Modeling of Aircraft Takeoff Weight Using Gaussian Processes

Yashovardhan S. Chati* and Hamsa Balakrishnan†
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

DOI: 10.2514/1.D0099

The takeoff weight of an aircraft is an important aspect of aircraft performance. However, the takeoff weight of a particular flight is generally not available to entities outside of the operating airline. The preceding observations motivate the development of accurate takeoff weight estimates that can be used for fuel-burn estimation or trajectory prediction. This paper proposes a statistical approach based on Gaussian process regression to determine both a mean estimate of the takeoff weight and the associated prediction interval, using observed data from the takeoff ground roll. The model development and validation are conducted using flight data recorder archives, which also provide ground-truth data. The models are found to have a mean absolute error in takeoff weight of 3.6%, averaged across nine different aircraft types, resulting in a nearly 35% smaller error than the models in the Aircraft Noise and Performance database. Finally, the developed models are used to predict aircraft fuel flow rate during climb out and approach. For the majority of the aircraft types studied, the statistical models of takeoff weight estimation are shown to result in a similar or better fuel flow rate predictive performance as compared to the Aircraft Noise and Performance models.

Nomenclature

a	=	acceleration, m · s^{-2}
C_D	=	coefficient of drag
C_L	=	coefficient of lift
D	=	drag, N
F_0	=	static engine thrust, N
F_{00}	=	maximum sea-level, static engine thrust, N
F_n	=	aircraft net thrust, N
f	=	generic function
f_r	=	frictional force, N
g	=	acceleration due to gravity, 9.81 m · s^{-2}
$\boldsymbol{I}_n$	=	$n \times n$ identity matrix
$\boldsymbol{K}$	=	covariance matrix
$k(\boldsymbol{x}, \boldsymbol{x}')$	=	covariance/kernel function
L	=	lift, N
ℓ	=	length scale
$\dot{m}_f$	=	averaged fuel flow rate per engine, kg/s
m_{TO}	=	takeoff weight, kg
N	=	normal reaction, N
N_{eng}	=	number of engines
$\mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$	=	normal/Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$
n	=	number of observations
p	=	probability distribution function
q	=	dynamic pressure, Pa
R	=	distance covered during takeoff ground roll, m
$\mathcal{S}$	=	wing reference area, m^2
t	=	time, s
V	=	true airspeed, m · s^{-1}
V_{GS}	=	ground speed, m · s^{-1}
$\boldsymbol{X}$	=	matrix of input vectors
x	=	predictor/input/independent variable
y	=	predicted/output/dependent variable
ϵ	=	noise
η	=	thrust deration level
μ_r	=	coefficient of friction
ρ_∞	=	ambient air density, kg · m^{-3}
σ^2	=	variance parameter

Selected as a Best-in-Track paper at the 12th USA/Europe Air Traffic Management R&D Seminar, Seattle, WA, 26–30 June 2017; received 2 October 2017; revision received 1 May 2018; accepted for publication 3 May 2018; published online 21 June 2018.

*Research Engineer, Department of Aeronautics and Astronautics, 77 Massachusetts Avenue. Student Member AIAA.

†Associate Professor, Department of Aeronautics and Astronautics, 77 Massachusetts Avenue. Associate Fellow AIAA.

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, TOW is not generally available outside the operating carrier, due to its dependence on proprietary information such as load factors and operational strategies. The preceding facts motivate the development of models to estimate the TOW of a flight from accessible information. As per the convention in aviation, the term weight in this paper refers to the physical mass of the aircraft.

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 [2–4]. 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 U.S. Department of Transportation [6]. However, this method cannot be easily extended to estimate the TOW of a particular flight because 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 [7–10]. 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. The equivalent TOW is computed using either an adaptive mechanism or least-squares algorithms. Machine learning techniques have also been applied to radar data to estimate the TOW to predict the future aircraft trajectory [11]. The methods proposed in these studies have been shown to be superior to the Eurocontrol's Base of Aircraft Data method for trajectory modeling [12]. However, because of the unavailability of ground-truth data, the accuracies of these TOW estimates have not been evaluated. Instead, these models have been evaluated based on the trajectory prediction accuracy.

Recent work has used runway automatic dependent surveillance–broadcast (ADS-B) data during takeoff to model the operational TOW, using least-squares methods [13]. However, these studies assumed no deration in the takeoff thrust and a standard coefficient of friction for the ground roll, and the resultant TOW estimates could not be validated due to the unavailability of ground-truth data. Phase-based models have recently been used for the Bayesian inference of TOW [14].

Table 1 FDR dataset: aircraft types and engines

Aircraft type	Engine type	MTOW, kg	OEW, kg	Number of flights
A319-112	2 × CFMI CFM56-5B6/2 or 2P	64,000	40,160	130
A320-214	2 × CFMI CFM56-5B4/2 or P/2P	73,500	42,100	169
A321-111	2 × CFMI CFM56-5B1/2 or 2P	89,000	48,500	117
A330-202	2 × GE CF6-80E1A4	230,000	120,500	84
A330-243	2 × RR Trent 772B-60	230,000	120,600	100
A330-343	2 × RR Trent 772B-60	230,000	124,600	182
A340-541	4 × RR Trent 553	372,000	170,900	52
B767-300	2 × GE CF6-80C2B7F	156,490	86,955	91
B777-300ER	2 × GE GE90-115B1	345,050	167,825	131

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 factors (such as manufacturing tolerances, turbulence, fluctuations in ambient conditions, component aging, and deterioration) motivate the development of a statistical model [15]. Data from the flight data recorders (FDRs) of real flight operations allow us to build and validate our models. The proposed Gaussian process regression (GPR) based techniques enable the estimation of the mean TOW of a flight as well as the underlying uncertainty distribution, which captures the cumulative effect of unmodeled factors and random effects. Although the models are built using FDR data, they can be used to estimate the operational TOW for flights, given trajectory data from surface surveillance sources. The models are shown to have a mean absolute error of 3.6% (averaged across different aircraft types) in predicting the TOW of flights in an independent test set, a nearly 35% reduction in mean absolute error compared to the aircraft noise and performance (ANP) database [16]. These GPR models are then applied to surface surveillance data from an airport surface detection equipment, model X (ASDE-X) system to demonstrate their practical utility. Finally, an application of using these TOW estimates to estimate the fuel flow rate in climb out and approach is demonstrated.

B. Outline

We start with briefly describing the flight data recorder 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 GPR. The application of GPR to the problem of modeling the TOW is explained in Sec. V. Section VI presents metrics used to evaluate the predictive performance of the models. In Sec. VII, the models are evaluated using an independent/unseen test dataset not used for model training. The model estimates are also compared to those given by the ANP database [16]. The ANP database is used for TOW estimation by the Federal Aviation Administration's aviation environmental design tool (AEDT) [4], a widely used aircraft performance modeling tool. Section VIII shows TOW model predictive performance when surface surveillance data from ASDE-X are used as the source of trajectory information. In Sec. IX, we show how the TOW estimates given by our models can be used to estimate fuel flow rates in climb out and approach. Finally, we present the main conclusions of this study and directions for future research in Sec. X.

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

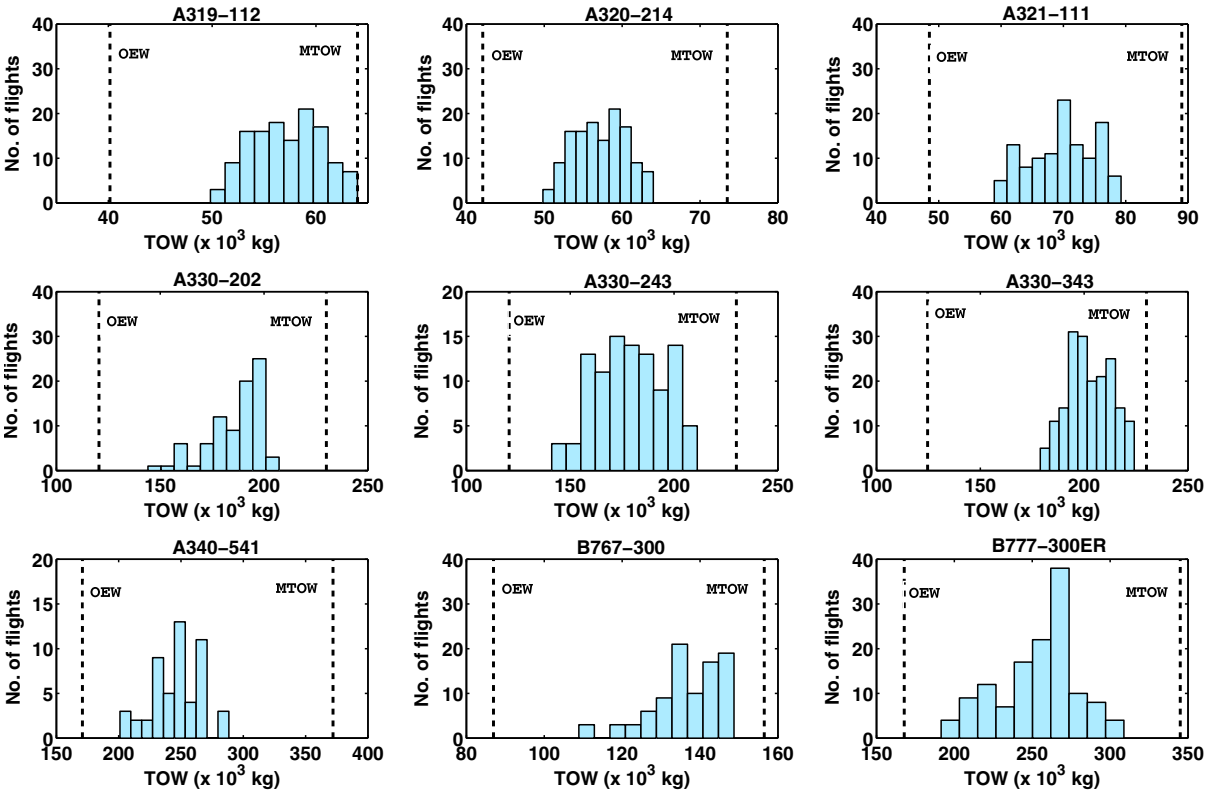

Fig. 1 Histograms of the takeoff weights in FDR data with MTOW and OEW values overlaid.

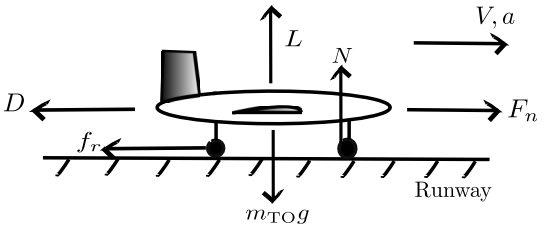

Fig. 2 Airplane dynamics during takeoff ground roll.

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 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.

The different aircraft and engine types included in the study are tabulated in Table 1. The approximate maximum takeoff weight (MTOW), the approximate operating empty weight (OEW), and the number of flights are also shown. The MTOW and the OEW values in the tables are approximate, representative numbers for the aircraft types and have been obtained from Jane's All the World's Aircraft database [17–19]. 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. This paper focuses only on the takeoff ground roll portion of the trajectory [20].

Figure 1 shows histograms of the takeoff weights observed in the FDR dataset for different aircraft types, with the MTOW and OEW values overlaid.

III. Model Features

Figure 2 shows the free-body diagram of an aircraft during takeoff roll on the runway with the forces acting on it. The equations of motion during takeoff ground roll are given in Eqs. (1–7):

$$L + N = m_{\mathrm{TO}} g \tag{1}$$

$$F_n - D - f_r = m_{\mathrm{TO}} a \tag{2}$$

$$L = q\mathcal{S}C_L \tag{3}$$

$$D = q\mathcal{S}C_D \tag{4}$$

$$f_r = \mu_r N \tag{5}$$

$$a = \frac{dV}{dt} \tag{6}$$

$$q = \frac{1}{2}\rho_\infty V^2 \tag{7}$$

Winds are neglected in this analysis. Neglecting wind speeds during takeoff ground roll, the aircraft airspeed is assumed to be equal to the ground speed ($V = V_{\mathrm{GS}}$). 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_{TO}) during the takeoff ground roll. The coefficients of lift and drag, governed by the aircraft configuration, are also assumed to be constant. The net thrust on the aircraft (F_n) 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 that 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 given by

$$F_n = f_{F_n}(N_{\mathrm{eng}}, V_{\mathrm{GS}}, \eta, \rho_\infty, F_{00}) \tag{8}$$

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

$$R = \int_{V_{\mathrm{GS}_1}}^{V_{\mathrm{GS}_2}} V_{\mathrm{GS}} \frac{\mathrm{d}V_{\mathrm{GS}}}{a} \tag{9}$$

Here, V_{GS_1} is the aircraft ground speed at the start of the takeoff ground roll, and V_{GS_2} is the aircraft ground speed at wheels-off at the end of the takeoff ground roll. Combining Eqs. (1–9), the TOW can be expressed by the following functional relation:

$$m_{\mathrm{TO}} = f_{m_{\mathrm{TO}}}(R, \rho_\infty, V_{\mathrm{GS}_1}, V_{\mathrm{GS}_2}, \mathcal{S}, F_{00}, C_L, C_D, \mu_r, \eta, N_{\mathrm{eng}}) \tag{10}$$

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

The modeling variables are now restricted to only those that can be obtained or derived from easily accessible databases. The ground roll distance and aircraft ground speed during ground roll can be derived from surface surveillance data, whereas 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 used as model features. Hence, for a particular aircraft type, the model uses the ground roll distance (R), the ambient air density (ρ_∞) during roll, the aircraft ground speed at the start of the takeoff ground roll (V_{GS_1}), and the aircraft ground speed at the end of the takeoff ground roll (V_{GS_2}) as the predictor/input variables. The predicted/output variable is the aircraft TOW (m_{TO}). In other words, the TOW prediction model has the following form:

$$m_{\mathrm{TO}} \approx f_{m_{\mathrm{TO}}}(R, \rho_\infty, V_{\mathrm{GS}_1}, V_{\mathrm{GS}_2}) \tag{11}$$

The unmodeled features will contribute to the uncertainty in the TOW estimate and will be reflected in the prediction intervals provided by the 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 remote sensing [21,22] and robotics [23].

Because the output variable (TOW) is a continuous (nondiscrete) variable, 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 [24,25].

A regression model is given by

$$y = f(\boldsymbol{x}) + \epsilon \tag{12}$$

where y is the predicted/output/dependent variable; $\boldsymbol{x}$ is the predictor/input/independent vector; $f(\boldsymbol{x})$ 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, the regression function $f(\boldsymbol{x})$ is assumed to follow a Gaussian process

(GP) prior, that is, the function values at any finite set of inputs $\boldsymbol{x}$ follow a joint Gaussian distribution [24]. Then,

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

where $m_e(\boldsymbol{x})$ is the mean function, and $k(\boldsymbol{x}, \boldsymbol{x}')$ is the kernel/covariance function over two inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$, which governs the covariance among function values as $k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{cov}(f(\boldsymbol{x}), f(\boldsymbol{x}'))$. 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, $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$, with mean 0 and noise variance σ_n^2. Assuming 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 $k_{\mathrm{noise}}(\boldsymbol{x}_p, \boldsymbol{x}_q)$ over d-dimensional input vectors $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$:

$$y \sim \mathrm{GP}(0, k_{\mathrm{noise}}(\boldsymbol{x}_p, \boldsymbol{x}_q)) \tag{14}$$

The noisy kernel function for the dependent variables $k_{\mathrm{noise}}(\boldsymbol{x}_p, \boldsymbol{x}_q)$ relates to the kernel function for the regression function values $k(\boldsymbol{x}_p, \boldsymbol{x}_q)$ as follows:

$$k_{\mathrm{noise}}(\boldsymbol{x}_p, \boldsymbol{x}_q) = k(\boldsymbol{x}_p, \boldsymbol{x}_q) + \sigma_n^2 \delta_{pq} \tag{15}$$

where δ denotes the Kronecker delta.

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

1) Dot product squared exponential (DPSE) kernel: This kernel function is used to model very smooth functions (where the smoothness of the GP is defined in terms of its mean square differentiability [24]). It is given by

$$k(\boldsymbol{x}_p, \boldsymbol{x}_q) = \sigma_0^2 + \boldsymbol{x}_p^T \Sigma \boldsymbol{x}_q + \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{\ell_i^2}\right)$$
$$\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2) \tag{16}$$

2) Dot product exponential (DPE) kernel: This kernel function is used to model very rough functions. It is given by

$$k(\boldsymbol{x}_p, \boldsymbol{x}_q) = \sigma_0^2 + \boldsymbol{x}_p^T \Sigma \boldsymbol{x}_q + \sigma_f^2 \exp\left(-\sqrt{\sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{\ell_i^2}}\right)$$
$$\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2) \tag{17}$$

In Eqs. (16) and (17), $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ 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 ith 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 $[\sigma_0^2 \quad \sigma_1^2 \quad \sigma_2^2 \quad , \ldots, \quad \sigma_d^2 \quad \sigma_f^2 \quad \ell]^T$.

Numerous other kernel functions exist, the details of which can be found in [25].

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

$$\hat{\theta} = \arg\max_{\theta} \log p(\theta | \boldsymbol{X}, \boldsymbol{y})$$
$$= \arg\max_{\theta} \left\{\log p(\theta) - \frac{1}{2}\boldsymbol{y}^T \boldsymbol{K}_y^{-1} \boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}_y| - \frac{n}{2}\log(2\pi)\right\} \tag{18}$$

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; $\boldsymbol{X}$ is the $n \times d$ matrix of d-dimensional inputs; $\boldsymbol{y}$ is the $n \times 1$ vector of the dependent variable values; and $\boldsymbol{K}_y$ is the $n \times n$ covariance matrix derived from the noisy kernel function over pairs of input variables [Eq. (15)].

This paper uses GPR to make predictions on new data. The predictive distribution of the dependent variable values $\boldsymbol{y}^*$ at a set of new inputs $\boldsymbol{X}^*$ is also a Gaussian distribution, given by

$$\boldsymbol{y}^* | \boldsymbol{X}^*, \ \mathcal{D} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\mathcal{C}}) \tag{19}$$

where $\boldsymbol{\mu} = \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})\boldsymbol{K}_y^{-1}\boldsymbol{y}$, and $\boldsymbol{\mathcal{C}} = \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X}^*) - \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})\boldsymbol{K}_y^{-1} \times \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})^T + \sigma_n^2 \boldsymbol{I}_{n^*}$. Here, n^* is the number of new inputs at which predictions are desired; $\boldsymbol{X}^*$ is the $n^* \times d$ matrix of the set of new inputs; $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})$ is the set of training inputs and dependent variable values (used for hyperparameter inference); $\mathcal{N}(\mu, \mathcal{C})$ refers to a multivariate Gaussian distribution with mean vector μ and covariance matrix $\mathcal{C}$; $\boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})$ is the $n^* \times n$ covariance matrix derived from the noisy kernel function over pairs of new and training input variables [Eq. (15)]; $\boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X}^*)$ is the $n^* \times n^*$ covariance matrix derived from the noisy kernel function over pairs of the new input variables; and $\boldsymbol{I}_{n^*}$ is the $n^* \times n^*$ identity matrix.

In contrast to other regression methods (such as ordinary least-squares regression, classification, and regression trees), GPR is a nonparametric method of regression and does not need the selection of suitable basis functions. Nonparametric methods are useful in problems with no prior knowledge of the exact functional form of the feature vectors (which may be nonlinear). 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. This uncertainty is a cumulative effect of system operational variability, modeling assumptions, unmodeled factors, and random noise in system performance. These factors, along with mathematical tractability of the Gaussian distributions involved, make GPR the regression method of choice in this paper.

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. Sixty-five percent 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. The GPR starts with hyperparameter inference for the different noisy kernel functions (described in Sec. IV) using the ground-truth FDR values of the predictor and the predicted variables during takeoff roll in the training dataset. 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 [26] based GPstuff Toolbox [27] is used for GPR in this study. Once the models are trained and hyperparameters are inferred, they can be used to determine the point estimates, the prediction intervals, and the predictive distributions of the TOW at a new input vector. Under GPR, the TOW predictive distribution is a normal/Gaussian distribution.

VI. Model Evaluation

The models are evaluated for their performance in predicting (estimating) the TOW of flights in an independent/unseen dataset not used for training. The mean TOW estimates and the 95% prediction intervals are calculated using the regression models developed in Sec. V. The mean estimates (point predictions) are the mean values of the TOW predictive distributions. The 95% prediction intervals are given by the 95% highest density intervals [28] of the predictive distributions for the TOW. A 95% highest density interval is an interval in the domain of a probability distribution such that 1) the probability mass within the interval is 0.95, and 2) 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.

The metrics used to evaluate the models are as follows:

1) Mean error (ME): This is the mean of the values of the relative prediction error on independent/unseen prediction data not used for model training:

$$\mathrm{ME} = \frac{1}{n^*}\sum_{i=1}^{n^*}\left(\frac{\widehat{m_{\mathrm{TO}i}} - m_{\mathrm{TO}_i}}{m_{\mathrm{TO}i}}\right) \tag{20}$$

Here, n^* is the number of observations in the prediction dataset, $\widehat{m_{\mathrm{TO}i}}$ is the mean (point) estimate of the TOW of flight i from the model, and $m_{\mathrm{TO}i}$ is the actual TOW of flight i in the prediction dataset. The ME reflects whether the modeled point estimate overpredicts or underpredicts the ground truth. A negative value of ME indicates that the model point predictions, on average, underpredict the ground truth. A positive value of ME indicates that the model point predictions, on average, overpredict the ground truth. Thus, the ME is an indicator of the bias in the model predictions. It does not reflect the accuracy of the model point predictions.

2) Mean absolute error (MAE): This is the mean of the absolute values of the relative prediction error on independent prediction data not used for model training:

$$\mathrm{MAE} = \frac{1}{n^*}\sum_{i=1}^{n^*}\left|\frac{\widehat{m_{\mathrm{TO}i}} - m_{\mathrm{TO}_i}}{m_{\mathrm{TO}i}}\right| \tag{21}$$

The MAE reflects the accuracy of the model point prediction. A model with a low MAE is desired (with zero being the lowest possible value).

3) Prediction coverage (PC): This is the fraction of the observations in the independent prediction set for which the ground-truth values of the TOW lie within the 95% prediction intervals given by the model. The PC indicates how well the prediction intervals capture the variability of the output variable. A PC value close to 95% indicates that the model has been properly specified and formulated.

4) Normalized length of prediction interval (NLPI): This is the mean of the length of the 95% prediction intervals expressed as a fraction of the point estimate. The NLPI indicates the extent of relative uncertainty present in the estimated output.

These metrics are used for model selection (using the validation dataset) as well as for evaluating the selected model (using the test dataset). A validation study using the validation datasets is done to choose the appropriate kernel functions to build the final GPR models to predict the TOW. The different kernels are compared in terms of these metrics using statistical multicomparison techniques. For each aircraft type, the kernel function that gives the overall statistically significantly (at a 5% significance level) best predictive performance on the validation dataset is selected to build the final model for TOW prediction.

VII. Model Results and Comparisons with Other Models

Table 2 shows the performance of the finally selected GPR models in predicting the TOW on the test datasets for the different aircraft types. The values of the predictor variables in the test data are obtained from FDR data. The FDR data also provide the ground-truth values of the TOW for model evaluation. The table also shows the performance of the TOW estimation model given by the ANP database. A part of AEDT, the ANP database models the TOW as a piecewise constant function of the flight stage/trip length [4]. The flight stage/trip length is determined by calculating the great circle distance between the flight origin and destination airports. Because the ANP model is a deterministic model, its PC and NLPI are not reported in the table. The metrics shown for the individual model performances are calculated across all flights in the test dataset of a particular aircraft type. All the metrics are calculated on destandardized data. For each metric, the table shows the mean value (and the standard deviation within parentheses). Table 2 also shows the statistical comparison of the model MEs with zero and the model MAEs with one another. For comparison of MEs, the null hypothesis is that the MEs are zero (unbiased model predictions). The alternative hypothesis is that the MEs are nonzero (biased model predictions, $\mathrm{ME} \neq 0$). For comparison of MAEs, 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) MAE. The alternative hypothesis is that the GPR model gives a better predictive performance than the ANP model in terms of a lower MAE ($\mathrm{MAE_{GPR}} < \mathrm{MAE_{ANP}}$). The ME and MAE used for this statistical comparison study are calculated on a per-flight basis (and not across flights) for a particular aircraft type. Table 2 shows the p-values obtained through the Wilcoxon signed rank test [29]. The p-value is the probability of observing a value as extreme or more extreme than the calculated test statistic under the null hypothesis. This test is appropriate because the performance of the GPR and the ANP models is to be compared on the same set of flights in the test data (matched data). The p-values indicating acceptance of the alternative hypothesis at the 5% significance level are highlighted in bold.

Figures 3 and 4 contain box plots showing the errors and the absolute errors, respectively, for the TOW prediction across the different flights in the test data. On each box, the central mark is the median; the edges of the box are the 25th and the 75th percentiles; the whiskers extend to the most extreme data points not considered outliers; and outliers are plotted as crosses. The notch in the box represents an interval that can be used for statistical comparison at a

Table 2 TOW estimation: performance metrics for the GPR and the ANP models, on the test datasets for different aircraft types

	ME, %		MAE, %		PC, %	NLPI, %	p-values		
Aircraft type	GPR	ANP	GPR	ANP	GPR	GPR	$\mathrm{ME_{GPR}}$	$\mathrm{ME_{ANP}}$	MAE
A319-112	0.6 (5.9)	1.0 (6.3)	5.0 (3.0)	5.3 (3.3)	84.6 (36.8)	18.4 (0.9)	0.657	0.439	0.319
A320-214	0.4 (4.9)	0.1 (5.1)	3.9 (2.9)	4.3 (2.7)	97.1 (17.1)	19.3 (0.9)	0.778	0.871	0.199
A321-111	−0.3 (7.5)	1.0 (8.5)	6.1 (4.2)	6.7 (5.1)	91.3 (28.8)	23.9 (1.0)	0.808	1.000	0.303
A330-202	−1.8 (2.6)	−4.1 (5.6)	2.4 (1.9)	6.0 (3.3)	94.1 (24.3)	11.2 (2.1)	**0.025**	**0.013**	**0.007**
A330-243	−0.4 (3.4)	0.4 (6.5)	2.6 (2.1)	4.5 (4.5)	85.0 (36.6)	10.2 (1.3)	0.681	0.737	0.058
A330-343	0.2 (4.3)	1.8 (4.1)	3.6 (2.3)	3.5 (2.7)	97.3 (16.4)	17.5 (0.8)	0.874	**0.021**	0.666
A340-541	2.5 (3.8)	12.5 (14.0)	4.0 (1.9)	12.7 (14.0)	90.0 (31.6)	14.4 (9.4)	0.093	**0.009**	**0.023**
B767-300	0.2 (2.5)	−6.5 (6.7)	2.0 (1.5)	8.4 (3.8)	94.4 (23.6)	10.9 (1.8)	0.528	**0.004**	**1.4e–4**
B777-300ER	0.1 (3.4)	3.5 (6.3)	2.0 (2.7)	5.6 (4.5)	92.3 (27.2)	8.2 (2.4)	0.517	**0.007**	**0.002**

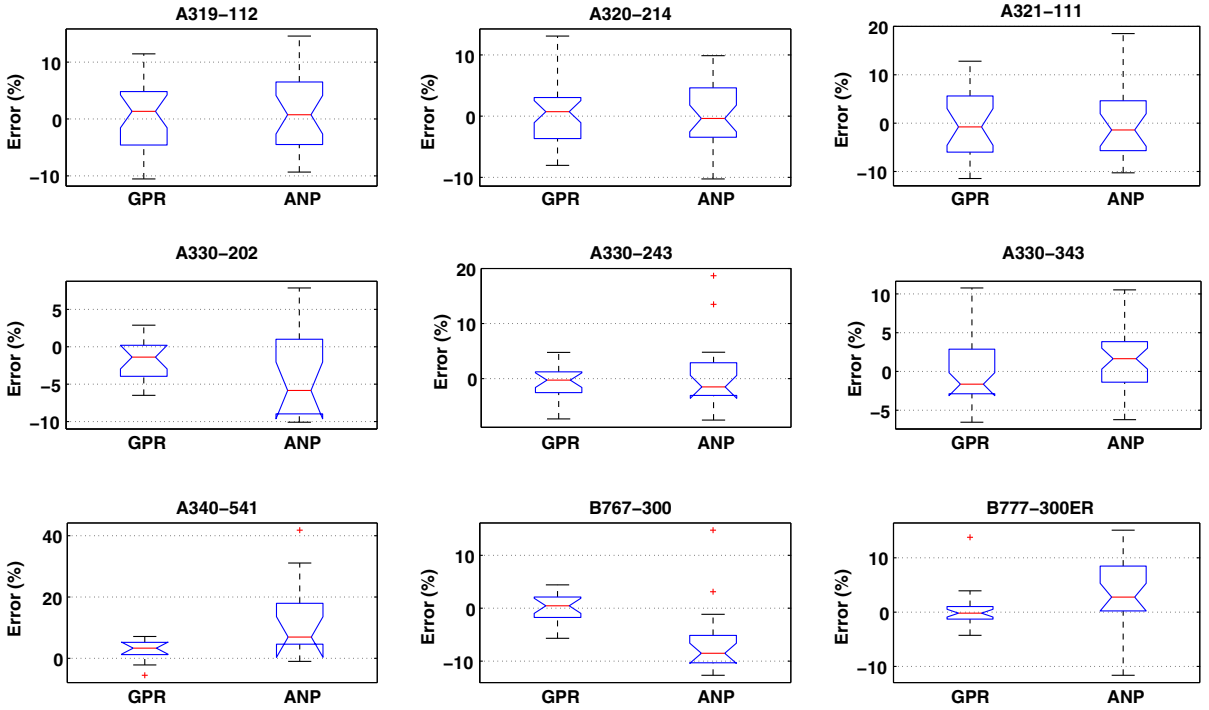

Fig. 3 Box plots showing error on TOW prediction for test data flights.

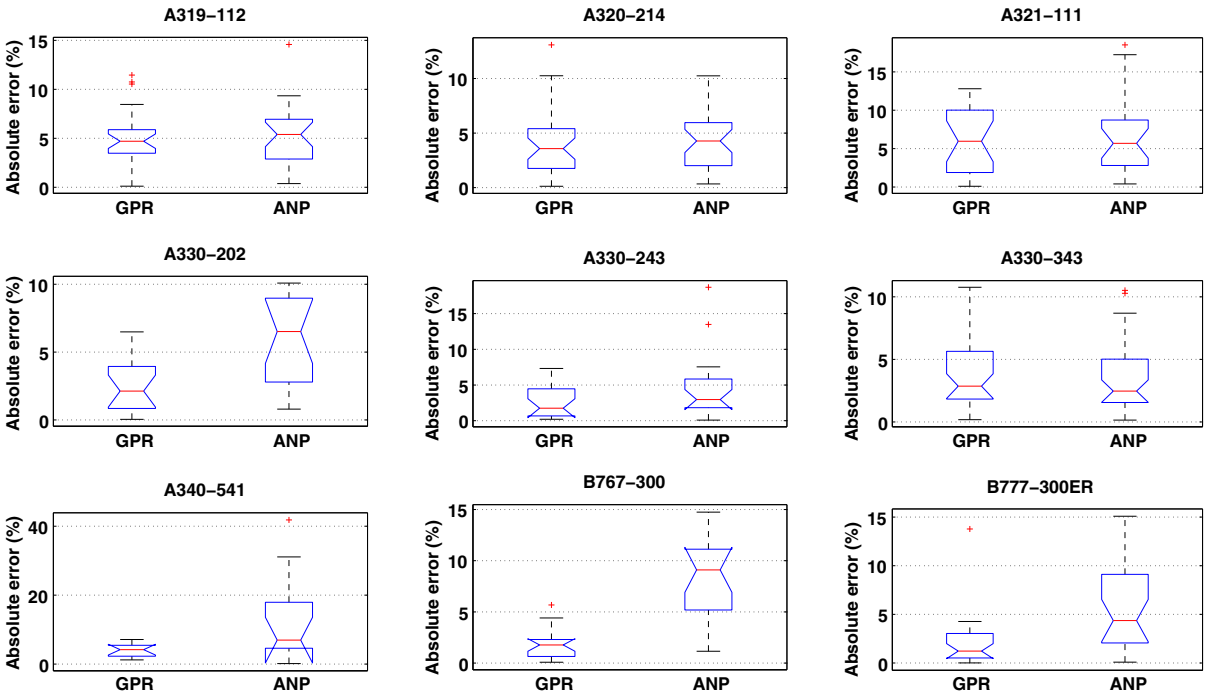

Fig. 4 Box plots showing absolute error on TOW prediction for test data flights.

5% significance level. Model predictions are unbiased (at a 5% significance level) if the error box plots in Fig. 3 contain zero in their notched regions. The notched regions of the absolute error box plots in Fig. 4 for the GPR models can be compared to those of the box plots for the corresponding ANP models to determine which model gives a statistically significantly lower MAE (at a 5% significance level) on the TOW prediction.

From Table 2, it can be seen that the GPR models give unbiased predictions of the TOW in most of the cases. The median of the MEs across the different aircraft types is 0.2%. The GPR models give a median MAE of 3.6% across the different aircraft types, whereas the ANP models give a median MAE of 5.6%. The median PC given by the GPR models is 92.3% across the different aircraft types. The median NLPI is 14.4%. The statistical multicomparison tests for the MAE indicate that the GPR models perform statistically significantly (at the 5% significance level) similar to or better than the ANP models for all the aircraft types in the study. The GPR models reduce the median MAE by 36% as compared to the ANP models. Moreover, unlike the ANP models, the GPR models also give prediction intervals that can quantify the uncertainty in TOW.

VIII. Performance on Ground Surveillance Data

In the previous section, the TOW model predictive performance was evaluated using test data drawing values of the predictor variables from the FDR dataset. In practice, however, the models will be applied to trajectory data obtained through ground surveillance systems. ASDE-X is one such surveillance system. In this section, TOW model predictive performance is evaluated using values of the trajectory variables obtained from ASDE-X. The values of the evaluation metrics determined in this section indicate the predictive performance of the models as they will be used in practice using ground surveillance data.

A. Airport Surface Detection Equipment, Model X

The ASDE-X is a surveillance system at major U.S. airports that is used primarily for collision avoidance on the airport surface [30]. Using data from surface radars, terminal radars, ADS-B sensors, and aircraft transponders, ASDE-X tracks all aircraft and ground vehicles on the airport surface and in the airspace within 8 km of the airport. The ASDE-X archives in this study comprise records for the A330-343, which are inferred to correspond to the flights in the corresponding FDR dataset (based on a maximum likelihood match).

For a particular flight, its ASDE-X record contains track information such as latitude, longitude, speed, heading, and altitude, as functions of time. There is also information about the departure fix, aircraft identification, and aircraft type. The trajectory variables in the ASDE-X dataset have a sampling rate of 1 Hz and resolutions of less than 10^{-6} deg (for latitude, longitude, and heading) and 1 kt for speed. The raw ASDE-X data are very noisy and have missing fields. These tracks, therefore, have been smoothed before use to obtain the noise-free estimates of the ground speeds and the distance covered during takeoff roll.

B. Predictive Performance with Airport Surface Detection Equipment, Model X Data Inputs

Models trained on A330-343 FDR data in Sec. V are applied to trajectory predictor variable (ground speed and takeoff roll distance) values drawn from the ASDE-X data to determine mean predictions and prediction intervals for the TOW. Values of the ambient air density and the ground-truth TOW are still obtained from the corresponding FDR records for the A330-343. Model evaluation is done on 33 flights for the A330-343, which have records in both the FDR and the ASDE-X datasets and which have not been used for TOW model training. Across these 33 flights, the median absolute errors between the smoothed ASDE-X-derived and the FDR values of R, V_{GS_1}, and V_{GS_2} are 3.2, 61.9, and 0.8%, respectively.

Table 3 tabulates the TOW predictive performance when the source of trajectory variables is ASDE-X. Each cell reports the mean and standard deviation (within parentheses) of the evaluation metric across all the flights in unseen data not used for model training. For reference, the predictive performance when the trajectory predictor variables are obtained from FDR data is also tabulated. This predictive performance on FDR-obtained variables may differ from that recorded in Table 2 because different sets of flights have been used for model evaluation in the two tables.

Table 3 Performance of the GPR and the ANP models for the A330-343 to predict TOW when the trajectory predictor variables are obtained from ASDE-X

	ME, %		MAE, %		PC, %	NLPI, %
Trajectory source	GPR	ANP	GPR	ANP	GPR	GPR
ASDE-X	1.9	1.3	3.6	3.4	97.0	17.5
	(4.1)	(4.0)	(2.7)	(2.3)	(17.4)	(1.0)
FDR	1.2	1.3	3.3	3.4	100.0	17.4
	(3.8)	(4.0)	(2.2)	(2.3)	(0.0)	(0.8)

Table 3 shows that, for the A330-343, model TOW predictive performance (in terms of MAE, PC, NLPI) when trajectory variables are obtained from ASDE-X is similar to that when trajectory variables are obtained from FDR. For the A330-343, the GPR model applied to surface surveillance data gives an ME of 1.9%, MAE of 3.6%, PC of 97%, and NLPI of 17.5%. The GPR model gives an MAE statistically similar to the ANP model. The ANP models do not need any trajectory variables as inputs and are therefore independent of the source of such data.

IX. 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 (the mass of fuel consumed per unit time) per engine. The analysis is demonstrated for different aircraft types in the climb out and approach phases of flight. These phases are the parts of ascent and descent, respectively, which take place below 3000 ft above field elevation. Because of its proximity to the ground, the fuel burn in these phases impacts the environment in the vicinity of airports.

A. Fuel Flow Rate Modeling

In previous research [31,32], we showed that the average fuel flow rate per engine in climb out can be statistically estimated by considering the aircraft dynamic pressure multiplied by the wing reference area, aircraft mass, ratio of the vertical speed to the ground speed, ground speed, and rate of change of the ground speed with time as predictor variables. In approach, the aircraft altitude above the mean sea-level elevation of the arrival airport is also included as an additional predictor variable. Using these variables (all in SI units), we develop an aircraft-type specific GPR model for the fuel flow rate in climb out and approach (each phase modeled separately) and use the aircraft TOW in place of the instantaneous aircraft mass as a predictor variable. For this fuel flow rate modeling, each point in the training, validation, and test sets represents one FDR observation in a particular flight. Each flight contributes to multiple observations. All the observations of a particular flight belong to the training, the validation, or the test sets. 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, explained in Secs. IV and V, is again employed to build the fuel flow rate models.

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 model [33]. The TOW, which is a predictor variable for the fuel flow rate models, needs to be estimated because 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. One method for analyzing the sensitivity of the fuel flow rate to TOW is presented here. The values of the TOW in the unseen prediction dataset not used for training (combined validation and test datasets) for each aircraft type are changed systematically by a fixed percentage of the true value to give a modified prediction dataset [34]. All the other predictors in the prediction set are held at their original values (which is an approximation because the other variables might depend on TOW too). GPR fuel flow rate models trained using the true TOW are run on the modified prediction dataset, and the mean absolute error in the predicted fuel flow rate is calculated. The variation of this error with the percentage deviation of the estimated TOW in the modified prediction dataset from its true value indicates the sensitivity of the model fuel flow rate predictions to the TOW.

Table 4 Increase in fuel flow rate MAE for a +3% deviation in TOW from its actual value

Aircraft type	Phase	Increase in MAE, %
A319-112	Climb out	12.9
	Approach	5.9
A320-214	Climb out	19.2
	Approach	5.2
A321-111	Climb out	18.0
	Approach	4.9
A330-202	Climb out	15.0
	Approach	4.3
A330-243	Climb out	3.9
	Approach	3.1
A340-541	Climb out	6.7
	Approach	1.0
B767-300	Climb out	12.8
	Approach	9.7
B777-300ER	Climb out	4.7
	Approach	3.1

Table 4 tabulates the percentage change in mean absolute error in fuel flow rate prediction in climb out and approach due to a +3% deviation in estimated TOW from its actual value. Fuel flow rate GPR models [32] are used in this sensitivity analysis. The table entries are obtained by averaging across all flights in the combined validation and test datasets (unseen data not used for training). The model predictions are seen to be sensitive to even a 3% deviation from the true TOW, with MAE changing by as much as about 19% for the A320-214 in climb out. This sensitivity of the MAE in fuel flow rate prediction to the deviation in the estimated TOW motivates the need to accurately predict the TOW, in order to accurately predict the fuel flow rate.

Using the TOW predicted by the ANP models as well as by the developed GPR models (Sec. V) as inputs to the fuel flow rate GPR models, their predictive performance on the flights in the unseen prediction dataset (combined validation and test datasets) is now evaluated. To incorporate uncertainty in the predicted TOW, the fuel flow rate predictive performance is evaluated using the predictive distribution of the fuel flow rate marginalized over the uncertain values of TOW. In other words, we are interested in computing the following:

$$p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, \boldsymbol{\phi}, \mathcal{D}_1, \mathcal{D}_2) = \int_{m_{\mathrm{TO}}} p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1) \times p(m_{\mathrm{TO}}|\boldsymbol{\phi}, \mathcal{D}_2)\, \mathrm{d}m_{\mathrm{TO}} \tag{22}$$

Here, p refers to the PDF; $\dot{m}_f$ is the fuel flow rate to be predicted; $\boldsymbol{x}_{-\mathrm{TOW}}$ is the vector of predictor variables in the fuel flow rate GPR model excluding the TOW; m_{TO} is the TOW; and $\mathcal{D}_1$ is the set of the training variables used for building the fuel flow rate GPR model. $p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1)$ is the PDF of the predictive distribution given by the fuel flow rate GPR model and is thus a Gaussian PDF. $p(m_{\mathrm{TO}}|\boldsymbol{\phi}, \mathcal{D}_2)$ is the distribution of the predicted TOW parameterized by $\boldsymbol{\phi}$ and $\mathcal{D}_2$.

The ANP model is a deterministic model giving a flight stage length-based point estimate of the TOW, $m_{\mathrm{TO,ANP}}$. Under the ANP model, Eq. (22) becomes

$$p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, \boldsymbol{\phi}, \mathcal{D}_1, \mathcal{D}_2) = p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, m_{\mathrm{TO,ANP}}, \mathcal{D}_1) \tag{23}$$

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

The GPR models for TOW prediction (Sec. V) give the complete predictive distribution for the TOW (which is a normal distribution). Therefore, under a GPR model for TOW estimation, Eq. (22) becomes

$$p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, \boldsymbol{\phi}, \mathcal{D}_1, \mathcal{D}_2) = \int_{m_{\mathrm{TO}}} p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1) \times p(m_{\mathrm{TO}}|\boldsymbol{\phi}, \mathcal{D}_2)\, \mathrm{d}m_{\mathrm{TO}} \tag{24}$$

$$\approx \frac{1}{n_s} \sum_{i=1}^{n_s} p(\dot{m}_f|\boldsymbol{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}_i}, \mathcal{D}_1) \tag{25}$$

When GPR models are used to predict the TOW, $\boldsymbol{\phi}$ and $\mathcal{D}_2$ hold specific meanings. $\boldsymbol{\phi}$ is the vector of predictor variables used in the GPR TOW prediction models (R, ρ_∞, V_{GS_1}, and V_{GS_2}, as mentioned in Sec. III). $\mathcal{D}_2$ is the training set used to build the GPR models to predict the TOW. Equation (25) approximates Eq. (24) 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 prediction. In this study, n_s is chosen to be 1000. Equation (25) therefore shows that the desired predictive distribution of the fuel flow rate under a GPR model of the TOW can be approximately modeled as a Gaussian mixture distribution with n_s equally weighted components.

B. Results

Depending on how the TOW variable is predicted, there are two variants of the GPR models developed to predict the fuel flow rate:

1) Model 1: This variant predicts the TOW predictor variable that is input to the GPR fuel flow rate model using the ANP method.

2) Model 2: This variant predicts the TOW predictor variable using the GPR models developed in Sec. V for TOW prediction.

The predictive distributions for the fuel flow rate marginalized over the TOW for model 1 [Eq. (23)] and model 2 [Eq. (25)] are used to calculate the mean predictions and the 95% highest density prediction intervals for the fuel flow rates. Table 5 tabulates the fuel

Table 5 Fuel flow rate predictive performance using estimated TOW

		MAE, %		PC, %		NLPI, %	
Aircraft type	Phase	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
A319-112	Climb out	3.8 (1.5)	3.9 (1.8)	83.9 (16.3)	92.4 (12.1)	13.6 (0.5)	17.0 (1.2)
	Approach	16.8 (5.4)	16.8 (5.4)	95.0 (4.8)	95.7 (4.3)	94.3 (12.1)	96.5 (12.5)
A320-214	Climb out	4.2 (2.2)	4.3 (2.4)	91.6 (14.5)	96.0 (9.0)	20.9 (1.3)	25.1 (2.2)
	Approach	16.2 (6.3)	16.6 (6.2)	94.6 (4.9)	95.0 (4.2)	104.6 (95.2)	99.3 (51.8)
A321-111	Climb out	7.8 (4.4)	6.5 (2.9)	72.5 (25.0)	93.7 (10.0)	21.2 (1.2)	29.8 (1.7)
	Approach	18.1 (4.7)	17.2 (4.5)	91.5 (5.5)	93.4 (5.0)	76.1 (11.0)	80.8 (12.6)
A330-202	Climb out	5.1 (1.9)	3.6 (1.5)	83.1 (11.9)	92.4 (8.2)	19.7 (1.4)	19.9 (1.4)
	Approach	27.3 (10.1)	28.0 (11.3)	91.0 (7.1)	91.1 (7.3)	146.9 (22.1)	141.7 (24.4)
A330-243	Climb out	3.0 (1.9)	2.8 (2.0)	89.8 (19.6)	90.8 (20.7)	12.6 (0.7)	12.9 (0.8)
	Approach	20.8 (11.5)	20.6 (11.5)	90.8 (9.0)	91.1 (9.0)	111.3 (20.6)	110.8 (21.3)
A340-541	Climb out	9.1 (12.3)	5.1 (5.4)	66.6 (32.9)	83.9 (22.4)	13.9 (0.5)	17.1 (4.9)
	Approach	19.6 (5.5)	17.6 (3.3)	95.7 (3.9)	96.5 (2.7)	120.9 (22.5)	120.8 (18.1)
B767-300	Climb out	5.3 (2.4)	3.0 (1.5)	83.2 (20.0)	98.1 (4.3)	19.2 (1.2)	19.5 (2.0)
	Approach	20.0 (7.4)	19.8 (6.7)	93.3 (7.5)	95.6 (5.2)	147.8 (58.6)	147.0 (212.6)
B777-300ER	Climb out	8.1 (4.3)	6.7 (2.5)	87.0 (10.1)	93.0 (7.7)	29.0 (3.4)	31.8 (5.8)
	Approach	17.4 (5.6)	16.5 (5.3)	93.8 (5.6)	94.1 (5.2)	99.5 (17.0)	101.8 (17.9)

flow rate predictive performance of these models on the unseen prediction dataset (combined validation and test datasets) in climb out and approach for different aircraft types. Each cell entry records the mean (and the standard deviation within parentheses) of the evaluation metric. Statistical multiple comparison tests (at a significance level of 5%) have been performed to determine which of the two models perform better on different evaluation metrics.

In climb out, for the majority of the aircraft types, model 2 gives a lower or similar MAE as compared to model 1 and a higher or similar PC as compared to model 1. Thus, for the majority of the aircraft types, using the GPR-TOW model gives a better or similar fuel flow rate predictive performance as compared to the ANP-TOW prediction model. However, the NLPI for model 2 is greater than or equal to that for model 1. This is expected as model 2 propagates uncertainty in TOW to that in fuel flow rate and hence leads to bigger prediction intervals. The ANP-TOW prediction model gives a fuel flow rate median MAE, PC, and NLPI of 5.2, 83.6, and 19.5%, respectively, in climb out. The GPR-TOW prediction model gives a fuel flow rate median MAE, PC, and NLPI of 4.1, 92.7, and 19.7%, respectively, in climb out.

In approach, for the majority of the aircraft types, model 2 gives a lower or similar MAE as compared to model 1 and a higher or similar PC as compared to model 1. Thus, for the majority of the aircraft types, using the GPR-TOW model gives a better or similar fuel flow rate predictive performance as compared to the ANP-TOW prediction model. However, the NLPI for model 2 is greater than or equal to that for model 1 for the majority of the aircraft types. This is again expected as model 2 propagates uncertainty in TOW to that in fuel flow rate and hence leads to bigger prediction intervals. The ANP-TOW prediction model gives a fuel flow rate median MAE, PC, and NLPI of 18.9, 93.6, and 108.0%, respectively, in approach. The GPR-TOW prediction model gives a fuel flow rate median MAE, PC, and NLPI of 17.4, 94.6, and 106.3%, respectively, in approach.

The difference in fuel flow rate predictive performance for the GPR-TOW and the ANP-TOW prediction models in climb out is starker than that observed in approach. A greater number of aircraft types give similar performance from the two models for approach than for climb out. This observation can also be made from Table 4, in which the increase in fuel flow rate MAE is greater in climb out than in approach for a similar deviation of the estimated TOW from its true value.

X. 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 linear, quadratic, or some other form of basis functions before 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 the present 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.

It was possible to validate and evaluate the present models using an independent test set of FDR data. Metrics were developed to quantify the accuracy of both the point (that is, mean) estimates as well as the uncertainty estimates of the TOW. The present GPR models gave a median MAE of 3.6% and a median PC of 92.3% on the test FDR data (averaged across the different aircraft types studied). The model performance was also compared with that of the Aircraft Noise and Performance model; the GPR models were shown to predict the TOW statistically significantly (at the 5% significance level) similar to or more accurately than the ANP model. When used on ASDE-X surface surveillance data for the A330-343, the GPR model was found to give an MAE of 3.6% and a PC of 97%, thereby demonstrating the practical utility of the TOW prediction models developed in this paper.

Finally, an application of the TOW models to estimate aircraft engine fuel flow rates in climb out and approach 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 fuel flow rate estimate. Therefore, the TOW was first estimated and further used to estimate the fuel flow rate. As a result, the accuracy of fuel rate estimate depended on the accuracy of the TOW model. The impact of using two different TOW models (the present GPR model and the ANP model) as input to a GPR model for the fuel flow rate was investigated. It was shown that the fuel flow rate model that used the GPR estimate of TOW (that is, model 2) performed similar to or better than the fuel flow rate model that used the ANP estimate of TOW (that is, model 1) in both climb out and approach for the majority of the aircraft types studied. Using the GPR estimates of TOW, the median MAE in predicting the fuel flow rate is shown to be 4.1% in climb out and 17.4% in approach. The median PC for the fuel flow rate predictions is 92.7% in climb out and 94.6% in approach. More detailed discussions of these results can be found in [35].

The amount of flight data used for TOW model building were 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.

Acknowledgments

This research was supported in part by the National Science Foundation under grant number 1239054, CPS: Frontiers: FORCES. The authors are also grateful to Sandeep Badrinath (Massachusetts Institute of Technology) for developing the algorithm to smooth the ASDE-X data.

References

[1] Tulapurkara, E., "National Programme on Technology Enhanced Learning (NPTEL) India: Lectures on Airplane Design (Aerodynamic): Weight Estimation," Indian Inst. of Technology Madras, Chennai, India, 2013, http://nptel.ac.in/courses/101106035/009_Chapter%03_L6_(02-10-2013).pdf [retrieved Jan. 2017].

[2] Raymer, D. P., *Aircraft Design: A Conceptual Approach*, 5th ed., AIAA, Reston, VA, 2012.

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

[4] Ahearn, M., Boeker, E., Gorshkov, S., Hansen, A., Hwang, S., Koopmann, J., Malwitz, A., Noel, G., Reherman, C., Senzig, D., Solman, G. B., Tosa, Y., Wilson, A., Zubrow, A., Didyk, N., DiPardo, J., Grandi, F., Majeed, M., Bernal, J., Dinges, E., Rickel, D., Yaworski, M., Hall, C., and Augustine, S., "Aviation Environmental Design Tool (AEDT) Technical Manual, Version 2b, Service Pack 3," Rept. DOT-VNTSC-FAA-16-11, U.S. Dept. of Transportation, Volpe National Transportation Systems Center, Cambridge, MA, June 2016.

[5] Sherry, L., and Neyshabouri, S., "Estimating Takeoff Thrust from Surveillance Track Data," *Transportation Research Board Annual Meeting*, Jan. 2014.

[6] "Air Carrier Statistics Database," U.S. Dept. of Transportation, Washington, D.C., 2016, https://www.transtats.bts.gov/Tables.asp?DB_ID=111 [retrieved Jan. 2017].

[7] Alligier, R., Gianazza, D., and Durand, N., "Energy Rate Prediction Using an Equivalent Thrust Setting Profile," *Proceedings of the International Conference on Research in Air Transportation*, May 2012.

[8] Alligier, R., Gianazza, D., and Durand, N., "Ground-Based Estimation of the Aircraft Mass, Adaptive vs. Least Squares Method," *Proceedings*

of the USA/Europe Air Traffic Management Research and Development Seminar, June 2013, Paper 198.

[9] Alligier, R., Gianazza, D., and Durand, N., "Learning the Aircraft Mass and Thrust to Improve the Ground-Based Trajectory Prediction of Climbing Flights," *Transportation Research Part C: Emerging Technologies*, Vol. 36, Nov. 2013, pp. 45–60. doi:10.1016/j.trc.2013.08.006

[10] Alligier, R., Gianazza, D., Hamed, M. G., and Durand, N., "Comparison of Two Ground-Based Mass Estimation Methods on Real Data," *Proceedings of the International Conference on Research in Air Transportation*, May 2014.

[11] Alligier, R., Gianazza, D., and Durand, N., "Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 6, Dec. 2015, pp. 3138–3149. doi:10.1109/TITS.2015.2437452

[12] Nuic, A., "User Manual for the Base of Aircraft Data (BADA) Revision 3.13," Rept. 15/04/02-43, Eurocontrol Experimental Centre, Brétigny-sur-Orge, France, 2015.

[13] Sun, J., Ellerbroek, J., and Hoekstra, J., "Modeling and Inferring Aircraft Takeoff Mass from Runway ADS-B Data," *Proceedings of the International Conference on Research in Air Transportation*, June 2016.

[14] Sun, J., Ellerbroek, J., and Hoekstra, J., "Bayesian Inference of Aircraft Initial Mass," *Proceedings of the USA/Europe Air Traffic Management Research and Development Seminar*, June 2017, Paper 25.

[15] Kulikov, G. G., and Thompson, H. A., (eds.), *Dynamic Modelling of Gas Turbines: Identification, Simulation, Condition Monitoring, and Optimal Control*, Springer–Verlag, London, 2004.

[16] "Aircraft Noise and Performance (ANP) Database," Eurocontrol Experimental Centre, https://www.aircraftnoisemodel.org [retrieved Jan. 2017].

[17] Jackson, P., (ed.), *Jane's All the World's Aircraft: Development and Production*, Jane's by Information Handling Services Markit, 2017–2018.

[18] Hunter, J., (ed.), *Jane's All the World's Aircraft: In Service*, Jane's by Information Handling Services Markit, 2017–2018.

[19] Jackson, P., (ed.), *Jane's All the World's Aircraft*, International Thomson Publ. Company, 1995–1996.

[20] Chati, Y. S., and Balakrishnan, H., "Analysis of Aircraft Fuel Burn and Emissions in the Landing and Take Off Cycle Using Operational Data," *Proceedings of the International Conference on Research in Air Transportation*, May 2014.

[21] Hultquist, C., Chen, G., and Zhao, K., "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, 2014, pp. 723–732. doi:10.1080/2150704X.2014.963733

[22] Pasolli, L., Melgani, F., and Blanzieri, E., "Gaussian Process Regression for Estimating Chlorophyll Concentration in Subsurface Waters from Remote Sensing Data," *IEEE Geoscience and Remote Sensing Letters*, Vol. 7, No. 3, July 2010, pp. 464–468. doi:10.1109/LGRS.2009.2039191

[23] Chai, K. M. A., Williams, C. K. I., Klanke, S., and Vijayakumar, S., "Multi-Task Gaussian Process Learning of Robot Inverse Dynamics," *Advances in Neural Information Processing Systems 21*, 2009, pp. 265–272.

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

[25] Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., and Vehtari, A., "Bayesian Modeling with Gaussian Processes Using the GPstuff Toolbox," arXiv:1206.5754 [stat.ML], 2015.

[26] Matlab, Software Package, Ver. R2012a, Mathworks, Natick, MA, 2012.

[27] Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., and Vehtari, A., "GPstuff: Bayesian Modeling with Gaussian Processes," *Journal of Machine Learning Research*, Vol. 14, No. 1, April 2013, pp. 1175–1179.

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

[29] Wilcoxon, F., "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, Vol. 1, No. 6, Dec. 1945, pp. 80–83. doi:10.2307/3001968

[30] "*ASDE-X Brochure*," Sensis Corp., East Syracuse, NY, 2008.

[31] Chati, Y. S., and Balakrishnan, H., "A Gaussian Process Regression Approach to Model Aircraft Engine Fuel Flow Rate," *Proceedings of the ACM/IEEE International Conference on Cyber-Physical Systems*, ACM, New York, April 2017, pp. 131–140. doi:10.1145/3055004.3055025

[32] Chati, Y. S., and Balakrishnan, H., "Data-Driven Modeling of Aircraft Engine Fuel Burn in Climb Out and Approach," *Transportation Research Record: Journal of the Transportation Research Board* (to be published).

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

[34] Chati, Y. S., and Balakrishnan, H., "Statistical Modeling of Aircraft Engine Fuel Flow Rate," *Proceedings of the Congress of the International Council of the Aeronautical Sciences*, ICAS Paper ICAS2016_0619, Sept. 2016.

[35] Chati, Y. S., "Statistical Modeling of Aircraft Engine Fuel Burn," Ph.D. Thesis, Dept. of Aeronautics and Astronautics, Massachusetts Inst. of Technology, Cambridge, MA, Feb. 2018.

J. Krozel
Associate Editor