2.99. A coefficient of more than 3 or less than -3, indicates strong correlations and therefore not accepted.

Table 2: Pearson Correlation Matrix

	CAN	**TEMP**	**DEW**	**HUM**	**SEA**	**VIS**	**WSPD**	**PRE**	**CLD**	**WDIR**	**CEIL**
CAN	**1**	-0.3 0.0	-0.2 0.0	0.3 0.0	0.3 0.0	-0.2 0.0	0.0 0.1	0.2 0.0	0.2 0.0	0.0 0.6	-0.2 0.0
TEMP	-0.3 0.0	**1**	0.9 0.0	-0.2 0.0	-0.6 0.0	0.3 0.0	0.0 0.7	0.0 1.0	-0.2 0.0	-0.1 0.0	0.0 0.2
DEW	-0.2 0.0	0.9 0.0	**1**	0.2 0.0	-0.6 0.0	0.1 0.0	0.0 0.6	0.1 0.0	0.1 0.0	-0.2 0.0	-0.2 0.0
HUM	0.3 0.0	-0.2 0.0	0.2 0.0	**1**	0.0 0.4	-0.6 0.0	0.0 0.6	0.4 0.0	0.7 0.0	-0.2 0.0	-0.6 0.0
SEA	0.3 0.0	-0.6 0.0	-0.6 0.0	0.0 0.4	**1**	-0.1 0.1	-0.2 0.0	-0.1 0.0	0.0 1.0	0.0 0.2	0.1 0.0
VIS	-0.2 0.0	0.3 0.0	0.1 0.0	-0.6 0.0	-0.1 0.1	**1**	0.0 0.4	-0.2 0.0	-0.2 0.0	0.1 0.0	0.2 0.0
WSPD	0.0 0.1	0.0 0.7	0.0 0.6	0.0 0.6	-0.2 0.0	0.0 0.4	**1**	0.0 0.2	0.1 0.0	0.1 0.0	-0.1 0.0
PRE	0.2 0.0	0.0 1.0	0.1 0.0	0.4 0.0	-0.1 0.0	-0.2 0.0	0.0 0.2	**1**	0.2 0.0	0.0 0.1	-0.2 0.0
CLD	0.2 0.0	-0.2 0.0	0.1 0.0	0.7 0.0	0.0 1.0	-0.2 0.0	0.1 0.0	0.2 0.0	**1**	-0.2 0.0	0.0 0.0
WDIR	0.0 0.6	-0.1 0.0	-0.2 0.0	-0.2 0.0	0.0 0.2	0.1 0.0	0.1 0.0	0.0 0.1	-0.2 0.0	**1**	0.0 0.2
CEIL	-0.2 0.0	0.0 0.2	-0.2 0.0	-0.6 0.0	0.1 0.0	0.2 0.0	-0.1 0.0	-0.2 0.0	0.0 0.0	0.0 0.2	**1**

Model: CAN = Cancelled flights, TEMP = Temperature (℉), DEW = Dew point (℉), HUM = Humidity (%), SEA = Sea level pressure (in), VIS = Visibility (mi.), WSPD = Wind speed (mph), PRE = Precipitation (in), CLD = Cloud cover, WDIR = Wind direction (degrees), CEIL = Cloud ceiling height (ft.)

RESEARCHERID
THOMSON REUTERS
[Maina* 4(12): December, 2017] ISSN 2349-4506
Impact Factor: 2.785
Global Journal of Engineering Science and Research Management

ANALYSİS OF PARAMETER ESTİMATES

The NB regression model is used as part of analysis because it is suitable for count data variables, which usually show over-dispersion tendencies. The results of the NB analysis are presented Table 3. The table shows the NB regression coefficients for the predictors together with their corresponding standard errors, Wald chi-square values, and p-values at 95% confidence intervals for the coefficients. All the predictors are statistically significant. In this case, when all the predictors are held at zero, the intercept coefficient is 4.35. For the other predictors (dew point, visibility, precipitation, cloud cover and cloud ceiling height), these are the NB regression estimates for a one-unit increase in weather related cancellations, given the other variables are held constant in the model. If the dew point, visibility, precipitation, cloud cover and the cloud height were to each increase by one unit, the difference in the logarithms of the expected cancellations would be expected to decrease by -0.03, -0.13, 1.21, 0.16 and -0.02 respectfully, while holding the other variables in the model constant.

This is reasonable because decreased dew point, visibility and cloud ceiling height together with increased precipitation, and cloud cover is most likely expected to lead to increased flight launch conflicts, hence increased chances of cancellations.

Table 3: Parameter Estimates

			95% Conf. Int.	Wald	Hypothesis Test		
Parameter	**B**	**Std. Error**	**Lower**	**Upper**	**Wald Chi²**	**df**	**Sig.**
(Intercept)	4.35	0.33	3.71	5.00	176.46	1	0.000
Dew Point (°F)	-0.03	0.00	-0.03	-0.03	158.14	1	0.000
Visibility (mi.)	-0.13	0.03	-0.19	-0.07	15.71	1	0.000
Precipitation (in)	1.21	0.14	0.94	1.48	74.90	1	0.000
Cloud Cover	0.16	0.02	0.12	0.20	55.48	1	0.000
Cloud Ceiling Height (ft.)	-0.00	0.00	-0.00	-0.00	8.92	1	0.003
NEGBIN Dispersion (k)	1.64	0.07	1.51	1.77			

The next step in this analysis involves using the EB model to determine the expected cancellations. This step corrects for the RTM phenomenon, which is the tendency for the occurrence of a number of cancellations of given days to fluctuate up and down, and to converge to a long-term average as demonstrated in Figure 1. The figure shows the distributions of the actual number of cancellations recorded, the predicted number of cancellations found using the NB distributions, and the expected number of cancellations as estimated by the EB model. As shown, unlike the NB found predicted cancellations, the actual and expected cancellations follow a similar distribution. This is because the EB model estimates the number of cancellations by applying a weight factor to adjust both the actual and predicted number of cancellations to account for both over-dispersion and RTM.

RESEARCHERID
THOMSON REUTERS
[Maina* 4(12): December, 2017] ISSN 2349-4506
Impact Factor: 2.785
Global Journal of Engineering Science and Research Management

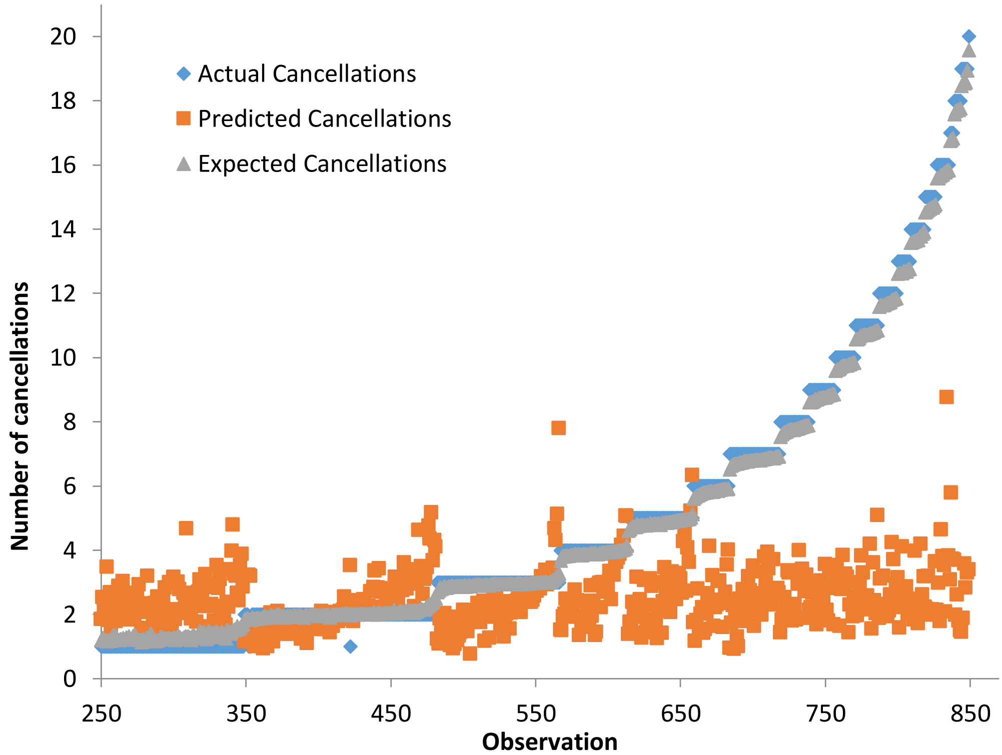

Figure 1: Relationships between actual, predicted and expected cancellations

5.1. Goodness-of-fit

Having established that the data exhibits over-dispersion and applying the Pearson correlation matrix criterion, a NB distribution model with cancelled flights as the response variable and dew point, visibility, precipitation, cloud cover and cloud ceiling height as the predictors is derived and applied. The link function of the NB is a logarithm, and all the 1,047 cases for each predictor are used in the analysis. To determine whether the data fit well in the model, a goodness-of-fit test is performed to compare the observed values to the predicted/fitted values as presented in Table 4. The criterion used to test is deviance, Pearson chi-square and the log likelihood. The values, degrees of freedom (*df*) and the ratios of value to df are presented. The model fits well since the ratios of the value to the *df* for both criteria are less than three. Therefore, the NB parameter estimates presented in Table 3 are acceptable.

Table 4: Goodness of Fit[b]

Criterion	Value	*df*	Value / df
Deviance	1575.1	1412	1.116
Pearson Chi-Square	3055.3	1412	2.164
Log Likelihood[a]	-4935.9		

1. ***a. The full log likelihood function is displayed and used in computing information criteria***
2. ***b. Information criteria are in small-is-better form.***

RESEARCHERID
THOMSON REUTERS
[Maina* 4(12): December, 2017]
ISSN 2349-4506
Impact Factor: 2.785
Global Journal of Engineering Science and Research Management

MODEL VALİDATİON

Once the mathematical model is developed and evaluated, the final step involves validation of the model. Here, validation is conducted by performing both qualitative and quantitative analysis using the 2016 expected or rather calculated weather related cancellations to the actual 2016 cancellations reported by BTS. A total of 185 observations are used for validation.

6.1. Qualitative Validation

As shown in figure 2, a simple linear regression plot is used for the qualitative analysis. The horizontal and vertical axes show the actual cancellations and expected cancellations, respectively. Results show a tight clustering of both the actual and expected number of cancellations along the diagonal line. This is an indication that both data sets agree. . In addition, an R^2 of 0.987 indicates a very strong and positive correlation between the expected and actual cancellations.

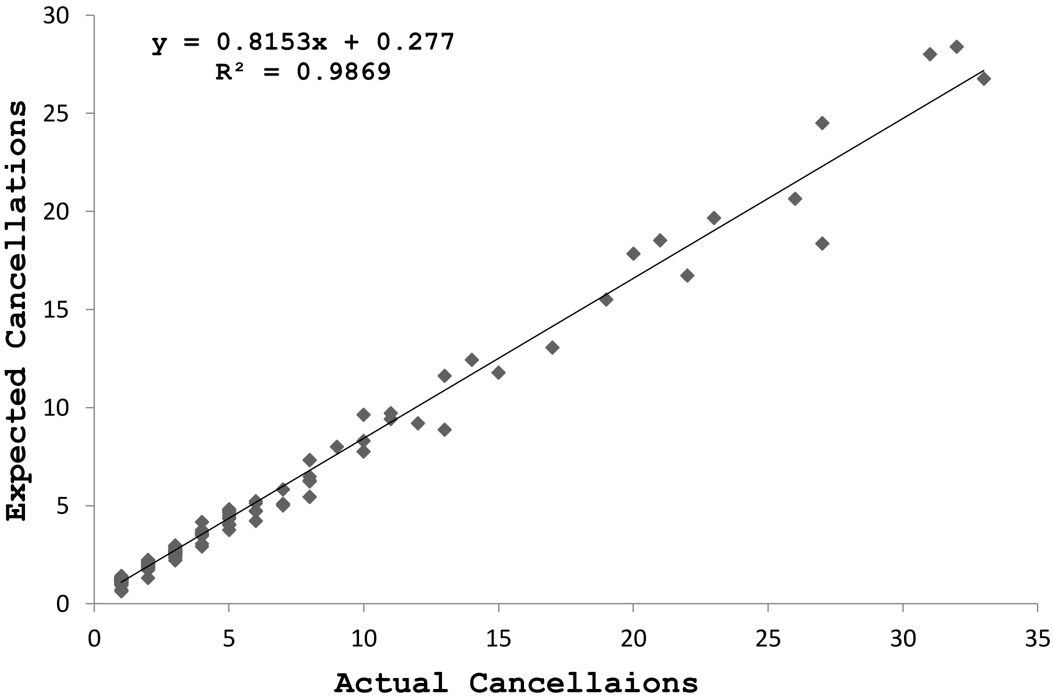

Figure 2: Correlation between actual and estimated flight cancellations model validation

6.2. Quantitative Validation

The paired t-test is used for the quantitative analysis. A paired t-test also referred to as a dependent sample t-test is usually used when the observations are not independent of one another. In this case, both the actual and expected cancellations are influenced by common weather factors. The paired t-test accounts for this. For each cancellation, the study is looking at the differences in the values of the two variables and testing if the mean of these differences is equal to zero.

Table 5 below, shows the list of variables (Actual and Expected Cancellations), the respective means of the variables (5.865 and 5.065), the 185 valid (i.e., non-missing) observations used in the t-test, the standard deviations (12.412 and 10.664) of the variables and the estimated standard deviation of the sample mean. The standard deviation of the sample mean provides a measure of the variability of the sample mean.

Table 5: Paired Samples Statistics

		Mean	N	s	SE Mean
Pair 1	Actual Cancellations	5.865	185	12.412	0.913
	Expected Cancellations	5.065	185	10.664	0.784

As Table 6 shows, the paired t-test forms a single random sample of the paired difference. The mean of these values between both cancellations is compared to 0 in the test. 0.8 is the mean within-cancellation difference between the two variables. 1.93 is the standard deviation of the mean paired difference. 0.142 is the estimated standard deviation of the sample mean. 0.52 to 1.08 is the lower and upper bound of the confidence interval (95%) for the mean difference. 5.637 is the t-statistic. It is the ratio of the mean of the difference to the standard error of the difference: (.8/1.143). 184 is the degrees of freedom for the paired observations, it is the number of observations minus 1. This is because the test is carried out on one sample of the paired differences. 0.00 is the two-tailed *p-value* computed using the t distribution. It is the probability of observing a greater absolute value of t under the null hypothesis. If the p-value is less than the pre-specified alpha level (0.05 for this study), it is then concluded that the mean difference between the two variables is statistically significantly different from zero and vice versa.

As presented in Table 6, the t-statistic is 5.637 with 184 *df*. The resultant two-tailed *p-value* is 0.000, which is less than 0.05. This study then concludes that the mean difference of the actual and expected flight cancellations is significantly different from zero and therefore acceptable.

Table 6: Paired Samples Test

		Paired Differences							
					95% CI of Diff.				
		Mean	s	SE Mean	Lower	Upper	t	df	Sig. 2tail
Pair 1	Actual-Expected Cancellations	0.800	1.930	0.142	0.520	1.080	5.637	184	0.000

CONCLUSİON

The main objective of this study was to develop and validate a model to estimate the expected number of cancellations originating from the DFW airport due to forecasted weather conditions. NB distributions were used to determine the predicted cancellations, which together with the actual cancellations were empirically weighted determine the expected cancellations by employing the EB model. The study site for this research is DFW International Airport, where 1,047 observations from 2011 to 2015 for each of the six variables are used for evaluation. The mathematical model was then validated using 185 observations of flight cancellations obtained in 2016. Both qualitative and quantitative statistical methods were used for validation. Both methods indicate that the approach and results used in this study are reasonable and acceptable. Although reliable, this study recommends that further research to be conducted by including additional data in the validation process. It is also recommended that data from other sites/airports be used in the evaluation process. The findings of this research can help both the airport and the airlines operators to optimize flight operations and resources. As a result, airlines will save funds and the airport will conduct flight operations effectively and efficiently by optimizing capacity. In addition, through these exercises, the airport can improve safety and minimize potential incursions hence minimizing or completely eradicating IROPS. The second part of this study will look at optimizing the model determined to allow for sensitivity analysis and efficient planning for both flight and airfield operations.

ACKNOWLEDGMENTS

This study would like to acknowledge these DFW personalities; Jim Crites (Former Executive Vice President of Operations), Paul Sichko (Vice President of Operations), Steve Tobey (Assistant Vice President Operations) and

Daniel Criswell (Operations Systems Research Analyst) for their insightful professional contribution ranging from assisting in identifying the problem to advising.

REFERENCES

1. Ausick, P.,October 29, 2014. *The High Cost of Cancelled Flights*. http://247wallst.com/services/2014/10/29/the-high-cost-of-cancelled-airline-flights/ (last Accessed June 20, 2017)
2. Biggeri, A., Catelan, D., & Baccini, M. (2013, June). *Regression to the mean*. In Advances in Latent Variables-Methods, Models and Applications.
3. Cao, B., Liu, Y., Qian, W. Y., & Wang, Y. H. (2015, August). *Construction of the evaluation system for regional road traffic safety and application based on a grey integrated evaluation model*. In Grey Systems and Intelligent Services (GSIS), 2015 IEEE International Conference on (pp. 395-403). IEEE.
4. Catalina, T., Virgone, J., & Blanco, E. (2008). *Development and validation of regression models to predict monthly heating demand for residential buildings*. Energy and buildings, 40(10), 1825-1832.
5. Cockrell Skinner, A., Goldsby, T. U., & Allison, D. B. (2016). *Regression to the Mean: A Commonly Overlooked and Misunderstood Factor Leading to Unjustified Conclusions in Pediatric Obesity* Research. Childhood Obesity, 12(2), 155-158.
6. Dickey, D. A. (2012, April). *Introduction to predictive modeling with examples*. In SAS Global Forum 2012.
7. Elvik, R. (2008). *The predictive validity of empirical Bayes estimates of road safety*. Accident Analysis & Prevention, 40(6), 1964-1969.
8. Gross, F., Persaud, B., & Lyon, C. (2010). *A guide to developing quality crash modification factors* (No. FHWA-SA-10-032).
9. Harewood, G. C. (2015). *Impact of "regression to the mean" on colonoscopy performance data*. Gastrointestinal endoscopy, 81(4), 1054-1055.
10. Jarrah, A. I., Yu, G., Krishnamurthy, N., & Rakshit, A. (1993). *A decision support framework for airline flight cancellations and delays*. Transportation Science, 27(3), 266-280.
11. Jones, C. USA TODAY February 2, 2015. *Canceled flights cost airlines, fliers millions*. http://www.usatoday.com/story/travel/flights/2015/02/02/canceled-flights-cost-airlines-and-fliers-millions/22750345/ (last Accessed September 28, 2016)
12. Lan, S., Clarke, J. P., & Barnhart, C. (2006). *Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions*. Transportation science, 40(1), 15-28.
13. Maina, E. V., Ford, A., & Robinson, M. (2015). *Estimating the Safety Benefits of Red Light Cameras at Signalized Intersections in Urban Areas Case Study: The City of Virginia Beach*. International Journal of Transportation Engineering, 3(1), 45-54.
14. Mao, X., Yang, Y., Cai, K. Q., & Yang, W. H. (2015, September). *A decision support method for flight cancellations in adverse weather: An airport perspective*. In 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC) (pp. 1E2-1). IEEE.
15. Niaparast, M., & Schwabe, R. (2013). *Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients.* Journal of Statistical Planning and Inference, 143(2), 296-306.
16. Poch, M., & Mannering, F. (1996). *Negative binomial analysis of intersection-accident frequencies*. Journal of transportation engineering, 122(2), 105-113.
17. Sargent, R. G. (2004, December). *Validation and verification of simulation models*. In Proceedings of the 36th conference on Winter simulation (pp. 17-28). Winter Simulation Conference.
18. Schoen, J. W., March 3, 2015. *Bad Weather Flight Cancellations Cost Americans $2.4 Billion*. http://www.thefiscaltimes.com/2015/03/03/Badr-Weather-Flight-Cancellations-Cost-Americans-24-Billion CNBC. (last Accessed June 20, 2017)
19. Seelhorst, M. (2014). *Flight Cancellation Behavior and Aviation System Performance*.http://www.nextor.org/pubs/NEXTOR-II-Flight-Cancellation-2014.pdf (last Accessed June 20, 2017)
20. Shephard, R. J. (2003). *Regression to the mean*. Sports medicine, 33(8), 575-584.

21. Sridhar, B., Wang, Y., Klein, A., & Jehlen, R. (2009, June). *Modeling flight delays and cancellations at the national, regional and airport levels in the United States*. In 8th USA/Europe ATM R&D Seminar, Napa, California (USA).
22. Tegge, R. A., Jo, J. H., & Ouyang, Y. (2010). *Development and application of safety performance functions for Illinois*. Research Report ICT-10-066. Illinois Center for Transportation. Illinois Department of Transportation, Springfield, IL.
23. TRB Online. *Continuous Dependent Variable Models*. http://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2chapter4.pdf (last Accessed May 2017)
24. Vida Maina, E., Forde, A., & Robinson, R. M. (2016). *Impact of optimally minimizing delay times on safety at signalized intersections in urban areas, case study: the city of Virginia Beach*. International Journal of Transportation Engineering, 3(4), 277-288.
25. Washington, S., & Wolf, J. (1997). *Hierarchical tree-based versus ordinary least squares linear regression models: theory and example applied to trip generation*. Transportation Research Record: Journal of the Transportation Research Board, (1581), 82-88.
26. Xiong, J., & Hansen, M. (2013). *Modelling airline flight cancellation decisions*. Transportation Research Part E: Logistics and Transportation Review, 56, 64-80.
27. Zhao, Y., James, W. H., & Cheryl, L. A. (2009). *A score test for overdispersion in Poisson regression based on the generalized Poisson-2*. Journal of Statistical Planning and Inference, 139(4), 1514-1521.
28. Zou, G. (2004). *A modified Poisson regression approach to prospective studies with binary data*. American journal of epidemiology, 159(7), 702-706.
29. Zou, Y., Zhang, Y., & Lord, D. (2013). *Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis*. Accident Analysis & Prevention, 50, 1042-1051.