Statistical Modeling of Aircraft Takeoff Weight

Yashovardhan S. Chati and Hamsa Balakrishnan
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: {yschati, hamsa}@mit.edu

***Abstract*—The Takeoff Weight (TOW) of an aircraft is an important aspect of aircraft performance, and impacts a large number of characteristics, ranging from the trajectory to the fuel burn of the flight. Due to its dependence on factors such as the passenger and cargo load factors as well as operating strategies, the TOW of a particular flight is generally not available to entities outside of the operating airline. The above observations motivate the development of accurate TOW estimates that can be used for fuel burn estimation or trajectory prediction.**

This paper proposes a statistical approach based on Gaussian Process Regression (GPR) to determine both a mean estimate of the TOW and the associated confidence interval, using observed data from the takeoff ground roll. The predictor variables are chosen by considering both their ease of availability and the underlying aircraft dynamics. The model development and validation are conducted using Flight Data Recorder archives, which also provide ground truth data.

The proposed models are found to have a mean TOW error of 3%, averaged across eight different aircraft types, resulting in a nearly 50% smaller error than the models in the Aircraft Noise and Performance (ANP) database. In contrast to the ANP database which provides only point estimates of the TOW, the GPR models quantify the uncertainty in the estimates by providing a probability distribution.

Finally, the developed models are used to estimate aircraft fuel flow rate during ascent. The TOW estimated by the GPR models is used as an input to the fuel flow rate estimation. The proposed statistical models of the TOW are shown to enable a better quantification of uncertainty in the fuel flow rate as compared to the deterministic ANP models, or to models that do not use the TOW as an explicit input.

***Index Terms*—statistical modeling; Takeoff Weight (TOW); fuel flow rate; Flight Data Recorder (FDR); takeoff ground roll**

I. Introduction

The Takeoff Weight (TOW) of an aircraft is an essential parameter for modeling or estimating its trajectory and fuel consumption, as well as other aircraft performance characteristics, such as, its rate of climb/descent, range, endurance, ceiling, and takeoff distance [1]. However, it is not generally available outside the operating carrier, due to its dependence on proprietary information such as load factors and operational strategies. The above facts motivate the development of models to estimate the TOW of a flight from accessible information.

Aircraft design studies have traditionally estimated the TOW by considering its components, namely, the payload weight, stage length fuel weight, operating empty weight, reserve fuel weight, and alternative fuel weight [2, 3, 4]. This approach is effective for studies in which the payload weight is an input. It can also be used to estimate the average TOW of an aircraft type over a set of operations for which the average passenger load factor is available [5]; for example, average passenger load factors for different origin-destination pairs are published in the United States by the Department of Transportation [6]. However, this method cannot be easily extended to estimate the TOW of a particular flight, as load factors of individual flights are not publicly known.

Prior studies have estimated the TOW for a particular flight using simulated or real aircraft trajectory information during the climb phase [7, 8, 9, 10]. They typically estimate an equivalent TOW such that the power in climb modeled using the equivalent TOW matches the energy rate observed on past trajectory points. They estimate the equivalent TOW using either an adaptive mechanism or least squares algorithms. Machine learning techniques have also been applied to radar data to estimate the TOW in order to predict the future aircraft trajectory [11]. The methods proposed in these studies are shown to be superior to the EUROCONTROL's Base of Aircraft Data (BADA) method for trajectory modeling [12]. However, due to the unavailability of ground truth data, the accuracies of the TOW estimates in the studies are not known. Instead, these models have been evaluated based on the trajectory prediction accuracy.

Recent work has used runway ADS-B data during takeoff to model the operational TOW, using analytical methods or methods based on least squares [13]. However, the resultant TOW estimates could not be validated due to the unavailability of ground truth data. Moreover, these studies assume no deration in the takeoff thrust, and a standard coefficient of friction for the ground roll, which result in approximate estimates of the operational TOW.

A. Contributions of this Paper

We apply statistical machine learning techniques to model the operational TOW by using flight data from the takeoff ground roll. Random disturbances affecting an aircraft's operation (for example, component manufacturing tolerances, turbulence, fluctuations in ambient atmospheric conditions, and component aging and deterioration) motivate the development of a statistical model over a deterministic one [14]. Data from the Flight Data Recorders (FDRs) of real flight operations of a major airline allow us to validate our models using ground truth data (the actual TOW for each flight). The proposed machine learning techniques, based on Gaussian Process Regression (GPR), enable the estimation of the mean TOW of

This research was supported in part by the U.S. National Science Foundation under CPS: Large: ActionWebs (award no. 0931843).

a flight as well as the underlying uncertainty distribution. The uncertainty distribution captures the cumulative effect of unmodeled factors and random effects on the estimate of the TOW.

Our key contribution is the development and validation of models that map the trajectory variables during takeoff ground roll to the TOW. Although the proposed models are built using FDR data (which also enable validation of the models), they can now be used to estimate the TOW and operational fuel burn for flights given trajectory data, for example, from ADS-B, ASDE-X, or other surface surveillance sources. The proposed models are shown to have a mean error of 3% (averaged across different aircraft types) in predicting the TOW of flights in an independent test set, which is a nearly 50% reduction in error compared to that given by the Aircraft Noise and Performance (ANP) database [15]. The ANP database gives default TOWs for different aircraft types based on the stage/trip length. In employing these TOW estimates in a GPR model of the fuel flow rate of A321-111 flights in ascent, we show that the averaged mean error in the fuel flow rate estimation is 4.4%, and the projected 95% confidence interval contains 94% of the observations in the independent test set on an average.

B. Outline

We start with briefly describing the dataset in Sec. II. In Sec. III, we describe the features selected for the regression model of the TOW. Section IV provides a brief primer on Gaussian Process Regression. The application of GPR to the problem of modeling the TOW is explained in Sec. V. Section VI presents the evaluation of our models using an independent test dataset. In this section, the model estimates are also compared to those given by the Aircraft Noise and Performance (ANP) database [15]. It is worth noting that the ANP database is used for TOW estimation by the FAA's Aviation Environmental Design Tool (AEDT) [4], a widely used aircraft performance modeling tool. In Sec. VII, we show how the TOW estimates given by our models can be used to estimate fuel flow rates in climb/ascent. Finally, we present the main conclusions of this study, and directions for future research in Sec. VIII.

II. Description of data

The operational flight data used in this study are obtained from the Flight Data Recorders (FDRs) of a major airline. The FDR records the values of aircraft and engine parameters during flight and is therefore, an accurate source of operational flight data.

The objective of our study is to use the limited amounts of FDR data available in order to develop models to estimate the TOW of a particular flight, given trajectory variables from the takeoff roll, and other accessible parameters (such as, the ambient weather conditions at the airport). The intent is that the proposed models could then be used even in the absence of FDR data, which are not generally available.

This study considers data from eight different aircraft types. The details of the aircraft and engine types, the Maximum Takeoff Weight (MTOW) specifications, and the number of flights for each aircraft type in our dataset are shown in Tab. I. The aircraft types included in this study represent a wide range of MTOW, from 75.5 metric tons for the A319 to 365 metric tons for the A340. The FDR dataset includes the aircraft trajectory, speeds, gross weight, acceleration, fuel flow rate, engine temperatures, ambient pressure and temperature, positions of auxiliary devices and control surfaces, etc. as a function of time. Each trajectory is divided into different flight phases [16]; this paper focuses only on the takeoff ground roll.

It is worth noting that as per the convention in aviation, the term *weight* in this paper refers to the *mass* of the aircraft in the physical sense.

TABLE I. FDR DATA: AIRCRAFT TYPES AND ENGINES.

Aircraft Type	**Engine Type**	**MTOW (kg)**	**# Flts.**
A319-112	2×CFMI CFM56-5B6/2 or 2P	75,500	130
A320-214	2×CFMI CFM56-5B4/2 or P/2P	77,000	169
A321-111	2×CFMI CFM56-5B1/2 or 2P	89,000	117
A330-202	2×GE CF6-80E1A4	217,000	84
A330-243	2×RR Trent 772B-60	233,000	100
A340-541	4×RR Trent 553	365,000	52
B767-300	2×GE CF6-80C2B7F	184,612	91
B777-3FX(ER)	2×GE GE90-115B1	351,534	131

III. Model variables

We use a physical understanding of aircraft dynamics during takeoff ground roll in order to select the appropriate features/variables to build our TOW models. Fig. 1 shows the free-body diagram of an aircraft with the forces acting on it.

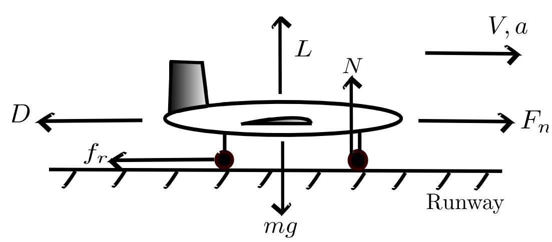

Figure 1. Schematic of airplane dynamics during takeoff ground roll.

The equations of motion during takeoff ground roll are as follows:

$$L+N = mg \quad (1)$$

$$F_n - D - f_r = ma \quad (2)$$

$$L = q\mathscr{S}C_L \quad (3)$$

$$D = q\mathscr{S}C_D \quad (4)$$

$$f_r = \mu N \quad (5)$$

$$a = \frac{dV}{dt} \quad (6)$$

$$q = \frac{1}{2}\rho V^2 \quad (7)$$

Here, L is the lift on the aircraft, N is the normal reaction from the ground, m is the aircraft TOW, g is the acceleration due to gravity, F_n is the net thrust on the aircraft, D is the air drag on the aircraft, f_r is the frictional force from the ground, a is the aircraft longitudinal acceleration, q is the dynamic pressure on the aircraft, $\mathscr{S}$ is the wing reference area, C_L and C_D are the coefficients of lift and drag, respectively during takeoff, μ is the coefficient of friction, V is the aircraft velocity, t is the time, and ρ is the ambient air density. Neglecting wind speeds during takeoff ground roll, the aircraft airspeed is assumed to be equal to the ground speed (V). The mass of fuel consumed during the takeoff ground roll is assumed to be small compared to the aircraft mass, so that the aircraft weight is effectively constant and equal to the TOW (m) during the takeoff ground roll. The coefficients of lift and drag, governed by the aircraft configuration, are also assumed to be constant during this flight phase. The net thrust on the aircraft is the averaged net thrust per engine times the number of engines (n_{eng}). The net thrust per engine is assumed to be a function of the static thrust (F_0) and the aircraft velocity [13]. The static thrust is the net thrust which would be produced by the engine if the aircraft were at rest at the set throttle setting. During the takeoff roll, the throttle setting does not change. The static thrust is assumed to be a function of the thrust deration level (η), the ambient air density during the takeoff roll (ρ) and the maximum sea level, static engine thrust (F_{00}). The net thrust on the aircraft is therefore, governed by the following functional relation:

$$F_n = F_n(n_{\text{eng}}, V, \eta, \rho, F_{00}) \tag{8}$$

The distance covered during takeoff roll (S) can be calculated by the following equation:

$$S = \int_{V_1}^{V_2} V \frac{dV}{a} \tag{9}$$

Here, V_1 is the aircraft velocity at the start of the takeoff ground roll and V_2 is the aircraft velocity at wheels-off at the end of the takeoff ground roll. Combining (1)–(9), the TOW can be expressed by the following functional relation:

$$m = m(S, \rho, V_1, V_2, \mathscr{S}, F_{00}, C_L, C_D, \mu, \eta, n_{\text{eng}}) \tag{10}$$

$\mathscr{S}$, F_{00}, and n_{eng} are constant for a given aircraft/engine type.

We restrict our modeling variables to only those which can be obtained or derived from easily accessible databases. The ground roll distance and aircraft velocities during ground roll can be derived from surface surveillance data, while the ambient air density can be obtained from airport weather data. By contrast, the values of the aircraft lift and drag coefficients, coefficient of friction, and thrust deration level are difficult to obtain, and are therefore, not included in the model. For a particular aircraft type, our model uses the ground roll distance (S, in m), the ambient air density (ρ, in kgm^{-3}) during roll, the aircraft velocity (ground speed) at the start of the takeoff ground roll (V_1, in ms^{-1}), and the aircraft velocity (ground speed) at the end of the takeoff ground roll (V_2, in ms^{-1}) as the predictor/input variables. The predicted/output variable is the aircraft TOW (m, in kg). In other words, we consider a model of the form

$$m = m(S, \rho, V_1, V_2). \tag{11}$$

The unmodeled variables will contribute to the uncertainty of the TOW estimate, and will be reflected in the prediction intervals provided by our statistical models.

IV. Gaussian Process Regression

The models proposed in this paper employ a machine learning technique known as Gaussian Process Regression (GPR). GPR is a powerful nonparametric Bayesian approach, with a Gaussian probabilistic framework. It has been successfully applied to diverse areas, including biomedical applications and health care [17, 18, 19], remote sensing [20, 21], music [22], robotics [23], cellular communications [24], and material microstructure analysis [25].

As all the model variables of interest to us are metric and continuous, the problem is well-suited to the use of regression. In this section, we briefly describe the GPR methodology, more details about which can be found in [26, 27].

A regression model is given by

$$y = f(\mathbf{x}) + \varepsilon, \tag{12}$$

where, y is the predicted/output/dependent variable, $\mathbf{x}$ is the predictor/input/independent vector, $f(\mathbf{x})$ is the underlying regression function that we wish to estimate, and ε is the noise with which the dependent variable is distributed about the regression function. Under GPR, we assume the regression function to follow a Gaussian Process (GP) prior. A function $f(\mathbf{x})$ is said to follow a Gaussian Process if the function values at any finite set of inputs $\mathbf{x}$ follow a joint Gaussian distribution [26]. Under a GP then,

$$f(\mathbf{x}) \sim GP(m_e(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{13}$$

where, $m_e(\mathbf{x})$ is the mean function, and $k(\mathbf{x}, \mathbf{x}')$ is the kernel/covariance function over two inputs $\mathbf{x}$ and $\mathbf{x}'$, which governs the covariance among function values as $k(\mathbf{x}, \mathbf{x}') = \text{cov}(f(\mathbf{x}), f(\mathbf{x}'))$. Under GPR, the mean function is often assumed to be the zero function. It is common to assume the noise to be drawn independently from a Gaussian distribution, $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, with mean 0 and noise variance σ_n^2. Under the assumption of a zero mean function for the GP governing the regression function and independent Gaussian noise, the dependent variable y also follows a GP with a zero mean function and a 'noisy' kernel function $k_{\text{noise}}(\mathbf{x_p}, \mathbf{x_q})$ over d-dimensional input vectors $\mathbf{x_p}$ and $\mathbf{x_q}$,

$$y \sim GP(0, k_{\text{noise}}(\mathbf{x_p}, \mathbf{x_q})). \tag{14}$$

The noisy kernel function for the dependent variables $k_{\text{noise}}(\mathbf{x_p}, \mathbf{x_q})$ relates to the kernel function for the regression function values $k(\mathbf{x_p}, \mathbf{x_q})$ as follows:

$$k_{\text{noise}}(\mathbf{x_p}, \mathbf{x_q}) = k(\mathbf{x_p}, \mathbf{x_q}) + \sigma_n^2 \delta_{pq} \tag{15}$$

Here, δ denotes the Kronecker delta.

The choice of different kernel functions affects the nature of the regression functions used for modeling, and gives GPR great modeling flexibility. Two commonly-used kernel functions are the following:

- Dot Product Squared Exponential (DPSE) kernel: This kernel function is used to model very smooth functions. It is given by:

$$
\begin{aligned}
k(\mathbf{x_p}, \mathbf{x_q}) &= \sigma_0^2 + \mathbf{x_p}^T \Sigma \mathbf{x_q} \\
&+ \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{\ell_i^2} \right) \\
\Sigma &= \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2) \quad (16)
\end{aligned}
$$

- Dot Product Exponential (DPE) kernel: This kernel function is used to model very rough functions. It is given by:

$$
\begin{aligned}
k(\mathbf{x_p}, \mathbf{x_q}) &= \sigma_0^2 + \mathbf{x_p}^T \Sigma \mathbf{x_q} \\
&+ \sigma_f^2 \exp\left(-\sqrt{\sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{\ell_i^2}} \right) \\
\Sigma &= \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2) \quad (17)
\end{aligned}
$$

In (16) and (17), $\mathbf{x_p}$ and $\mathbf{x_q}$ are d-dimensional input column vectors, σ_0^2 is the constant variance parameter, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2$ are the variance parameters for each of the d input dimensions, σ_f^2 is a variance parameter governing the magnitude of the exponential part of the kernel, ℓ is the d-dimensional vector of length scales (one for each input dimension), and the subscript i refers to the i^{th} component of the vector. These kernel parameters are referred to as *hyperparameters* in GPR. Thus, the hyperparameter vector for both the DPSE and the DPE kernels is $[\sigma_0^2\ \sigma_1^2\ \ \sigma_2^2\ \ \ldots\ \ \sigma_d^2\ \ \sigma_f^2\ \ \ell]^T$.

Numerous other kernel functions exist, details of which can be found in [27].

The noisy kernel hyperparameter vector θ is the kernel hyperparameter vector mentioned above with the noise variance σ_n^2 appended. It is estimated as the vector which maximizes the log posterior probability of the hyperparameter vector, given the matrix of input vectors $\mathbf{X}$ and the vector of dependent variable values $\mathbf{y}$,

$$
\begin{aligned}
\hat{\theta} &= \underset{\theta}{\mathrm{argmax}} \log p(\theta | \mathbf{X}, \mathbf{y}) \\
&= \underset{\theta}{\mathrm{argmax}} \left\{ \log p(\theta) - \frac{1}{2} \mathbf{y}^{\mathbf{T}} \mathbf{K_y^{-1}} \mathbf{y} - \frac{1}{2} \log |\mathbf{K_y}| - \frac{n}{2} \log(2\pi) \right\}.
\end{aligned} \quad (18)
$$

Here, $p(.)$ refers to the Probability Distribution Function (PDF) over the argument, $p(\theta)$ is the prior distribution on the hyperparameter vector, n is the number of observations, $\mathbf{X}$ is the $n \times d$ matrix of d-dimensional inputs, $\mathbf{y}$ is the $n \times 1$ vector of the dependent variable values, and $\mathbf{K_y}$ is the $n \times n$ covariance matrix derived from the noisy kernel function over pairs of input variables (15).

In this paper, the aim of GPR is to make predictions on new data points (and not hyperparameter inference, for example). For GPR, the predictive distribution of the dependent variable values $\mathbf{y}^*$ at a set of new inputs $\mathbf{X}^*$ is also a Gaussian distribution, given by:

$$
\begin{aligned}
\mathbf{y}^* | \mathbf{X}^*, \mathscr{D} &\sim \mathscr{N}(\mu, \mathscr{C}) \\
\mu &= \mathbf{K}(\mathbf{X}^*, \mathbf{X}) \mathbf{K_y^{-1}} \mathbf{y} \\
\mathscr{C} &= \mathbf{K}(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{K}(\mathbf{X}^*, \mathbf{X}) \mathbf{K_y^{-1}} \mathbf{K}(\mathbf{X}^*, \mathbf{X})^{\mathbf{T}} + \sigma_n^2 \mathbf{I_{n^*}}
\end{aligned} \quad (19)
$$

Here, n^* is the number of new inputs at which predictions are desired, $\mathbf{X}^*$ is the $n^* \times d$ matrix of the set of new inputs, $\mathscr{D} = (\mathbf{X}, \mathbf{y})$ is the set of training inputs and dependent variable values (used for hyperparameter inference), $\mathscr{N}(\mu, \mathscr{C})$ refers to a multivariate Gaussian distribution with mean vector μ and covariance matrix $\mathscr{C}$, $\mathbf{K}(\mathbf{X}^*, \mathbf{X})$ is the $n^* \times n$ covariance matrix derived from the noisy kernel function over pairs of new and training input variables (15), $\mathbf{K}(\mathbf{X}^*, \mathbf{X}^*)$ is the $n^* \times n^*$ covariance matrix derived from the noisy kernel function over pairs of the new input variables, and $\mathbf{I_{n^*}}$ is the $n^* \times n^*$ identity matrix.

The advantage of GPR lies in the fact that it is a nonparametric method of regression, thereby doing away with the need to choose basis functions suitable for the model. Moreover, being probabilistic in nature, GPR directly gives the complete predictive distribution as part of the model development. This predictive distribution enables the easy quantification of uncertainty in the predicted variable.

V. Regression Methodology

In this section, the regression methodology used for TOW model building is explained. The FDR dataset for each aircraft type is divided into three sets, namely, the training, the validation, and the test sets. 65% of the flights are randomly chosen to constitute the training set which is used for model building, 15% of the flights are randomly chosen to constitute the validation set which is used for selection from a group of candidate models, and the remaining 20% flights constitute the test set which is used for testing the predictive performance of the selected model. Each observation (data point) in the training, validation and test sets corresponds to the takeoff of one flight. All the variables chosen for regression in Sec. III are standardized, that is, they are shifted by the sample mean and then scaled by the sample standard deviation of the respective variables in the training datasets. Although different regression techniques including ordinary least squares regression [28], Classification and Regression Trees (CART) [29], Least Squares Boosting (LSB) using regression trees [30], and GPR were investigated, the last of these performed the best, and is therefore, presented in this paper.

The GPR starts with hyperparameter inference for the different noisy kernel functions (described in Sec. IV) using the training data. The hyperparameters, being all positive, are given a broad gamma prior with mode 1 and variance 100 (for lack of specific prior knowledge). The MATLAB® [31] based 'GPstuff Toolbox' [32] is used for GPR in this study. Once the models are trained and the hyperparameters are inferred, they are used to determine the model predictive distribution of the TOW for inputs in unseen data not used for training.

Under GPR, the predictive distribution is a normal/Gaussian distribution.

VI. Model performance

The models are evaluated for their performance in predicting (estimating) the TOW of flights in an independent dataset not used for training. The mean TOW estimates and the 95% confidence (prediction) intervals are calculated using the regression models developed in Sec. V. The 95% confidence intervals are given by the 95% highest density intervals [33] [1] of the predictive distributions for the TOW. The metrics used to evaluate the models are as follows:

- **Mean Absolute Relative Prediction Error (MARPE) or Mean Error (ME):** This is the mean of the absolute value of the relative prediction error on independent prediction data (validation or test data).

$$\mathrm{ME} = \frac{1}{n^*}\sum_{i=1}^{n^*}\left|\frac{m_i - \hat{m}_i}{m_i}\right| \tag{20}$$

  Here, n^* is the number of observations in the prediction dataset, m_i is the actual TOW of flight i in the prediction dataset, and $\hat{m}_i$ is the mean estimate of the TOW of flight i from the model. The ME indicates the L_1-norm accuracy of the mean prediction.
- **Normalized Root Mean Squared Prediction Error (NRMSPE) or the Root Mean Squared Error (RMSE):** The RMSE indicates the L_2-norm accuracy of the mean prediction.

$$\mathrm{RMSE} = \frac{\sqrt{\frac{1}{n^*}\sum_{i=1}^{n^*}(m_i - \hat{m}_i)^2}}{\mathrm{sd}(\hat{\mathbf{m}})} \tag{21}$$

  Here, $\mathrm{sd}(\hat{\mathbf{m}})$ is the standard deviation of the vector of the mean predicted TOWs in the prediction dataset.
- **Prediction Coverage (PC):** This is the percentage of the observations in the prediction set for which the actual values of the TOW lie within the 95% confidence intervals given by the model. The PC reflects the accuracy of the predicted uncertainty estimates.
- **Predictive Log Likelihood (PLL):** For a probabilistic model giving the complete predictive distribution, the PLL calculates the log of the likelihood of the actual TOW occurring under the predictive distribution given by the model. PLL indicates the accuracy of the overall predictive distribution given by the model. The approximate PLL for the GPR models (where the predictive distribution is a Gaussian) is calculated as follows:

$$\mathrm{PLL} = \sum_{i=1}^{n^*}\left[-\frac{(m_i - \hat{m}_i)^2}{2\sigma_i^2} - \frac{1}{2}\log\sigma_i^2 - \frac{1}{2}\log(2\pi)\right] \tag{22}$$

  Here, σ_i is the standard deviation of the predictive distribution of the TOW for flight i given by the model.

[1] A 95% highest density interval is an interval in the domain of a probability distribution such that, (i) the probability mass within the interval is 0.95, and (ii) every point inside the interval has a probability density not less than every point outside it. The interval is unique and the shortest among all possible 95% confidence intervals for that distribution.

Models having low ME, low RMSE, high PC, and high PLL are desired. These metrics are used for model selection (using the validation dataset), as well as for evaluating the selected model (using the test dataset). Different models are compared in terms of these metrics using statistical multi-comparison techniques. Among the various models developed in Sec. V using different regression techniques, the GPR model with the Dot Product Squared Exponential (DPSE) kernel is found to statistically (at the 5% significance level) give the overall best predictive performance on the validation datasets for the different aircraft types. Hence, it is selected as the final model for estimating the operational TOW.

A. Model Evaluation

Table II shows the performance of the GPR model with DPSE kernel in estimating (or predicting) the TOW on the test datasets for the different aircraft types. The table also shows the performance of the TOW estimation model given by the Aircraft Noise and Performance (ANP) database. A part of FAA's AEDT aircraft performance modeling tool, the ANP database models the TOW as a piecewise constant function of the flight stage/trip length [4]. We determine the flight stage/trip length by calculating the great circle distance between the flight origin and destination airports. Due to the deterministic nature of the ANP model, the PC of the ANP model is zero. The PLL of the ANP model is $-\infty$, and while the GPR model yields significantly higher values of PLL, we do not present them in this paper due to limited space.

The metrics shown for the individual model performances are calculated across all flights in the test dataset of a particular aircraft type. Table II also shows the comparison of the predictive performance of the GPR and the ANP models on test data. The comparative study is undertaken by using a one-sided Wilcoxon signed rank test [34]. This test is appropriate because the performance of the GPR and the ANP models are to be compared on the same set of flights in the test data (matched data). The null hypothesis is that the GPR model gives a similar (or worse) predictive performance as compared to the ANP model in terms of a similar (or higher) ME, and a similar (or lower) PC. The alternate hypothesis is that the GPR model gives a better predictive performance than the ANP model in terms of a lower ME (i.e., $\mathrm{ME}_{\mathrm{GPR}} < \mathrm{ME}_{\mathrm{ANP}}$) and a higher PC (i.e., $\mathrm{PC}_{\mathrm{GPR}} > \mathrm{PC}_{\mathrm{ANP}}$). The ME and PC used for this statistical comparison study are calculated on a per-flight basis (and not across flights) for a particular aircraft type. Table II shows the p-values [2] obtained through the Wilcoxon test.

From Tab. II, it can be seen that the proposed GPR models give a mean error of about 6% or less for all the aircraft types. The average of the ME values across the different aircraft types is 3% for the GPR models, as compared to 5.6% for the corresponding ANP models. The prediction coverage given by the GPR models is 95% or more for all the

[2] The p-value is the probability of observing a value as, or more, extreme than the calculated test statistic under the null hypothesis.

TABLE II. TOW ESTIMATION: PERFORMANCE METRICS OF THE GPR WITH DPSE KERNEL REGRESSION MODEL, AND THE ANP MODEL, ON THE TEST DATASETS FOR DIFFERENT AIRCRAFT TYPES. ALL THE EVALUATION METRICS ARE CALCULATED ON DESTANDARDIZED DATA (THAT IS, DATA AT THEIR ORIGINAL LOCATION AND SCALE, AND NOT SHIFTED BY THE MEAN AND SCALED BY THE STANDARD DEVIATION OF THE TRAINING DATASET). THE P-VALUES INDICATING ACCEPTANCE OF THE ALTERNATE HYPOTHESIS AT THE 5% SIGNIFICANCE LEVEL ARE HIGHLIGHTED IN BOLD.

Aircraft Type	No. of Points		GPR with DPSE Kernel			ANP			Model Comparison: p-values	
	Training	*Test*	*ME (%)*	*RMSE*	*PC (%)*	*ME (%)*	*RMSE*	*PC (%)*	$ME_{GPR} < ME_{ANP}$	$PC_{GPR} > PC_{ANP}$
A319-112	85	26	4.57	1.99	96	5.47	7.88	0	**0.041**	**4.4e-6**
A320-214	110	34	3.58	1.39	100	4.72	2.80	0	**0.019**	**1.8e-7**
A321-111	76	23	6.15	1.80	96	6.73	1.9e14	0	0.303	**1.4e-5**
A330-202	55	17	2.22	0.43	100	5.97	2.33	0	**0.004**	**1.5e-4**
A330-243	65	20	1.85	0.32	95	3.60	0.89	0	**0.017**	**4.8e-5**
A340-541	34	10	1.67	0.31	100	4.61	0.82	0	**0.011**	**0.003**
B767-300	59	18	1.93	0.34	100	8.34	2.72	0	**1.2e-4**	**9.8e-5**
B777-3FX(ER)	85	26	1.99	0.39	96	5.48	0.96	0	**0.001**	**4.4e-6**

aircraft types. As indicated by the p-values, our GPR model also performs statistically significantly (at the 5% significance level) better than the ANP model, in terms of both the mean error (indicative of the accuracy of the point estimate) and the prediction coverage (indicative of the accuracy of the uncertainty estimate) for nearly all the aircraft types in the study. The only exception is the mean error for the A321, where the models are similar in performance. Fig. 2 presents a graphical depiction of the better predictive performance of our GPR model over the ANP model for the B777 on its test dataset.

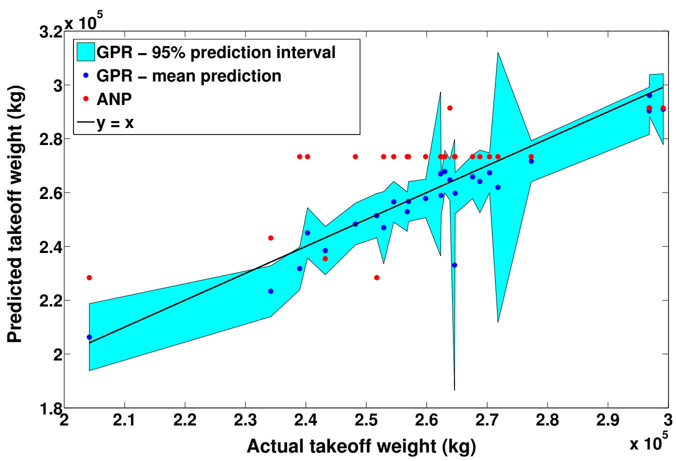

Figure 2. TOW estimation for the B777: Performance of the GPR and the ANP models on the test dataset. The blue dots are the GPR mean predictions and the red dots are the ANP model predictions. The light blue region represents the 95% confidence (prediction) intervals given by the GPR model. The black line is the $y = x$ line on which the mean predictions would lie for a perfect prediction of the mean TOW.

VII. Application of the Takeoff Weight estimation model to estimate the fuel flow rate

As mentioned in Sec. I, the TOW is an essential parameter for estimating the fuel consumption of an aircraft. In this section, we show how our GPR models for TOW estimation can be used for estimating the average fuel flow rate (that is, the mass of fuel consumed per unit time) per engine. The analysis is demonstrated for the A321-111 aircraft in the ascent phase of flight (the phase just after wheels-off to top of climb), but can be easily extended to other phases of flight and other aircraft types.

A. Fuel Flow Rate Modeling

In previous research, we showed that the average fuel flow rate per engine in ascent can be statistically estimated by considering the aircraft dynamic pressure multiplied by the wing reference area, the aircraft mass, the ratio of the vertical speed to the ground speed, the ground speed, and the rate of change of the ground speed with time as predictor variables [35]. Using these variables (all in SI units), we develop a GPR model for the fuel flow rate of the A321-111 in ascent, and use the aircraft TOW in place of the instantaneous aircraft mass as a predictor variable. Each point in the training, validation, and test sets represents one FDR observation during ascent in a particular flight. Each flight contributes to multiple observations during ascent. All the observations in the ascent phase of a particular flight belong to either the training, or the validation, or the test sets. The total number of points in the training, validation, and test sets is 18,261, 4,241, and 6,171, respectively. The flights in each of the training, validation, and test sets used in fuel flow rate modeling are the same as those in the sets used for TOW modeling in Sec. V. The GPR modeling methodology is the same as that explained in Secs. IV and V. A GPR model with a DPSE kernel is built using the training dataset.

The true values of the variables from the FDR data are used to train the models. However, as before, we desire models that can be used to estimate the fuel flow rate of operations even in the absence of FDR data. Therefore, only variables derivable through more accessible data (such as, ground-based track data) are used as model features, and as inputs while evaluating model performance. Subsequently, the air density is assumed to be a function of the aircraft altitude according to the International Standard Atmosphere (ISA) model [36]. The TOW, which is a predictor variable for the fuel flow rate models, also needs to be estimated as its actual value for a particular flight is not available. The accuracy of the fuel flow rate estimation therefore, depends on the accuracy of the TOW estimation. Fig. 3 shows the sensitivity of the fuel flow rate to the TOW. The figure plots the Mean Error (ME) in the fuel flow rate on the test dataset from the GPR model developed for the A321-111 in ascent, as a function of the deviation of the estimated TOW from the actual TOW. A positive deviation means that the estimated TOW is greater than the actual TOW, while a negative deviation implies that the estimated TOW is less than the actual TOW [30]. Fig. 3 shows that as expected, the ME is the least when the estimated TOW is close to the actual TOW, and increases with the magnitude of the deviation in TOW. The ME increases from 2.89% to 3.11% (a percentage increase of 7.6%) for just a 2% deviation in the estimated TOW from the actual value. This sensitivity of the ME in fuel flow rate to the deviation in the estimated TOW motivates the need to accurately estimate the TOW, in order to accurately estimate the fuel flow rate.

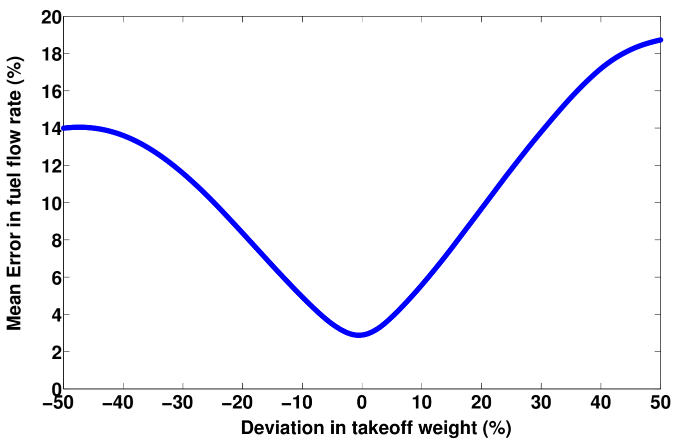

Figure 3. A321-111 in ascent: Sensitivity of the Mean Error (ME) in fuel flow rate estimation to the deviation in the estimated TOW, for the test dataset.

Using the TOW estimated by the ANP model as well as by our GPR model (Sec. V) as inputs to a fuel flow rate GPR model, their predictive performances on the flights in the unseen test dataset are now evaluated. To incorporate uncertainty in the estimated TOW, this predictive performance is evaluated using the predictive distribution of the fuel flow rate in ascent marginalized over the estimated values of TOW. In other words, we are interested in computing the following:

$$p(\dot{m}_f|\mathbf{x_{-m}},\phi,\mathscr{D}_\mathbf{1},\mathscr{D}_\mathbf{2}) = \int_m p(\dot{m}_f|\mathbf{x_{-m}},m,\mathscr{D}_\mathbf{1})p(m|\phi,\mathscr{D}_\mathbf{2})dm \tag{23}$$

Here, p refers to the Probability Distribution Function (PDF), $\dot{m}_f$ is the fuel flow rate to be predicted, $\mathbf{x_{-m}}$ is the vector of predictor variables in the fuel flow rate GPR model excluding the TOW (that is, the aircraft dynamic pressure multiplied by the wing reference area, the ratio of the vertical speed to the ground speed, the ground speed, and the rate of change of the ground speed with time during ascent), m is the TOW, and $\mathscr{D}_\mathbf{1}$ is the set of the training variables used for building the fuel flow rate GPR model. $p(\dot{m}_f|\mathbf{x_{-m}},m,\mathscr{D}_\mathbf{1})$ is the PDF of the predictive distribution given by the fuel flow rate GPR model and is thus, a Gaussian PDF. $p(m|\phi,\mathscr{D}_\mathbf{2})$ is the distribution of the estimated TOW parametrized by ϕ and $\mathscr{D}_\mathbf{2}$.

The ANP model is a deterministic model giving a flight stage length-based point estimate of the TOW, m_{ANP}. Under the ANP model, (23) becomes

$$p(\dot{m}_f|\mathbf{x_{-m}},\phi,\mathscr{D}_\mathbf{1},\mathscr{D}_\mathbf{2}) = p(\dot{m}_f|\mathbf{x_{-m}},m_{\text{ANP}},\mathscr{D}_\mathbf{1}) \tag{24}$$

which is the PDF of a normal distribution under the GPR formulation.

Our GPR models for TOW (Sec. V) give the complete predictive distribution for the TOW (which is a normal distribution). Therefore, under our GPR models for TOW, (23) becomes

$$p(\dot{m}_f|\mathbf{x_{-m}},\phi,\mathscr{D}_\mathbf{1},\mathscr{D}_\mathbf{2}) = \int_m p(\dot{m}_f|\mathbf{x_{-m}},m,\mathscr{D}_\mathbf{1})p(m|\phi,\mathscr{D}_\mathbf{2})dm \tag{25}$$

$$\approx \frac{1}{n_s}\sum_{i=1}^{n_s} p(\dot{m}_f|\mathbf{x_{-m}},m_i,\mathscr{D}_\mathbf{1}). \tag{26}$$

When our GPR models are used to estimate the TOW, ϕ and $\mathscr{D}_\mathbf{2}$ hold specific meanings. ϕ is the vector of predictor variables used in the GPR models to estimate the TOW (i.e., V_1, V_2, ρ, and S as mentioned in Sec. III). $\mathscr{D}_\mathbf{2}$ is the training set used to build the GPR models to estimate the TOW. Equation (26) approximates (25) through a Monte Carlo approximation with n_s samples of the TOW drawn from its Gaussian predictive distribution given by the GPR models for TOW estimation. In this study, we choose n_s to be 1000. Equation (26) therefore, shows that the desired predictive distribution of the fuel flow rate under a GPR model of the TOW is approximately a Gaussian Mixture distribution with n_s equally weighted components.

A question that naturally arises is how well a GPR model for the fuel flow rate in ascent that is built without using the TOW explicitly as a predictor variable at all, would perform in terms of prediction. To answer this question, a GPR model with the DPSE kernel function is trained using the aircraft dynamic pressure multiplied by the wing reference area during ascent, the ratio of the vertical speed to the ground speed during ascent, the ground speed during ascent, the rate of change of the ground speed with time during ascent, and the aircraft ground speed at the start of the takeoff ground roll, the aircraft ground speed at wheels-off, the ambient air density during takeoff, the takeoff ground roll distance as predictor variables and the averaged fuel flow rate per engine as the predicted variable (all in SI units). This set of predictor variables is a combination of the predictor variables used to model the fuel flow rate in Sec. VII-A (excluding the TOW), and the predictor variables that were used to build the TOW models in Sec. V. The predictive distribution for the fuel flow rate under this model, $p(\dot{m}_f|\psi,\mathscr{D}_\mathbf{3})$, is also a Gaussian/normal distribution. Here, ψ is the vector of predictor/input variables used in the model, and $\mathscr{D}_\mathbf{3}$ is the training set used to build the model. In other words:

$$p(\dot{m}_f|\mathbf{x_{-m}},\phi,\mathscr{D}_\mathbf{1},\mathscr{D}_\mathbf{2}) = p(\dot{m}_f|\psi,\mathscr{D}_\mathbf{3}) \tag{27}$$

B. Evaluation of Fuel Flow Rate Model

Depending on how the TOW variable is estimated, there are three variants of the GPR models developed in Sec. VII-A to estimate the fuel flow rate:

1) **Model 1:** This variant estimates the TOW predictor variable (an input to the GPR model for fuel flow rate) using our GPR models for TOW;
2) **Model 2:** This variant estimates the TOW predictor variable that is input into the GPR model for fuel flow rate using the ANP model; and
3) **Model 3:** This variant employs a GPR model that does not use the TOW as an explicit predictor variable at all.

The predictive distributions for the fuel flow rate marginalized over the TOW for Model 1 (26), Model 2 (24) and Model 3 (27) are used to calculate the mean predictions and the 95% highest density prediction intervals for the fuel flow rates. Table III tabulates the predictive performance of these models on the unseen test data in ascent for the A321-111, using the evaluation metrics described in Sec. VI. Each evaluation metric is calculated for each flight in the test dataset using all the points in the ascent phase of that flight. In Table III, the columns under 'Individual Model Performance' show the mean and the standard deviation (within parentheses) of the per-flight evaluation metrics, averaged across all the flights in the test dataset.

Table III also shows the results of a study to compare the predictive performance of Model 1 with that of Model 2 and Model 3. The comparison is done using a one-sided Wilcoxon signed rank test on the flights in the test dataset. The null hypotheses are that Model 1 gives a similar (or worse) predictive performance as compared to Model 2 or Model 3. The alternate hypotheses are that Model 1 gives a better predictive performance than Model 2 (Model 1 $>$ Model 2), or Model 3 (Model 1 $>$ Model 3), in terms of a lower ME, lower RMSE, higher PC, or higher PLL. The ME, RMSE, PC, and PLL given by Model 1 are compared with those given by Model 2 or Model 3, on the same set of flights in the test data (matched data). Table III shows the p-values obtained through the Wilcoxon test. Since two statistical tests are being simultaneously performed (comparison of Model 1 with Model 2 and with Model 3), a Bonferroni correction is applied to get a family significance level of atmost 5%, giving a per-test significance level of 2.5%.

Table III shows that the p-values for ME and RMSE are greater than 0.025 for a one-tailed Wilcoxon test. A two-tailed Wilcoxon test is then conducted to test if the ME and RMSE given by Model 1 are statistically significantly different from those given by Model 2 or Model 3. The p-values of the two-tailed test (not shown here) are also found to be greater than 0.025. Therefore, in terms of the accuracy of the point estimates of the fuel flow rates, Model 1 is statistically similar in its predictive performance to Model 2 and Model 3 at a family significance level of atmost 5%. However, Model 1 is statistically significantly better than Model 2 and Model 3 in terms of the accuracy of estimating the uncertainty of the fuel flow rates (as the p-values for PC using a one-tailed Wilcoxon test are less than 0.025). Model 1 is also statistically significantly better than Model 2 in terms of the accuracy of estimating the overall predictive distribution of the fuel flow rates (as the p-value for PLL is less than 0.025). Thus, a statistical model of TOW estimation (Model 1) which takes the uncertainty in TOW into account is able to model the fuel flow rate more accurately than the deterministic ANP model of TOW estimation (Model 2) which does not consider the uncertainty in TOW at all. The superior predictive performance of Model 1 over Model 3 also shows that a model explicitly incorporating the TOW as a predictor variable is able to estimate the fuel flow rate more accurately than one not using the TOW explicitly, even though the model might indirectly capture the effect of the TOW through the inclusion of other variables which govern the TOW. Using the GPR models to estimate the TOW gives an averaged mean error in the fuel flow rate of about 4.4%, and an averaged prediction coverage of more than 90% on the A321-111 test data in ascent.

VIII. Summary and Conclusions

This paper presented a statistical approach to model aircraft Takeoff Weight (TOW), given trajectory variables from the takeoff roll and other data that are often available to analysts. Gaussian Process Regression, a nonparametric probabilistic method, was selected to build the regression models. By virtue of being nonparametric, GPR does not need the assumption of basis functions of the input/predictor variables (unlike methods like least squares regression, where one has to assume either linear, quadratic, or some other form of basis functions prior to carrying out the regression). Being a probabilistic method, GPR can provide the complete predictive distribution of the TOW rather than just a point estimate. The uncertainty estimates given by the predictive distribution quantify the cumulative effect of unmodeled factors in TOW modeling as well as random disturbances in aircraft operation. The model variables or features were selected to reflect a physical understanding of aircraft dynamics during ground roll, as well as their ease of availability. The result of our research is, for the first time, validated models that can provide a probabilistic estimate of the TOW, given trajectory data from the takeoff ground roll.

In contrast to prior methods proposed for TOW estimation, we were able to validate and evaluate our models using an independent test set of FDR data. Metrics were developed to quantify the accuracy of both the point (i.e., mean) estimates as well as the uncertainty estimates of the TOW. Our GPR models gave a mean error of 3% (on average across all the aircraft types), and a prediction coverage of more than 95% on the test data for all the aircraft types studied. The model performance was also compared with that of the Aircraft Noise and Performance (ANP) model; the proposed GPR models were shown to predict the TOW statistically significantly (at the 5% significance level) more accurately than the ANP model.

Finally, an application of the TOW models to estimate aircraft engine fuel flow rates was also described. Fuel flow rate is highly dependent on the weight of the aircraft; therefore the TOW estimate is a valuable predictor/input variable of the

TABLE III. FUEL FLOW RATE ESTIMATION FOR THE A321-111 IN ASCENT: COMPARATIVE PERFORMANCE OF THREE DIFFERENT MODELS. EACH ENTRY UNDER 'INDIVIDUAL MODEL PERFORMANCE' SHOWS THE MEAN AND THE STANDARD DEVIATION (WITHIN PARENTHESES) OF THE PER-FLIGHT EVALUATION METRIC, AVERAGED ACROSS ALL THE FLIGHTS IN THE TEST DATA. ALL THE EVALUATION METRICS ARE CALCULATED ON DESTANDARDIZED DATA (THAT IS, DATA AT THEIR ORIGINAL LOCATION AND SCALE AND NOT SHIFTED BY THE MEAN AND SCALED BY THE STANDARD DEVIATION OF THE TRAINING DATASET). P-VALUES INDICATING ACCEPTANCE OF THE ALTERNATE HYPOTHESIS AT THE 5% FAMILY SIGNIFICANCE LEVEL ARE HIGHLIGHTED IN BOLD.

	Individual Model Performance			**Model Comparison: p-values**	
	Model 1	***Model 2***	***Model 3***	***Model 1 > Model 2***	***Model 1 > Model 3***
ME (%)	4.40 (2.59)	5.16 (3.66)	4.04 (1.98)	0.092	0.644
RMSE	0.32 (0.15)	0.36 (0.21)	0.30 (0.10)	0.116	0.722
PC (%)	93.78 (10.04)	80.15 (21.09)	90.99 (8.78)	**9.8e-5**	**0.007**
PLL	427.73 (166.68)	180.25 (510.70)	388.36 (163.31)	**2.7e-4**	0.024

fuel flow rate estimate. We therefore, first estimated the TOW, and further used it to estimate the fuel flow rate. As a result, the accuracy of fuel rate estimate depended on the accuracy of the TOW model. We investigated the impact of using two different TOW models (our proposed GPR model and the ANP model) as input to a GPR model for the fuel flow rate. In addition, we developed a third GPR model of fuel flow rate that did not require the TOW as an explicit predictor variable. We showed that the fuel flow rate model that used the GPR estimate of TOW (i.e., Model 1) performed similarly to the other two models (Model 2: ANP+GPR and Model 3: TOW-less GPR) in determining point estimates of the fuel flow rate on the A321-111 test dataset studied. The averaged mean error in predicting the fuel flow rate in ascent was shown to be approximately 4.4%. However, we found that Model 1 gave a statistically significantly better estimate of the uncertainty (as reflected by the prediction coverage) in the fuel flow rate on the A321-111 test dataset studied, as compared to the other two methods. In short, a statistical model that accounts for the uncertainty in TOW and propagates this uncertainty properly to the fuel flow rate, is also able to quantify the uncertainty in the fuel flow rate more accurately than a deterministic model (like the ANP). Moreover, a fuel flow rate model that explicitly includes the TOW as a predictor variable performs better than a model that excludes the TOW.

It is worth noting that the amount of flight data used for TOW model building was quite limited. The same methodology can be applied to more flight data from varied operations to increase the accuracy of the TOW models. The model accuracy can be further increased by including more variables – for example, the takeoff thrust deration level and the coefficient of friction during the takeoff ground roll, if available – as predictors. Despite these limitations, this paper has highlighted the potential of modern statistical methods to estimate the TOW, and proposed a class of more accurate, validated models that are capable of estimating the TOW of an aircraft from its takeoff ground roll trajectory. Future work will include the demonstration of our statistical models on ASDE-X data corresponding to the takeoff roll, in order to estimate the TOW of the flight being tracked.

References

[1] E. Tulapurkara, "National Programme on Technology Enhanced Learning (NPTEL) India: Lectures on airplane design (aerodynamic): Weight estimation," 2013, [Online; accessed 22 January 2017].

[2] D. P. Raymer, *Aircraft Design: A Conceptual Approach*, 5th ed. Reston, Virginia: American Institute of Aeronautics and Astronautics (AIAA), 2012.

[3] J. Roskam, *Airplane Design Part V: Component Weight Estimation*. Ottawa, Kansas: Roskam Aviation and Engineering Corp., 1985.

[4] M. Ahearn et al., *Aviation Environmental Design Tool (AEDT) Technical Manual, Version 2b, Service Pack 3*, U.S. Department of Transportation Volpe National Transportation Systems Center, 2016, Report No. DOT-VNTSC-FAA-16-11.

[5] L. Sherry and S. Neyshabouri, "Estimating takeoff thrust from surveillance track data," in *Transportation Research Board Annual Meeting*, January 2014.

[6] U.S. Department of Transportation, "Air Carrier Statistics Database," 2016.

[7] R. Alligier, D. Gianazza, and N. Durand, "Energy rate prediction using an equivalent thrust setting profile," in *International Conference on Research in Air Transportation*, May 2012.

[8] ——, "Ground-based estimation of the aircraft mass, adaptive vs. least squares method," in *USA/Europe Air Traffic Management Research and Development Seminar (ATM 2013)*, June 2013.

[9] ——, "Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 45–60, 2013.

[10] R. Alligier, D. Gianazza, M. G. Hamed, and N. Durand, "Comparison of two ground-based mass estimation methods on real data," in *International Conference on Research in Air Transportation*, May 2014.

[11] R. Alligier, D. Gianazza, and N. Durand, "Machine learning and mass estimation methods for ground-based aircraft climb prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3138–3149, December 2015.

[12] A. Nuic, *User Manual for the Base of Aircraft Data (BADA) Revision 3.13*, EUROCONTROL Experimental Centre, 2015, EEC Technical/Scientific Report No. 15/04/02-43.

[13] J. Sun, J. Ellerbroek, and J. Hoekstra, "Modeling and

inferring aircraft takeoff mass from runway ADS-B data," in *International Conference on Research in Air Transportation*, June 2016.

[14] G. G. Kulikov and H. A. Thompson, Eds., *Dynamic Modelling of Gas Turbines: Identification, Simulation, Condition Monitoring, and Optimal Control*. London: Springer-Verlag, 2004.

[15] EUROCONTROL Experimental Centre, "Aircraft Noise and Performance (ANP) Database," [Online; accessed 22 January 2017].

[16] Y. S. Chati and H. Balakrishnan, "Analysis of aircraft fuel burn and emissions in the landing and take off cycle using operational data," in *International Conference on Research in Air Transportation*, May 2014.

[17] R. Dürichen, M. A. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton, "Multi-task Gaussian process models for biomedical applications," in *IEEE-EMBS International Conference on Biomedical and Health Informatics*, June 2014.

[18] L. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson, and L. Tarassenko, "Gaussian process regression in vital-sign early warning systems," in *Annual International Conference of the IEEE EMBS*, August 2012.

[19] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "Probabilistic estimation of respiratory rate using Gaussian processes," in *Annual International Conference of the IEEE EMBS*, July 2013.

[20] C. Hultquist, G. Chen, and K. Zhao, "A comparison of Gaussian process regression, random forests, and support vector regression for burn severity assessment in diseased forests," *Remote Sensing Letters*, vol. 5, no. 8, pp. 723–732, 2014.

[21] L. Pasolli, F. Melgani, and E. Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3, pp. 464–468, July 2010.

[22] J. C. Platt, C. J. C. Burges, S. Swenson, C. Weare, and A. Zheng, "Learning a Gaussian process prior for automatically generating music playlists," in *Advances in Neural Information Processing Systems 14*, 2002, pp. 1425–1432.

[23] K. M. A. Chai, C. K. I. Williams, S. Klanke, and S. Vijayakumar, "Multi-task Gaussian process learning of robot inverse dynamics," in *Advances in Neural Information Processing Systems 21*, 2009, pp. 265–272.

[24] A. Schwaighofer, M. Grigoraş, V. Tresp, and C. Hoffmann, "GPPS: A Gaussian process positioning system for cellular networks," in *Advances in Neural Information Processing Systems 16*, 2004, pp. 579–586.

[25] C. A. L. Bailer-Jones, T. J. Sabin, D. J. C. MacKay, and P. J. Withers, "Prediction of deformed and annealed microstructures using Bayesian neural networks and Gaussian processes," in *Proceedings of the Australasia-Pacific Forum on Intelligent Processing and Manufacturing of Materials*, 1997.

[26] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[27] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari, "Bayesian modeling with Gaussian processes using the GPstuff toolbox," in *arXiv:1206.5754 [stat.ML]*, 2015.

[28] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. New Delhi: McGraw Hill Education (India) Private Limited, 2013.

[29] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York: CRC Press, 1993.

[30] Y. S. Chati and H. Balakrishnan, "Statistical modeling of aircraft fuel flow rate," in *Congress of the International Council of the Aeronautical Sciences*, September 2016.

[31] "The Mathworks, Inc., MATLAB® R2012a documentation," 2012, software package.

[32] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari, "GPstuff: Bayesian modeling with Gaussian processes," *Journal of Machine Learning Research*, vol. 14, pp. 1175–1179, 2013.

[33] J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed. San Diego, California: Academic Press, 2015.

[34] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, December 1945.

[35] Y. S. Chati and H. Balakrishnan, "A Gaussian process regression approach to model aircraft engine fuel flow rate," in *ACM/IEEE International Conference on Cyber-Physical Systems*, April 2017.

[36] J. D. Anderson, Jr., *Introduction to Flight*, 5th ed. McGraw-Hill Higher Education, 2004.

Author Biographies

Yashovardhan S. Chati is a doctoral candidate in the Department of Aeronautics and Astronautics, MIT. He applies statistical methods and machine learning algorithms to develop data-driven models of aircraft engine performance. He obtained his Bachelor of Technology and Master of Technology degrees from IIT Bombay. He has received the Institute Gold Medal from IIT Bombay in 2012 for being the most outstanding graduating student and a Graduate Research Award from the Transportation Research Board's Airport Cooperative Research Program in 2016. Website: web.mit.edu/yschati/www

Hamsa Balakrishnan is an Associate Professor of Aeronautics and Astronautics at MIT. Her research is in the design, analysis, and implementation of control and optimization algorithms for large-scale cyber-physical infrastructures, with an emphasis on air transportation systems. She is a recipient of the NSF CAREER Award in 2008, the Kevin Corker Award for Best Paper of ATM-2011, the inaugural CNA Award for Operational Analysis in 2012, the AIAA Lawrence Sperry Award in 2012, and the American Automatic Control Council's Donald P. Eckman Award in 2014. Website: web.mit.edu/hamsa/www