Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)

Quantile Regression Based Estimation of Statistical Contingency Fuel

Lei Kang, Mark Hansen
Institute of Transportation Studies
University of California, Berkeley
Berkeley, CA, USA
lkang119@gmail.com, mhansen@ce.berkeley.edu

***Abstract*—Reducing fuel consumption is a unifying goal across the aviation industry. One fuel-saving opportunity for airlines is the possibility of reducing contingency fuel loading. Airline flight planning system (FPS) provides recommended statistical contingency fuel (SCF) for dispatchers. However, due to limitations of the current SCF estimation procedure, the application of SCF is limited. In this study, we propose to use quantile regression based method to estimate more reliable SCF values. Utilizing a large fuel burn dataset from a major U.S. based airline, we find that the proposed quantile regression method outperforms airline's FPS. We also quantify the impact of implementing improved SCF values in fuel saving. While maintaining a same safety level, it has been found that for our study airline, the benefit of applying improved SCF models is estimated to be in the magnitude of $8 million annual savings as well as 89 million kilogram CO_2 emission reduction. This study also builds a link between SCF95 estimation and aviation system predictability in which the proposed models can also be used to predict benefits from reduced fuel loading enabled by improved air traffic management (ATM) targeting on improved system predictability.**

Keywords- Flight Fuel Planning; Statistical Contigency Fuel; Quantile Regression; Gradient Boosting Machine; Random Quantiel Forests; Benefit Assessment

I. INTRODUCTION

Reducing fuel consumption is a unifying goal across the aviation industry. Air transportation contributes 8% of transportation greenhouse gas (GHG) emissions in the U.S. [1] and 10.6% of transportation emissions globally [2]. The global GHG emissions by 2020 from aviation are projected to be around 70% higher than in 2005 even if fuel efficiency improves by 2% per year [3]. Moreover, fuel cost accounts for about 15.5% of total operating expenses for U.S. passenger airlines [4]. As these shares are expected to increase dramatically, there is consequently intense focus on reducing fuel consumption from many stakeholders (e.g. governments, aircraft manufacturers, and airlines), who have undertaken a wide range of efforts and initiatives. On the government side, enhanced air traffic management (ATM) aiming at aviation system efficiency has been estimated to provide 6% to 12% savings in fuel consumption [5]. For instance, the benefit of the enhanced capacity of Next Generation Air Transportation System (NextGen) in the United States has been estimated at $132.5 billion from delay savings and environmental emissions reductions over the period 2013–2030 [6]. Another way of reducing fuel consumption is through improved fuel efficiency by employing alternative fuels or new designs of aircraft and engines. One example in this regard is that the addition of winglets to the wingtips of aircraft has been shown to improve the aerodynamics of aircraft and hence reduce fuel burn by 2.5%-5% [6]. However, given the difficulty, cost, and long timeline of reducing aviation fuel and GHG emissions with new procedures, new technologies, and alternative fuels, airlines are turning to a simple method to reduce fuel consumption: reducing aircraft weight [7]. Since the lighter the aircraft is, the less thrust is required from the engine and hence the less fuel would be consumed. For instance, airlines are purchasing aircraft made with lightweight materials [8], and charging passengers for luggage [9]. Other efforts include accommodating lighter weight seats and galleys and reducing drinking water loads [10]. However, though the biggest source of excess weight added to the aircraft is excess fuel [11, 12], there has been little discussion on reducing unnecessary fuel loading in flight fuel planning. Among very few studies in this regime, reference [7] finds that reducing unnecessary fuel loading by dispatchers could result in fuel savings on the order of $400 million per year based on the analysis of a major U.S. airline. In line with [7], by levering fuel loading and consumption data from six major U.S. airlines, reference [13] estimates that annual cost to carry unused fuel would range from $59 million to $667 million across airlines. Motivated by the significant fuel saving potentials of reducing unnecessary fuel loading, the goal of this paper is to explore fuel saving opportunities in fuel planning stage. Specifically, we will target on one key aspect of fuel loading, named contingency fuel (to be discussed in details later). By estimating reliable recommendation values for contingency fuel using quantile regression based method, we show that our study airline could achieve better fuel performance in terms of monetary saving and GHG emission reduction. In the next section, we will introduce fuel planning basics and identify opportunities to improve upon current contingency fuel loading practice.

II. INDUSTRY PRACTICE

In practice, airlines rely on flight dispatchers to perform the duty of flight planning including fuel planning. US Federal

Aviation Regulations [14] (FARs) require a domestic commercial flight to uplift enough fuel to complete the flight to the intended destination airport (mission fuel), as well as fly from the destination airport to the alternate airport (if required based on the weather forecast at the scheduled time of arrival) and hold in the air for 45 min at normal cruising speed (reserve fuel). These quantities are automatically calculated by the airline's flight planning system (FPS) after the dispatcher chooses a route of flight among several possible routes. Even though it is not required by the FARs, on top of mission fuel and reserve fuel, airline dispatchers uplift contingency fuel to be on the aircraft to hedge against various uncertainties (e.g. weather uncertainty, traffic congestion uncertainty, air traffic control uncertainty etc.) to ensure flight safety and reduce the risk of diversions. Contingency fuel uplift is based on a combination of corporate fuel policies and airline dispatchers' own experience and judgment. Fuel uplifted for alternate airports that are not required can serve much the same purpose as contingency fuel if the alternate is dropped from the flight plan during the course of the flight. (In rare instances, which are not considered here, extra fuel is loaded because it is less expensive to carry fuel into particular airports for subsequent use than it is to purchase it at these airports. This is known as "tinkering".)

To provide consistent and objective fuel planning, some FPS provides recommended contingency fuel numbers based on a statistical analysis of historical fuel consumption for similar flights. Carriers usually term this Statistical Contingency Fuel (SCF) [7, 15]. When a dispatcher goes to dispatch a flight (about 2 hours prior to departure), the FPS pulls historical data (the number of years prior to be specified by the airline) of all flights between the same Origin-Destination (OD) pair that were scheduled to depart in the same "hour bank" or time window specified by the airline. For each historical flight, the difference between the actual trip fuel consumption and the planned mission fuel consumption is calculated. If this difference is negative, it is then called "under-burn"; otherwise, it will be termed as "over-burn". We will call this the under/over-burn value. The FPS converts the under/over-burn value in pounds to minutes and estimates a normal approximation of the distribution of this excess required fuel burn. The 95th and 99th percentiles of the distribution, which are also called the SCF95 and the SCF99, will be provided to dispatchers by the FPS as guidelines for contingency fuel loading. The interpretation of SCF95 (SCF 99) is that based on historical fuel consumption, loading the quantity of contingency fuel specified by SCF95 (SCF99) would result in a flight being able to land without dipping into any reserve fuel 95% (99%) of the time. More details regarding SCF could be found in references [7, 15, 16].

SCF has been widely used in airline industry. Based on a survey [17], many airlines have SCF estimation functionalities (as described above) embedded in their FPSs. These include Air India, British Midland International, United Airlines, Virgin America, Virgin Atlantic, SAS Group of Airlines, to name a few. In the case of our study airline (a major U.S.-based network carrier), when it comes to compute SCF value, the set of similar historical flights are defined as those that took place over the previous year and have the same OD and scheduled hour of departure. However, there are several limitations of the current SCF estimation procedure. First of all, from a statistical perspective, the procedure assumes that under/over-burn is normally distributed, which may not be the case. Second, unless the sample is quite large, the estimate of a 95^{th} or 99^{th} percentile based on the sample mean and standard deviation is subject to considerable sampling error. Likewise, it is of course impossible to calculate SCF values in the case of serving a new OD market with no similar historical flights. Thirdly, although the SCF calculation has implicitly accounted for weather and other events in history by using actual fuel consumption information [15], dispatchers might still have low confidence in applying those numbers due to oversimplified grouping criterion (e.g. OD-hour). For example, aircraft type is missing from the grouping criterion, even though the aircraft performance models used for different aircraft types may have varying predictive performance. Additionally, in order to increase the confidence level of dispatchers in SCF values, weather forecast should also be explicitly taken into account. While the high percentiles used in SCF are intended to account for adverse weather, dispatchers are reluctant trust SCF values in such conditions. A previous analysis based on the same study airline reveals that dispatchers would almost always load extra fuel above recommended SCF values [7]. As a result, it has been found that 1.04% of the fuel consumed by an average flight is due to carrying additional contingency fuel above a reasonable buffer. Similar behavior has also been observed in other airlines. For instance, based on a survey of 50 U.S. pilots and dispatchers about their fuel loading practices, reference [18] finds that airline dispatchers and pilots always load contingency fuel above the suggested contingency value by the airline.

A companion paper on the behavioral aspect of dispatchers' contingency fuel loading decisions finds that contingency fuel loading is related to weather uncertainty as well as aviation system predictability [16]. These results suggest that since contingency fuel reflects a dispatcher assessment of flight uncertainty, improving system predictability can lead to reduction in contingency fuel loading. In this paper, instead of modeling the contingency fuel decision directly, we consider how to provide reliable SCF values that dispatchers believe. Ideally, their faith in these values would be such that they would generally adhere to them in setting contingency fuel. To overcome the limitations of widely used SCF estimation method described above, we propose a new SCF estimation procedure that relies on quantile regression models, focusing on SCF95. Quantile regression models have several desirable properties: 1) it models the 95^{th} quantile of under/over-burn value directly rather than employing simplified grouping criterion and assuming a normal distribution; 2) it allows covariates to be added into the estimation function so that characteristics such as weather and traffic can be explicitly controlled for; 3) this method also allows us to estimate SCF values for flights where the old method cannot be used because there is not an adequate sample of similar flights.

III. METHODOLOGICAL APPROACH

The widely used SCF95 estimation method relies on a simplified grouping criterion and normal approximation. In this section, we propose to use quantile regression based machine learning techniques to tackle these issues.

A. Data Collection

Data were collected from three sources: the fuel and flight statistics data from a major U.S.-based air carrier, the flight level performance data from the Federal Aviation Administration (FAA) Aviation system Performance Metrics (ASPM) database, and the terminal weather forecasts (TAF) information for major airports from the National Oceanic and Atmospheric Administration (NOAA) database. A major U.S.-based airline provided data for this study. This carrier operates an extensive domestic network. The dataset provided from the airline includes all U.S. domestic flights between April 2012 and July 2013. There are altogether 663,757 flights with eight major aircraft types. In addition to basic flight operation characteristics (i.e. aircraft type and OD airports), this dataset also contains flight-level fuel uplift, aircraft pushback weight, planned mission fuel (in minutes and pounds), contingency fuel, alternate airport and corresponding alternate fuel, reserve fuel, tinkering fuel. It also provides actual fuel burn quantities (in pounds and minutes) by flight phase, including taxi-out, airborne, and taxi-in. The FAA ASPM flight level database includes individual flight data for the 77 large airports in US. This dataset is used to capture historical traffic conditions. For a given flight in the airline data, we look at historical flights with same OD pair, scheduled departure hour, and month that occurred in the previous year. Then we calculate the mean and standard deviation of airborne time based on these historical similar flights. By assuming stable weather patterns for a given OD, month, and hour of day, this historical airborne time information will provide good approximation of possible weather conditions for a current flight. The deviation of historical airborne times and the corresponding flight plan airborne times serves as another measure of flight time variability. Hence, we also compute the mean and standard deviation of the difference between actual airborne time and planned airborne time based on a current flight's previous year's counterparts. The TAF data contains forecasted information about ceiling, visibility as well as indicators of the presence of thunderstorms and snow by hour, date, and airport. The most recent updated TAF weather data at the time of two hours prior to scheduled departure time has been merged with airline data to recreate the conditions seen by dispatchers during the time of flight planning.

In this study, we do not include tinkering fuel flights since it is an economic decision with less influence from dispatchers' personal behavior. After merging all three datasets, we end up with 335,394 flights with SCF95 values. There are also 31,261 flights with no SCF95 values in our dataset which presumably due to the difficulty of applying old SCF estimation method in those cases.

B. Summary Statistics

The relationship between contingency fuel loading and its corresponding SCF95 value is shown in Figure 1. The minimum observed SCF95 is 10 minutes. For each SCF95 category, its corresponding contingency fuel in general varies considerable and is found to be systematically higher than SCF95. This is consistent with the findings from reference [7] that dispatchers seldom trust recommended SCF95 values. Table 1 presents the mean, standard deviation, and 95th percentile of under/over-burn statistics for eight aircraft types. It can be observed that fuel performance differs across aircraft types. This also suggests the need to incorporate aircraft type into SCF estimation. It is also noted that the standard deviations of the under/over-burn distributions for different aircraft types stay relatively constant.

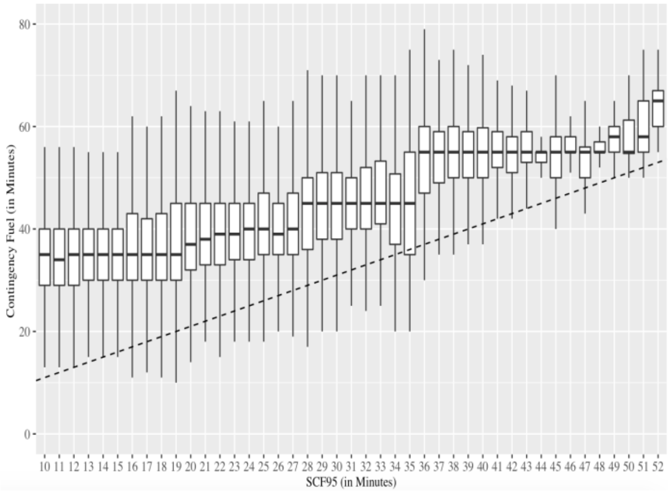

Figure 1 Contingency fuel loading (Minutes) vs SCF95 (Minutes)

TABLE 1 UNDER/OVER-BURN SUMMARY (IN MINUTES)

Aircraft type	Mean	Standard Deviation	95th percentile
A319	-9.5	9.4	5.9
A320	-7.6	9.7	8.1
B737-800	-4.0	9.8	12.7
B757-300	3.4	8.9	17.7
B757-200	5.0	10.1	21.7
DC9	7.2	7.6	19.6
MD88	7.5	9.0	22.2
MD90	1.4	8.0	14.2

C. Quantile Regression Method

In this section, we introduce our SCF95 estimation procedure based on quantile regression method. The dependent variable Y is the under/over-burn value (in minutes). The covariates X include terminal area weather forests, historical traffic conditions, aircraft types, departure hour window, departure month, and dummies for major airports. We focus on three techniques: 1) parametric quantile regression (QR); 2) gradient boosting machine (GBM) based quantile regression tree; 3) random quantile forests (RQF). Just as standard linear regression models conditional mean functions $E(Y \mid X)$, a quantile regression can be used to model conditional quantiles of Y : $Q_q(Y \mid X)$ and q denotes a specific quantile. It can be shown that the quantile regression estimator for *q*-th quantile minimizes the following loss function [19]:

$$J_q(\beta) = \sum_{i=1}^{N} \rho_q(y_i - f(x_i, \beta)) \qquad (1)$$

where $\rho_q(t) = t(q - \mathrm{I}_{(t<0)})$. In our case, we choose q to be 0.95. If we specify $f(x_i, \beta) = x_i'\beta$, minimizing (1) with respect to β would produce parameter estimates of the q-th conditional quantile function. Sometimes, in order to achieve better prediction performance, we could also leverage machine learning algorithms like gradient boosting machine (GBM) to estimate $f(x_i, \beta)$ in a non-parametric manner [20]. GBM is an ensemble learning algorithm which combines different base learners in a sequential fashion and gradually improves model fit, see equation (2).

$$\hat{f}_t \leftarrow \hat{f}_{t-1} + \alpha h_t(x^k) \tag{2}$$

The main idea of applying GBM in the quantile regression setting is to iteratively add simple regression tree models $h_t(x^k)$ (also called base learners) to existing model $\hat{f}_{t-1}$ so that the updated model $\hat{f}_t$ can further reduce the quantile loss function specified in equation (1). Here, $h_t(x^k)$ is constructed based on a splitting variable x^k which reduces the loss function the most at iteration t. α is called learning rate and it is usually set to be a small value. We use α = 0.005 in this paper. Number of iterations is a tuning parameter that we need to determine. If we set iteration to be a large number, we will be likely to overfit the data. Algorithm details could be found in [21].

Another powerful ensemble learning algorithm is called random forests (RF) which has been widely used in the area of conditional mean prediction. The idea of RF is to average the prediction outputs from a large number of decision trees [22]. For conditional mean, the prediction of a single decision tree for a new data point X = x is the mean response of Y in a particular leaf that contains X = x. Then, RF computes a final prediction by averaging predictions from all trees. Drawing analogy to conditional mean, when it comes to predicting conditional quantiles, instead of using mean response of Y in a leaf, we can report q-th empirical quantiles of Y in that leaf and then average the obtained quantiles across all trees. This algorithm is called random quantile forests (RQF) which has been proposed by [23]. The tuning parameter is RQF is the minimum node size which is related to how deep we should grow a decision tree. If we set the minimum node size to be a small number, then we would obtain a deep tree which is very likely to overfit the data.

To test the performance of three proposed quantile regression based models, in line with common practice in machine learning literature, we randomly divide flights with SCF values (335,394) into training set (60%), validation set (20%), and test set (20%). The 31,261 flights without SCF95 values will also serve as another test set. The idea is to build models using training set, select optimal tuning parameters with validation set, and compare model performance on test set. To evaluate the performance of new SCF95 predictions as compared to FPS SCF95, we will look at a loss function based goodness-of-fit measure [24]:

$$R(\beta) = 1 - \frac{J_q(f(\beta))}{J_q(\tilde{f}(\beta))} \tag{3}$$

where $J_q(f(\beta))$ is the value of loss function on test set using SCF95 estimation function $f(\beta)$. $\tilde{f}(\beta)$ is used to denote a model with the constant term only which is equivalent to use 95^{th} quantile of Y in the training set as prediction for every flight. Three proposed SCF95 estimation models plus FPS SCF95 will be measured against a constant only model developed using the training set. Regarding tuning parameters, in order to achieve the best prediction performance on test set, we should avoid overfitting the training data. Therefore, we need to find the optimal number of iterations in GBM and the minimum node size in RQF. Parameter tuning results will be provided in later sections.

IV. ESTIMATION RESULTS

In this section, we will present the estimation results and prediction performance of three proposed methods: QR, GBM, and RQF. Parametric QR model gives us the best model interpretation compared to machine learning models, thus we will focus on parameter estimates in the QR model. Prediction performance on test set across three models will also be discussed.

The estimation results of QR model are presented in Table 2. A positive parameter estimates means that an increase in the variable results in higher SCF95 prediction and vice versa. Regarding aircraft type, A319 is treated as baseline. The relative magnitudes of parameter estimates are consistent with Table 1 which suggests that heavy aircraft in general has higher SCF95 than smaller aircraft. Longer flights are also found to result in higher SCF95. This is partly because longer flights are more likely to experience en-route weather and traffic uncertainty. The signs of parameter estimates of historical traffic condition variables are all positive except for the mean airborne time which is statistically insignificant. Since we don’t have en-route weather forecast information, historical traffic predictability measures can serves as a good proxy for weather condition for a current flight assuming stable weather patterns for a given OD, month, and hour of day. The estimated results suggest that if historical traffic condition is less predictable as represented by large standard deviation of mean airborne time and large deviation between actual and planned airborne time, then SCF95 is higher. This also indicates that SCF95 could be reduced through enhanced ATM targeting on improving system predictability. Forecasted weather conditions for destination airports are found to have bigger impact than origin airports. Among terminal area weather forecast, forecasted thunderstorm is found to have the biggest impact on SCF95, followed by forecasted low ceiling and low visibility condition. The construction of low ceiling and low visibility variables is in accordance with the adverse weather definition of FARs. FARs require a flight to carry enough fuel to travel to an alternate airport if the weather conditions are such that visibility is less than 3 miles and the ceiling at the destination airport is less than 2000 feet at the flight’s estimated time of arrival plus/minus 1 hour. Forecasted low visibility and thunderstorm indicators at origin airports also have positive impact of SCF95. However, forecasted snow event at origin airports is found to have a counter-intuitive sign. One possible explanation is that snow effects at origin airports are very small and being cancelled out/or even reversed by month effects. Month indicator is also included in the model to capture seasonality effects. It can be found that winter season requires relative higher SCF95 to account for wind effects.

TABLE 2 ESTIMATION RESULTS FOR QUANTILE REGRESSION MODEL

Category	Variable	Estimates	T-stat
--	Intercept	-6.851	-8.00
Aircraft type (Baseline is A319)	A320	2.110	10.48
	B737-800	1.019	4.54
	B757-300	11.458	42.86
	B757-200	13.868	65.97
	DC9	16.133	49.13
	MD88	17.157	88.12
	MD90	9.427	44.18
Distance	Flight distance (in nautical miles)	0.004	3.21
Historical traffic condition	Mean of historical airborne time	-0.047 *	-1.57
	Standard deviation of historical airborne time	0.022	2.62
	Mean of difference between historical actual and planned airborne time	0.235	8.24
	Standard deviation of difference between historical actual and planned airborne time	0.188	13.31
TAF weather forecast for destination airports	Low visibility indicator (1-if lower than 3 miles, 0-otherwise)	4.249	7.51
	Low ceiling indicator (1-if lower than 2000 feet, 0-otherwise)	5.293	25.67
	Thunderstorm indicator (1-if thunderstorm presents, 0-otherwise)	9.534	17.49
	Snow indicator (1-if snow presents, 0-otherwise)	3.579	15.20
TAF weather forecast for origin airports	Low visibility indicator (1-if lower than 3 miles, 0-otherwise)	0.482	2.45
	Low ceiling indicator (1-if lower than 2000 feet, 0-otherwise)	-1.087 *	-1.40
	Thunderstorm indicator (1-if thunderstorm presents, 0-otherwise)	0.692	2.28
	Snow indicator (1-if snow presents, 0-otherwise)	-0.480	-2.69
Month (Baseline is January)	February	-0.374 *	-1.50
	March	-1.926	-8.66
	April	-0.580	-2.86
	May	-0.118 *	-0.57
	June	-1.533	-6.43
	July	-1.301	-5.20
	August	-1.255	-5.39
	September	-1.755	-7.91
	October	-1.430	-6.89
	November	-1.780	-8.69
	December	-0.384 *	-1.71
Number of observations	201,236		

Note:
1) To save space, airport fixed effects and departure hour fixed effects estimates are not presented in this table.
2) * denotes insignificant at 95% confidence level.

Turning to GBM training results, we have tested the number of iterations ranging from 1000 to 15000, and 8000 iterations is found to produce the smallest loss value on validation set. Regarding the node size selection in RQF, 50 is found to achieve the smallest validation set error. In addition, 200 decision trees are trained in RQF. Using the best model tuning parameters, the goodness-of-fit measure of three proposed models on test set and non-SCF flights test set are presented in Table 3. RQ, GBM and RQF are found to have similar model fitting performance, although RQF performs slightly better than the other two. Airline's FPS SCF95 is found to provide a poor fit on test set. As another performance comparison, if we load contingency fuel exactly as the recommended SCF95 values, our proposed methods yield a smaller percentage of flights landing with reserve fuel being used compared to airline's FPS (almost half of the FPS percentage). This demonstrates that our proposed models outperform airlines' FPS estimation procedure.

TABLE 3 GOODNESS-OF-FIT MEASURE

	Test set		Non-SCF test set	
	Goodness-of-fit measure	Percentage of landing with reserve fuel being used	Goodness-of-fit measure	Percentage of landing with reserve fuel being used
Quantile Regression	0.216	1.6%	0.205	1.6%
Gradient Boosting Machine	0.220	1.6%	0.205	1.6%
Random Quantile Forests	0.235	1.3%	0.219	1.5%
Airline FPS SCF95	0.073	3.0%	--	--

In Figure 2, we plot predicted SCF95 values against airline's FPS SCF95 values. We also consider the breakdown of terminal weather conditions. A weather impacted flight is defined as a flight with the following TAF weather forecast at destination airport: forecasted ceiling below 2000 feet, or visibility below 3 miles, or forecasted thunderstorm presence. For weather impacted flights, the quantile regression based models tend to predict higher SCF95 values than FPS. This is because terminal weather forecast has been explicitly taken into account in the SCF95 estimation process. This property is desirable for dispatchers because more confidence will be gained in making contingency fuel decisions. For non-weather impacted flights, three proposed methods tend to predict lower SCF95 values than FPS. Again, by adding weather, traffic, aircraft type information into SCF estimation, dispatchers can also potentially loading less contingency fuel which would lead to less fuel consumption. It is also noted that some SCF95 predictions are negative. In these few cases, the fuel burn predicted by the FPS will be high than the actual fuel burns more than 95% of the time. However, in order to evaluate potential fuel saving of applying the new SCF95, we will follow our study airline's SCF practice and set SCF95 values to be exactly 10 minutes if the corresponding prediction is less than 10.

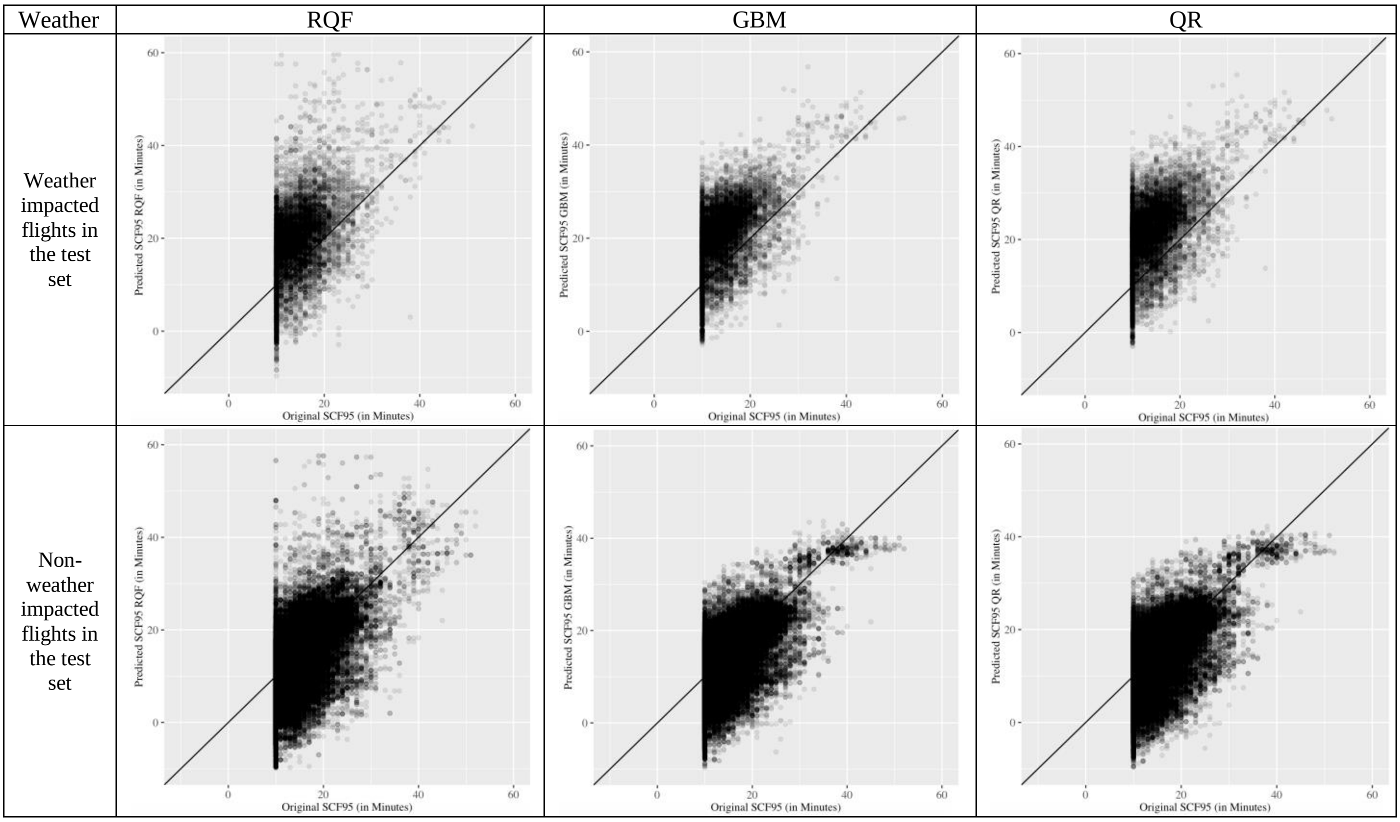

Figure 2 Model Predicted SCF95 v.s FPS SCF95

V. BENEFIT ASSESSMENT

In this section, we consider the possible fuel saving of implementing new SCF95 estimation procedure. As suggested in Figure 1, dispatchers in general load more contingency fuel than the SCF95 recommendation. Here we assume that because the SCF95 values obtained from our model are more believable (for example by being higher then there is adverse weather), dispatchers will follow them. Thus, the difference between contingency fuel loading and new SCF95 value defines our opportunity in fuel saving. This is because if we reduce contingency fuel loading, then the cost to carry contingency fuel would also decrease. If the difference is negative, we set it to zero. By assuming dispatchers follow new SCF95 recommendation perfectly in loading contingency fuel, we can compute fuel saving in terms of cost-to-carry (CTC) contingency fuel reduction. CTC is defined as the pounds of fuel consumed per pound of fuel carried per mile and it varies across aircraft types and flight distance. We borrow the estimated CTC factors from reference [7]. Those factors were estimated using Piano-5, an aircraft analysis tool and its estimates are not airline-specific. Therefore, the estimated CTC factors have wide applicability.

For a given flight *i* with aircraft type *a*, flight distance *d* (in miles), weight m (in pounds), and gate-to-gate fuel consumed *b* (in pounds), its CTC factor (in lbs/lbs-mile) is estimated by reference [7] in the following way. Gate-to-gate fuel consumption is assumed to a function of distance and weight as shown in equation 4.

$$b_{i,a} = \gamma_{1,a} m_{i,a} + \gamma_{2,a} m_{i,a} d_{i,a} + \gamma_{3,a} d_{i,a} \tag{4}$$

Then, the CTC factor can be expressed as

$$\gamma_{i,a} = \frac{\gamma_{1,a}}{d_{i,a}} + \gamma_{2,a} \tag{5}$$

where γ_1 and γ_2 are parameters associated with mass and mass-distance. Another way to interpret the fuel saving is in the form of CO_2 emission. We utilize the U.S. Environmental Protection Agency conversion factor for Jet Fuel [25] to translate fuel saving into the saving in CO_2 emission, in the unit of kg. For weather impacted flights in the regular test set and non SCF value test set, the per flight saving of fuel ranges from 109 to 121 lbs across different prediction methods. The estimated fuel savings for non-weather impacted flights are more consistent across three methods which are systematically higher than the saving estimates of weather impacted flights. This suggests that the current contingency fuel loading might be too conservative in good weather conditions hence leaving us more fuel saving opportunities. Since RQF performs slightly better than the other two methods on test set, we present the total savings in fuel consumption and CO_2 emission based on the RQF estimates. Since the test set is only a 20% random sample from the merged data set, if we extrapolate back to the original dataset with operations over 14 months for our study airline, we can further estimate an airline-wide total saving. The extrapolation is carried out is the following way: we first compute per flight basis weighted average CTC savings based on RQF estimates shown in Table 4 with number of flights as weights. By using $2/gallon as jet fuel price, we can extrapolate to airline-wide monetary saving by flight count. The airline-wide benefits are about $24 million fuel saving as well as 256 million kilogram CO_2 emission reduction. More careful extrapolation differentiating aircraft types, or airports would be an interesting future research direction; however, it is not the focus of this study, which concentrates on the flight-level impact.

Safety is a dispatcher's major consideration in contingency fuel loading. As shown in Table 3, if we load contingency fuel exactly as the proposed SCF95 values, we would still encounter a small proportion of flights using reserve fuel which is undesirable to airlines. To better apply our proposed SCF estimation method in practice, we try to find a safe buffer on top of proposed SCF95 that guarantees a same safety margin for our study airline. The safety benchmark is the percentage of flights landing with some reserve fuel being used based on actual contingency fuel loading. Based on the combined test set (including non-SCF value test set and regular test set), this safety benchmark is 0.10%. If we add a 15 minutes buffer to SCF95 prediction, we can reach a similar safety level (see Table 5). In Table 6, we compute the second order fuel saving by adding 15 minutes fuel to recommend SCF95 values. Based on RQF estimates, we can still achieve an airline-wide benefit of $8 million fuel saving as well as 89 million kilogram CO_2 emission reduction.

TABLE 5 SAFETY CHECK

Test set + Non SCF test set	Percentage of flights landing with reserve fuel being used	
	Use predicted SCF95 as contingency fuel	Use predicted SCF95 + 15 minutes as contingency fuel
Quantile Regression	1.62%	0.12%
Gradient Boosting Machine	1.61%	0.11%
Random Quantile Forests	1.41%	0.11%
Benchmark: Contingency fuel	0.10%	

TABLE 4 FIRST ORDER FUEL SAVING ESTIMATES

Data set		Fuel savings (lbs)			Number of flights	Estimates based on RQF			
		Quantile Regression	Gradient Boosting Machine	Random Quantile Forests		Test set Monetary savings ($)	Test set CO_2 emission (kg)	Airline-wide Monetary savings ($)	Airline-wide CO_2 emission (kg)
Test set	Weather impacted flights	110.5	115.8	122.1	9,812	3.55×10^5	3.70×10^6	2.45×10^7	2.56×10^8
	Non-weather impacted flights	125.6	126.3	124.8	57,267	2.13×10^6	2.22×10^7		
Non-SCF test set	Weather impacted flights	111.4	116.3	116.3	3,591	1.25×10^5	1.30×10^6		
	Non-weather impacted flights	118.9	119.4	118.2	27,670	9.76×10^5	1.02×10^7		

TABLE 6 SECOND ORDER FUEL SAVING ESTIMATES

Data set		Fuel savings (lbs)			Number of flights	Estimates based on RQF			
		Quantile Regression + 15 minutes	Gradient Boosting Machine + 15 minutes	Random Quantile Forests + 15 minutes		Test set Monetary savings ($)	Test set CO_2 emission (kg)	Airline-wide Monetary savings ($)	Airline-wide CO_2 emission (kg)
Test set	Weather impacted flights	38.0	41.5	47.6	9,812	1.39×10^5	1.45×10^6	8.56×10^6	8.94×10^7
	Non-weather impacted flights	43.3	43.7	43.7	57,267	7.47×10^5	7.80×10^6		
Non-SCF test set	Weather impacted flights	28.0	28.1	23.1	3,591	2.48×10^4	2.59×10^5		
	Non-weather impacted flights	32.1	33.2	32.7	27,670	2.70×10^5	2.82×10^6		

VI. CONCLUSIONS

This analysis shows the possibility to reduce fuel consumption through an improved SCF95 estimation procedure. A quantile regression based SCF95 estimation procedure has been proposed. Three estimation models including parametric quantile regression, gradient boosting machine, and random quantile forests are found to outperform airline's FPS in SCF95 estimation. RQF is also found to perform slightly better than the other two proposed models. The new SCF estimation procedure overcomes the limitations of the widely used SCF estimation method which relies on simplified grouping criterion and normal approximation. The proposed method can also incorporate terminal weather forecast and historical traffic condition into the SCF95 estimation.

With the help of cost-to-carry factors proposed by reference [7], we are also able to calculate the extra fuel burn to carry the difference between actual contingency fuel and new SCF95 value based on model prediction. The extra fuel burn then can be translated into monetary costs and CO_2 emission. The estimated benefit pool for our study airline is in the magnitude of $24 million fuel saving and 256 million kilogram CO_2 emission reduction over 14 months operation. We further investigate the impact of adding a practical safety buffer (15 minutes) which helps achieve a similar safety level, as measured by the fraction of flights landing without their full 45 minute fuel reserve, as the current practice for our study airline. Even after adding 15 minutes, the estimated benefits are still significant: $8 million fuel saving and 89 million kilogram CO_2 emission reduction. In addition, this study also builds a link between SCF95 estimation and aviation system predictability in which the proposed models can also be used to predict benefits from reduced fuel loading enabled by improved ATM.

In this study, we focus on SCF95 estimation. A similar analysis based on SCF99 also deserves further investigation. Our proposed method can be easily extended to other SCF estimation problems including SCF99. Based on the fuel burn data obtained from a major U.S.-based airline, a significant benefit has been estimated by improving SCF estimation. It is also expected that a large scale system-wide benefit could also be gained by applying new SCF estimation in flight fuel planning. Moreover, given the link between system predictability and SCF95 estimation, a system-wide fuel saving benefit due to improved ATM could also be assessed.

REFERENCES

[1] Enviornmental Protection Agency. Fast Facts: U.S. Transportation Sector Greenhouse Gas Emissions 1990-2014, 2016. Available at: https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100ONBL.pdf

[2] Intergovernmental Panel on Climate Change. Climate Change 2014: Mitigation of Climate Change, Cambridge University Press, 2014.

[3] International Civil Aviation Organization. ICAO Environment Report 2013: Aviation and Climate Change, 2014. Available at: http://cfapp.icao.int/Environmental-Report-2013/files/assets/common/downloads/ICAO_2013_Environmental_Report.pdf

[4] Bureau of Transportation Statistics. Airline Financial Data, 2016. Available at: https://www.rita.dot.gov/bts/press_releases/bts061_16 [Accessed 25 January 2017]

[5] Intergovernmental Panel on Climate Change. Aviation and the Global Atmosphere, 1999. Available at: https://www.ipcc.ch/pdf/special-reports/spm/av-en.pdf

[6] Federal Aviation Administration. Next Gen: the Business Case for the Next Generation Air Transportation System, 2014. Available at: https://www.faa.gov/nextgen/media/BusinessCaseForNextGen-2014.pdf

[7] M.S. Ryerson, M. Hansen, L. Hao, M. Seelhorst. Landing on Empty: Estimating the Benefits from Reducing Fuel Uplift in U.S. Civil Aviation. Environmental Research Letters 10, 2015, 094002.

[8] D.S. Lee, D. Fahey, P. Forster, P., Newton, R. Wit, L. Lim, B. Owen, and R. Sausen. Aviation and Global Climate Change in the 21st Century, Atmospheric Environment 43, 2009, pp: 3520-3537.

[9] R. Abeyratne. Air Carrier Responses to the Fuel Price Crisis: Some Issues. Global Trade and Customs Journal 4, 2009, pp: 31-43.

[10] European Commission. Fuel and Air Transport, 2015. Available at: http://ec.europa.eu/transport/modes/air/doc/fuel_report_final.pdf

[11] M.H. Sadraey. Aircraft Design: A Systems Engineering Approach. John Wiley & Sons, 2012.

[12] M. Irrgang. A Look at the State of Airline Fuel Conservation, 2011. Available at: http://www.agifors.org/document.go?documentId=2131&action=download

[13] L. Kang, M. Hansen, and M.S. Ryerson. Predictability Evaluation Based on Gate-In Fuel Prediction and Cost-to-Carry Estimation, presented at the Transportation Research Board 95th Annual Meeting, 2016. Available at: https://trid.trb.org/view.aspx?id=1394310

[14] E-CFR: Title 14: Aeronautics and Space Part 121, 2015. Operating Requirements: Domestic, Flag, and Supplemental Operations. Retrieved from http://www.ecfr.gov/cgi-bin/text-idx?SID=a8d3c4800d167b64bbfa2349ec337755&mc=true&node=pt14.3.121&rgn=div5

[15] S.E. Karisch, S.S. Altus, G. Stojković, and M. Stojković. Operations Quantitative Problem Solving Methods in the Airline Industry, A Modeling Methodology Handbook Series (International Series in Operations Research & Management Science), New York, Springer, 2012.

[16] L. Hao, M. Hansen, and M.S. Ryerson. Fueling for Contingencies: The Hidden Cost of Unpredictability in the Air Transportation System, Transportation Research Part D: Transport and Environment 44, 2016, pp: 199-210.

[17] G. Schiefer and T. Samuel. Managing Flight Operational Costs, 2011. SABRE. Available at: http://www.sabreairlinesolutions.com/pdfs/ManagingFlightOperationalCosts.pdf

[18] A. Trujillo. Uncertainties that Flight Crews and Dispatchers Must Consider When Calculating the Fuel Needed for a Flight. NASA Technical Memorandum 110240, 1996. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19960042496

[19] R. Koenker. Quantile Regression. Cambridge University Press. ISBN 0-521-60827-9, 2005.

[20] J.H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine, Annals of Statistics 29(5), 2001, pp: 1189-1232.

[21] G. Ridgeway. Generalized Boosted Models: A guide to the gbm package, Available at: http://www.saedsayad.com/docs/gbm2.pdf

[22] L. Breiman. Random forests. Machine Learning 45, 2001, pp: 5–32.

[23] N. Meinshausen. Quantile Regression Forests, Journal of Machine Learning Research 7, 2006, pp: 983–999.

[24] R. Koenker, J. Machado. Goodness of Fit and Related Inference Processes for Quantile Regression 94, 1999, pp: 1296-1310.

[25] EPA, 2013. Emission Factors for Greenhouse Gas Inventories, US Environmental Protection Agency. http://www.epa.gov/climateleadership/documents/emission-factors.pdf. Last accessed: January 2015.

AUTHOR BIOGRAPHY

Lei Kang is a Ph.D. candidate of the Institute of Transportation Studies in the Department of Civil and Environmental Engineering, University of California, Berkeley. He received a Master of Arts degree in Biostatistics from the Division of Biostatistics at University of California, Berkeley. He also obtained his Master's degree in Transportation and Infrastructure Systems Engineering from Purdue University. He is currently a member of the Committee on Airfield and Airspace Capacity and Delay, Transportation Research Board. His research interests are in the application of statistical methods and machine learning techniques to air traffic management and airline fuel loading decisions

Mark Hansen (BA'80, MCP'84, PhD'88) is a Professor of Civil and Environmental Engineering at the University of California, Berkeley. He graduated from Yale with a Bachelor's degree in Physics and Philosophy in 1980, and has a PhD in Engineering Science and a Masters in City and Regional Planning from UC Berkeley. Prior to graduate school, Dr. Hansen worked as a physicist at the Environmental Protection Agency. Since joining the Berkeley faculty in 1988, he has led transportation research projects in urban transportation planning, air transport systems modeling, air traffic flow management, aviation systems performance analysis, aviation safety, and air transport economics. He has taught graduate and undergraduate transportation courses in economics, systems analysis, planning, probability and statistics, and air transportation. Professor Hansen is the Berkeley co-director of the National Center of Excellence in Aviation Operations Research, a multiuniversity consortium sponsored by the Federal Aviation Administration. He is the former Chair of Transportation Research Board Committee AV-060, Airport and Airspace Capacity and Delay.