# Predicting Conceptual Aircraft Design Parameters Using Gaussian Process Regressions on Historical Data

Maxfield Arnson [1], Rawan Aljaber [2], and Gokcin Cinar [3]
*Integrated Design of Environmentally-friendly Aerospace Systems Laboratory*
*Department of Aerospace Engineering*
*University of Michigan, Ann Arbor, Michigan, 48109, USA*

**This paper presents a novel methodology for predicting key aircraft design parameters using Gaussian Process Regressions (GPR) applied to a comprehensive, open-source database of over 400 aircraft and 200 engines. The database, made freely available through the Future Aircraft Sizing Tool (FAST), enables aircraft designers to apply detailed historical data in early-stage conceptual design and provides full visibility into the data underlying the regressions—unlike traditional regression models, where the training data and model fit are often not disclosed. The non-parametric GPR models developed in this work allow for flexible input configurations, significantly improving the accuracy of predicting critical parameters, such as Operating Empty Weight (OEW) and engine weight, compared to established regressions from the literature. By incorporating a broader range of inputs—including maximum takeoff weight, payload, range, and sea level static thrust—these models reduce prediction errors and provide tighter error distributions, leading to more reliable estimates during early design phases. This paper outlines a methodology for adapting conventional aircraft data to explore hybrid-electric and fully electric aircraft designs, ensuring that historical data can be leveraged effectively for novel propulsion systems. Additionally, the flexibility of the GPR framework allows users to create their own regressions and update predictions as new data becomes available, making it a useful tool for researchers working with their own datasets.**

## Nomenclature

| | | |
|---|---|---|
| APM | = | Airport Planning Manual |
| ARD | = | Automatic Relevance Determination |
| CAA | = | Civil Aviation Authority (United Kingdom) |
| EASA | = | European Union Aviation Safety Agency |
| EIS | = | Entry-into-Service |
| EPFD | = | Electrified Powertrain Flight Demonstration |
| FAA | = | Federal Aviation Administration |
| FAST | = | Future Aircraft Sizing Tool |
| GPR/GPM | = | Gaussian Process Regression/Model |
| IDEAS | = | Integrated Design of Environmentally-friendly Aerospace Systems (Laboratory) |
| MTOW | = | Maximum Takeoff Weight |
| OEW | = | Operating Empty Weight |
| TCDS | = | Type Certificate Data Sheet(s) |

## I. Introduction

Conceptual aircraft aimed at sustainably advancing the aviation industry currently lack the formalized design processes that traditional aircraft benefit from. Typically, modern aircraft design begins with an analytical approach based on historical regressions, drawing from extensive aircraft databases. Major aerospace companies, such as Boeing

[1] Ph.D. Candidate, Department of Aerospace Engineering, University of Michigan, Ann Arbor, Michigan 48109, AIAA Student Member
[2] Graduate Student, Department of Aerospace Engineering, University of Michigan, Ann Arbor, Michigan 48109, AIAA Student Member
[3] Assistant Professor, Department of Aerospace Engineering, University of Michigan, Ann Arbor, Michigan 48109, AIAA Senior Member

or Airbus, maintain their own proprietary databases. While popular aircraft design methods, like those by Roskam [1] and Raymer [2], are widely used, they primarily focus on conventionally (i.e. tube-and-wing, Jet-A powered) aircraft. To bridge the gap between current and future designs, "transition" models are expected to emerge, linking today's aircraft evolution to more advanced, sustainable designs of the future.

The transition designs will likely be modified and improved versions of modern aircraft. Electrification, technological efficiency improvements, and sustainable drop-in fuels will be integrated into the industry before large scale alternative airframe or propulsion systems are commonplace [3] . Sigmoid curves (also referred to as "S-curves") are often used to describe the improvement of technology over time, starting slowly then rapidly progressing before seeing diminishing improvements towards the end of a lifecycle. Technologically, it is expected that transition aircraft designs will push the limits of current component maturation until breakthrough technology defines an entirely new Sigmoid curve to follow.

Traditionally, conceptual aircraft designs are developed based on historical data, as outlined by renowned authors such as Roskam, Raymer, and Jenkinson [1, 2, 4]. This historical data, applied through regression methods, is used to estimate key performance metrics, such as weight and range, in the early stages of aircraft design. This approach continues until a design layout is finalized, enabling more detailed analyses. However, conventional regression models are limited when it comes to novel aircraft concepts, such as those employing fully or partially electric (i.e., electrified) propulsion systems, or those designed to use alternative fuels like hydrogen. The absence of historical data for such unconventional designs presents a significant challenge in accurately predicting their performance.

Despite historical data only being available for conventional aircraft, recent studies have shown that by revisiting definitions of weight ratios, historical data can provide insights into feasibility of alternative aircraft as well [5]. Thus, even for novel systems, historical data still holds value in conceptual design. By developing more granular regressions, it is possible to integrate emerging trends—such as new weight and performance metrics—into conventional methods, thereby retaining the decades of experience embedded in historical data. This work provides the foundation for such an approach by leveraging historical regressions and adapting them for novel aircraft concepts, especially those which bridge the gap between contemporary and advanced designs.

This paper presents the foundation for this approach, demonstrating how historical data can be used in a more flexible and granular manner. By developing Gaussian Process Regression (GPR) models, this work enables the use of historical data to predict design parameters for electrified and alternative fuel aircraft while preserving the benefits of traditional regression-based methods. Importantly, these models are integrated into the Future Aircraft Sizing Tool (FAST) [6, 7], which provides open-source access to both the historical database and the regression tools, allowing designers to apply these methods to their own data or models.

Several publicly available databases offer information on aircraft and engine specifications, with Janes' *All the World's Aircraft* [8] being the most widely recognized. This database is updated annually, but it requires manual data entry for analysis, which can be time-consuming. Similarly, Roux's database handbooks on turbofan and turboprop engines [9, 10] provide detailed engine information, but they also require manual entry and are often focused on specific subsets of aircraft. Table 1 summarizes the key existing aircraft and engine databases and their limitations.

**Table 1 Examples of existing aircraft/engine databases and their limitations.**

| Author | Year | Database | Type | Limitations |
|---|---|---|---|---|
| Roux [9, 10] | 2007/2011 | Turbofan/Turboprop | Book | Format/Date |
| Janes [8] | 2023 | Aircraft | Book | Accessibility |
| FAA [11] | 2023 | Aircraft | Excel | Scale |
| Eurocontrol [12] | 2024 | Aircraft | Website | Scale |
| Jenkinson, Simpkin, Rhodes [13] | 2001 | Aircraft/Engine | Book/Website | Scale |
| Meier [14] | 2023 | Aircraft/Engine | Website | Incomplete |

To address these limitations, this work compiles a new historical aircraft and engine database, drawing from Type Certificate Data Sheets (TCDS) published by the FAA, EASA, and other aviation authorities, supplemented with information in existing databases. This database, originally developed for FAST, is an open-source resource and designed to be collaboratively updated, and it is intended to predict key aircraft parameters in the conceptual design phase. FAST uses GPR models to predict unknown parameters based on fundamental aircraft requirements, such as range and payload weight, along with designer-specified constraints, such as thrust-to-weight-ratio or wing loading.

This work aims to provide up-to-date, publicly available data on conventional aircraft that can be easily accessed, manipulated, and analyzed by aircraft designers of all experience levels. The methodology for creating and tuning the regressions used in FAST is fully documented, and the models are designed to be adaptable for novel aircraft designs. The results show that the GPR-based regressions offer improved accuracy over historical baselines and are well-suited for hybrid electric aircraft design. Section II discusses the methodology for this work. Methods describe data collection, regression procedures, and hyperparameter tuning. Section III connects the regressions to the FAST tool itself, showing where they are used for conceptual aircraft design. Section IV shows the performance of FAST regressions against historical references. These example regressions include OEW, engine weights, and fuselage length. After the concluding remarks, a short appendix in Section V.A describes how to run a basic regression in FAST.

## II. Methodology

This section describes in detail how data was collected for use in the database, in addition to any post processing that was performed on the data. Additionally, this section describes the methodology used to develop a probabilistic regression to utilize the data in aircraft sizing. It goes on to explain tuning methods which are essential to the functionality of the regressions. Finally, this section explains how the data is used within the FAST tool to assist in quick conceptual aircraft design, with the goal of improving conventional regressions on common parameter estimates in the design process.

### A. Data Collection

Historical data serves as a foundational element for forecasting future trends in aircraft and engine key performance parameters (KPPs). To ensure reliable projections, a substantial database was required. Many existing databases lack proper source citations and seldom offer the level of specificity needed for the intended analysis. For example, while a database might list general technical specifications for the Airbus A380-800, it often overlooks distinctions between variants, such as the -841, -842, and -861, which use different engines. A key contribution of this database is the recording of each variant as a unique vehicle, accounting for performance variations resulting from different engine options and configurations. This level of granularity distinguishes the database from other publicly available resources, making it particularly valuable for detailed performance analysis and conceptual design.

External databases vary in their formality and completeness. For instance, the Jet Engine Specification Database is an informally documented, "blog-style" database that provides ample parameters but lacks comprehensive data collection [14]. A more formal source is the database by Jenkinson, Simpkin, and Rhodes [13], used in their aircraft design textbook, *Civil Jet Aircraft Design* [4]. Although well-documented, this database focuses on fewer categories comprised of Aircraft, Engines, Airports, Atmosphere and Airspeeds.

The Integrated Design of Environmentally-friendly Aircraft Systems (IDEAS) Laboratory at the University of Michigan aimed to make a densely populated database without compromising the number of recorded parameters [6]. As a result, a custom database of commercial aircraft was developed, consisting of over 400 aircraft and over 200 turbofan engines, along with a select number of turbojet engines. The aircraft database interacts with the engine database to describe a complete vehicle. Each aircraft in the database is represented by 87 parameters, and each engine by 54 parameters. To ensure accuracy, only parameters directly gathered from reliable sources were recorded into the collection. A post-processing script was created and runs in MATLAB to extract any information required from the raw data. Examples of calculated parameters include Operating Empty Weight to Maximum Takeoff Weight ratio (OEW/MTOW), Thrust-to-Weight ratio, Taper Ratio, and Lift-to-Drag ratio, among others.

Data was primarily collected from TCDS published online by aviation authorities such as the FAA, EASA, CAA, and others. Additional information was sourced from aircraft manufacturer materials and airport planning manuals. However, fragmented version histories and proprietary information often resulted in an incomplete database. Inconsistencies arose from aircraft variants, certification sheet disparities, and inconsistent conventions. Despite these challenges, a standardized format was adhered to during data collection.

As an example, the data collection process for the Boeing 737-200 began by locating FAA TCDS A16WE on the FAA's website. A portion of this TCDS is seen in Fig. 1. This document lists every certified variant of the 737, including all versions of the aircraft family and certified engine options. The TCDS commonly reports certification year, aircraft weights, maximum operating passenger limits, altitudes, velocities, and thrusts for each engine option. If the data sheet guides a user to notes or flight manuals, those sources are also investigated, such as the payload range diagram shown in Fig. 2. A diagram like this one can be found in a airport planning manual (APM) or "flight manual", which may also contain information about nuanced differences between aircraft. For example, an APM could include the weight differences between two versions of the same aircraft, one with passenger entertainment centers installed and one without. These manuals often show how the payload-range diagrams change as different engines are installed on the same airframe, or have information on the aircraft geometry (e.g. wing area, aircraft binding box dimensions, etc.) if CAD files are not available on a manufacturer's website.

<u>II - Model 737-200 (Approved December 21, 1967) Transport Aircraft</u>

| | | | |
|---|---|---|---|
| **Engines:** | 2 Pratt and Whitney Turbofan Engines JT8D-7, JT8D-7A, JT8D-7B, JT8D-9, JT8D-9A, JT8D-15, JT8D-15A, JT8D-17, and JT8D-17A; Refer to the FAA Approved Airplane Flight Manual for aircraft engine and engine intermix eligibility. (Engine Type Certificate No. E2EA) | | |
| Fuel: | See NOTE 4 for authorized types of fuel. | | |
| **Engine Ratings:** | | Takeoff static thrust, standard day, sea level conditions (5 min) lb. | Maximum continuous static thrust, standard day, sea level conditions lbs. |
| | JT8D-7, -7A, -7B | 14,000 | 12,600 |
| | JT8D-9, -9A | 14,500 | 12,600 |
| | JT8D-15, -15A | 15,500 | 13,750 |
| | JT8D-17, -17A | 16,000 | 15,200 |

**Fig. 1 FAA TCDS A16WE's section describing the 737-200 [15].**

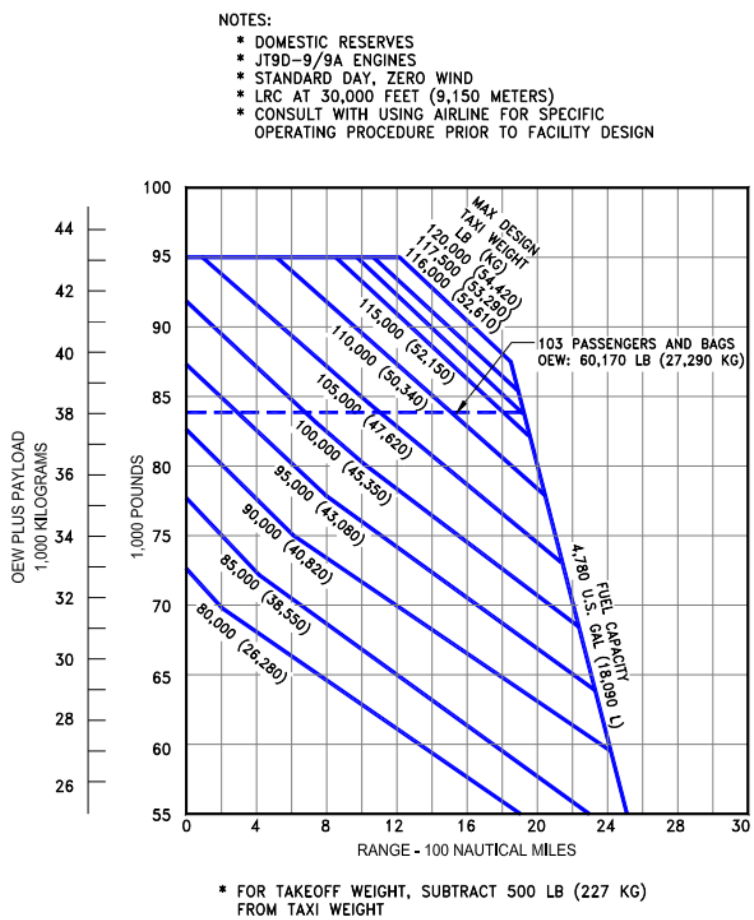


**Fig. 2 Boeing Document D6-58325-6, pp. 86: Payload/Range Diagram for a 737-100 with PW JT9D-9 Engines [16].**

Once all available aircraft parameters were recorded, the corresponding engines were investigated. For example, the 737 family has used many engines over the years. As shown in Fig. 1, the -200 alone has used nine different engines, and the 737 family consists of 13 commercial variants (-100 through -900 and the MAX -7 through -10). Every engine that has been used by the 737 family was documented in the database.

Engine TCDS are typically more comprehensive than aircraft TCDS. Most engine parameters are described in the data sheets. The secondary source of engine information used was Roux's *Turbofan and Turbojet Engines Database Handbook* [9, 10], where information such as fuel consumption, turbine entry temperature, and air mass flow rate can be found. In some cases, additional data was drawn from sources such as a 2008 gas turbine engine data table published in Aviation Weekly [17].

In cases where reliable data could not be found for a specific parameter, no value was recorded. However, thanks to the large sample size, missing entries did not significantly impact the overall analysis. This large collection of aircraft and engine data allowed for robust analysis, even when some entries were incomplete.

### B. Regression Overview

FAST utilizes historical regressions during aircraft sizing. GPRs were chosen for their adaptability to new data and dimensional flexibility. When using historical data for aircraft design, the number of known point performance metrics may vary between designs. Regressions without predetermined inputs are desirable, because they allow for input flexibility and can make use of all available data. For example, if an aircraft is being designed to carry 150 passengers over 2,000 nautical miles, a good initial guess for the aircraft's MTOW will expedite the weight iteration process. If little additional information is available, the regression can use just these two parameters (i.e., payload and range requirements) to predict an initial MTOW. However, if more information—such as wing loading—is available, including these inputs will allow the regression to capture more nuanced trends, leading to a more accurate prediction and faster convergence during aircraft sizing. This is illustrated in Fig. 3, where a notional aircraft, designed to carry 150 passengers 6,000 km, is shown to converge on an MTOW in less iterations with a more reasonable initial guess.

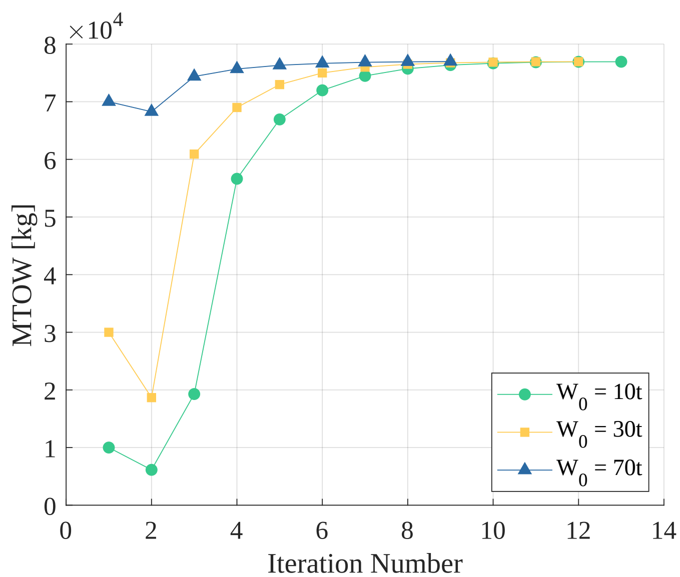


**Fig. 3 Notional aircraft MTOW convergence using different initial guesses.**

The regression employs a non-parametric GPR, which offers an advantage over a parametric GPR by not assuming a predetermined functional output form. This flexibility allows the regression to dynamically learn relationships between aircraft parameters, with the data and the covariance kernel determining the shape of the output space [18]. This capability is especially valuable in aircraft design, where relationships between design parameters are often nonlinear and influenced by multiple factors. Although both parametric and non-parametric GPRs scale computationally with $O(n^3)$, the term $n$ represents different quantities in each case. In parametric GPR, $n$ refers to the number of functional coefficients in the output form, while in non-parametric GPR, $n$ refers to the number of data points used for prediction. Given that the database contains several hundred aircraft, the computational cost of using a non-parametric GPR remains manageable.

Non-parametric GPR models the relationship between inputs and outputs by assuming that the data points and any subsets of them are jointly distributed according to a multivariate Gaussian distribution. In FAST, GPR is trained on a set of known input-output pairs from an aircraft database. This process encodes the relationships between the variables, allowing GPR to make probabilistic predictions for unknown outputs. Since the database contains over 100 variables for each aircraft, FAST first selects the relevant inputs and outputs for each specific regression task. It then creates a targeted subset of variables from the larger set, enabling the regression to focus on the desired relationships.

Non-parametric GPR assumes that the predicted output, given the inputs, follows a Gaussian distribution. As a result, the regression provides a Gaussian distribution for the output, defined by a mean and standard deviation. Equations 1 and 2 present these predicted values. Importantly, GPR does not assume that the input variables themselves follow a normal distribution.

$$\mu = \mu_0 + K(\bar{x}, x^*)\big(K(\bar{x}, \bar{x}) + \sigma^2 \times I\big)^{-1}(\bar{y} - \mu_0)^T \tag{1}$$

$$\Sigma = K(\bar{x}, \bar{x}) - K(\bar{x}, x^*)\big(K(\bar{x}, \bar{x}) + \sigma^2 \times I\big)^{-1} K(\bar{x}, x^*)^T \tag{2}$$

where:
- $\mu_0$ is the prior mean. It is a default assumption for an output variable if the data proves unhelpful.
- $\bar{x}$ is the set of input variable data for all aircraft used in the regression.
- $x^*$ is the set of input variable data corresponding to the unknown output value. It is the question being asked to the regression.
- $\sigma^2$ is the so-called "noise variance." It represents a distrust in the data collected. Analogously, in a physical system this variable captures the effect of noise introduced by a sensor or measuring equipment.
- $I$ is an identity matrix with size $N \times N$, where $N$ is the number of aircraft datasets used in the regression.
- $\bar{y}$ is the set of (known) output variable data for all the aircraft used in the regression.
- $K(x_1, x_2)$ is the covariance kernel. It is a measure of similarity between two sets of data (aircraft in this case), $x_1$ and $x_2$. This function and its tuning will be discussed in Section II.C.

### C. Covariance Kernel

While non-parametric GPR does not assume a specific functional form for the output, it does require a functional form for the covariance kernel. A covariance kernel is a measure of similarity between two data points, in this case aircraft, with respect to each parameter that defines them. The squared exponential kernel (also called the radial basis function) is widely regarded as a good choice for several reasons. A good kernel should effectively capture appropriate similarity between data points and be computationally efficient [19]. Moreover, the kernel should exhibit decaying similarity as data points points become sufficiently different [18, 20].

The squared exponential kernel satisfies these requirements and is adaptable to higher-dimensional spaces while maintaining a relatively low level of computational complexity. The $D-$dimensional form is presented in Equation 3.

$$K(\boldsymbol{x}, \boldsymbol{x^*}) = \tau \exp\left(-\gamma \sum_{i=1}^{D} \frac{(x_i - x_i^*)^2}{\ell_i^2}\right) \tag{3}$$

where:
- $\boldsymbol{x}$ and $\boldsymbol{x^*}$ are example data points. Each is an $D-$dimensional vector.
- $\tau$ is the output scale factor. It determines the absolute value of the impact that two "similar" data points will have on the posterior distribution.
- $\gamma$ is the input scale factor. It governs the sensitivity of the kernel to input similarities.
- $\ell_i$ is the parameter length scale. This takes unique values depending on the parameter being compared. It serves to normalize between parameters so they can be directly summed.

The squared exponential kernel fulfills all requirements for a good choice of covariance kernel. It compares all dimensions of two data points, normalizing each dimension such that all contribute equally. It decays to little contribution as two data points become dissimilar. In effect, this kernel will not modify a prior belief about the regression output if there does not exist data to claim that the prior belief is flawed. It scales computationally with $O(D)$, and in practice the dimensionality of a regression is limited to $D \sim 10$ as there are rarely aircraft design parameters with strong relationships to more than 10 other parameters without redundancy.

### D. Parameter and Hyperparameter Tuning

Thus far, many regression parameters and kernel hyperparameters have been introduced and defined. This section presents studies and discussions on tuning these parameters within the context of regression models. In FAST, the tuning methods are often parameterized based on the available data from the aircraft database, as the regressions are dynamic. Rather than optimizing parameters for a specific regression, tuning studies aim to minimize average errors across multiple regressions.

#### *1. Prior Mean*

The prior mean, shown by $\mu_0$ in Equation 1, represents a prior belief about the output of the regression, which may be adjusted depending on the data's relevance to the desired output parameter. This prior value can be based on an educated guess, a lower-fidelity numerical model, or a physics-based model. In FAST, since the regression does not have prior knowledge of which dataset corresponds to the output parameter, the prior mean is tuned to the mean of the desired output parameter (as shown in Equation 4). This approach provides the regression with a sense of scale, adjusting it according to the available data without requiring additional models. FAST does allow for an alternative prior to be input into a regression should a user decide they would like to overwrite the average value provided by the database.

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_i \tag{4}$$

where $\bar{y}$ is defined as it was in Equation 1, the set of output parameter values from all aircraft use in the regression.

#### *2. Length Scales and Output Scale Factor*

The length scale hyperparameter, denoted as $\ell$ in Equation 3, determines how similar two aircraft parameter values must be for the regression to treat them as "alike." This hyperparameter is highly sensitive to the specific aircraft parameter it applies to, and a distinct length scale is required for each dimension being compared. For example, when predicting operational empty weight (OEW) as a function of maximum takeoff weight (MTOW) and entry-into-service (EIS) year, the regression evaluates the MTOW, EIS, and OEW of aircraft in the database to estimate an appropriate OEW for the given MTOW and EIS. While comparisons between aircraft within 10,000 pounds of the target MTOW may be reasonable, comparisons to aircraft that entered service within 10,000 years of the target EIS would clearly be inappropriate.

There are two primary approaches to tuning the length scale parameter(s). One option is to perform empirical testing or optimization to determine the values that minimize error using a set of validation data. Alternatively, the parameters can be learned using automatic relevance determination (ARD). ARD requires that all available parameters in the database be used for every output, and the length scales are learned by tuning the input variables. Inputs with little relevance to the output (i.e., those showing weak correlation with the output) are naturally assigned near-zero contributions [21].

Since length scales are not universal across different regressions, optimization would require fixing the inputs and outputs of specific regressions, which undermines the flexibility of a dynamic regression model. While ARD could dynamically learn the length scales, this approach would be computationally expensive, negating the benefit of a quick regression call within the aircraft sizing code. However, the two desirable functions that ARD performs, parameter relevance and tuning, can be approximated as follows:

1) When performing a regression, it is assumed that not all input parameters are relevant. Either a select few are chosen based on known physical relationships between aircraft parameters, or a small subset of parameters is used and assumed to be relevant. The former method is employed for regressions with predefined outputs, such as OEW or propulsion system weight. The latter approach is used when the output is set dynamically based on user interaction. When a user inputs information describing an aircraft, these values are used to create a regression that predicts any missing data the user has not provided. In this case, required information such as payload and range is supplemented with optional data to predict the remaining inputs necessary for the sizing iteration.
2) Regarding the tuning of parameters deemed relevant, the goal is not to find the perfect value that minimizes error for a specific regression. Instead, the tuning value should be parameterized based on the data and tailored to the

**Table 2 Regression inputs and outputs for a $\gamma$ tuning parameter study, shown in Fig. 4**

| Output | Inputs |
|---|---|
| Aircraft Length | MTOW |
| OEW | MTOW, Range |
| Engine Weight Fraction | MTOW, Thrust |
| MTOW | Payload, Range, EIS |
| Lift-to-drag Ratio | Payload, Range, MTOW |
| Wing Loading | MTOW, EIS |

specific input variable. Using the standard deviation of the input variable provides a length scale for the kernel that is both appropriate in scale and intuitive. If two aircraft have input parameter values within one standard deviation of each other, it is reasonable to consider them "similar," as their difference is essentially negligible during the large data processing that the regressions employ.

Equation 5 illustrates how the length scale hyperparameters are tuned.

$$\ell_i^2 = \mathrm{var}(\bar{x}_i) \tag{5}$$

Since all inputs are normalized by their respective length scales, the kernel output will reflect changes on a consistent order of magnitude. To scale this to the appropriate output variable, an additional tuning parameter, denoted by $\tau$ in Equation 3, is required. The exponential term determines the similarity between two data points, while the output scale factor $\tau$ dictates the extent to which this similarity influences the adjustments made to the prior belief. Similar to the length scales, $\tau$ is tuned to one standard deviation of the output variable, which provides an appropriate scale for the kernel to modify the prior with. This is shown in Equation 6.

$$\tau = \sqrt{\mathrm{var}(\bar{y})} \tag{6}$$

*3. Noise Variance*

In a physical system, the noise in recorded data may be defined through testing sensors. However, in a simulated system, noise variance is treated as an additional hyperparameter and thus requires tuning as well. Despite being collected from reputable sources, it is ill-advised to assign a small universal value to the noise variance. In addition, the noise should be scaled depending on the regression's output parameter. During data collection, disagreements between sources for parameter values were not uncommon. For example, the range reported in a TCDS may differ from that in an Aircraft Planning Manual. TCDS figures typically reflect values at their extremes (i.e. do-not-exceed speeds or weights), whereas Aircraft Planning Manuals offer detailed diagrams that present multiple specific ranges under various conditions. Typically, these differences were on the order of about 5 to 10 percent. Therefore, it is assumed that the standard deviation of the data noise is 7.5% of the mean of the data for the desired output parameter.

*4. Input Scale Factor*

The input scale factor $\gamma$ is commonly accepted be on the order of $\gamma = 0.5$. This parameter is meant to be constant for any regression regardless of the input and output spaces. This was tested empirically by considering multiple regressions which are commonly used in aircraft design. The aircraft database was divided into two sets of data, 90% being used to train the model while 10% was reserved to test the regression output. The data was assigned to validation and training sets randomly. Six common regressions, summarized in Table 2, were then performed using the test data's inputs. This was repeated as the input scale factor $\gamma$ was varied. Then, a new 90-10% split was taken and the regressions were repeated (100 times) for each $\gamma$. All of the errors for each value of $\gamma$ were averaged for each regression. Additionally, the average of the six regression's error was taken at each gamma to confirm if there exists an optimum value of $\gamma$ such that the average error across all regressions is minimized. The results from this study are shown in Fig. 4, where it can be shown that for values of $0.2 < \gamma < 5$ the resultant average error across all regressions appears to be somewhat constant. A slight optimum was found at $\gamma = 2.27$, which is the value assumed to universally minimize the error output for any regression in FAST.

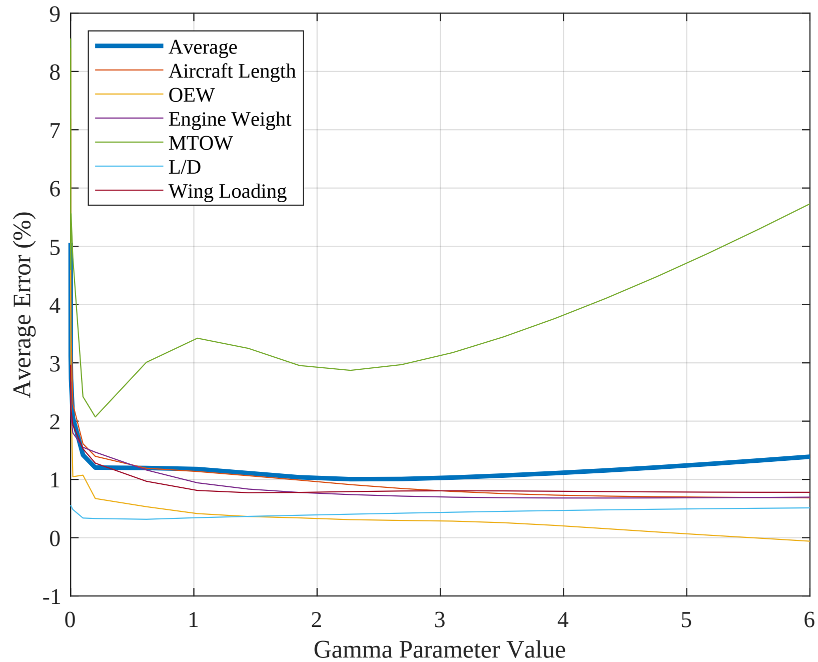


**Fig. 4 Example regression errors as function of the hyperparameter $\gamma$**

### E. Database Usage for Alternative Energy Aircraft

Although FAST's database consists solely of conventionally powered aircraft, the tool is built to design, analyze, and evaluate advanced aircraft with novel propulsion systems, including electrified aircraft propulsion [7]. Accordingly, when using conventional aircraft data to draw conclusions about hybrid-electric or fully electric concepts, modifications must be made to account for biases inherent in conventional regressions. For example, when predicting aircraft OEW, conventional data introduces biases about airframe shape, fuel tank locations and weights, material choices, and other factors. The following methodology, summarized here and described in detail by Arnson et al. [22], outlines an approach for modifying the inputs and outputs of a conventional OEW regression for use in conceptual design of electrified architectures.

1) Use existing conventional data to define a new parameter, referred to in this paper as "airframe weight." This parameter is defined in Equation 7 as:

$$\begin{aligned} W_{\text{Airframe}} &= W_{\text{MTO}} - W_{\text{Fuel}} - W_{\text{Payload}} - W_{\text{Crew}} - W_{\text{Engines}} - W_{\text{Other}} \\ &= W_{\text{OEW}} - W_{\text{Engines}} \end{aligned} \tag{7}$$

For unconventional designs, additional terms may be added to account for components such as fuel cells, batteries, or electric motors.

2) Perform a regression on $W_{\text{Airframe}}$ using input parameters that are not specific to conventional aircraft. Examples include maximum takeoff weight, aircraft thrust-to-weight ratio, aircraft range, payload weight, etc.
3) If necessary, perform post-processing to the regression output. For example, if liquid hydrogen is intended to replace kerosene, modify airframe weight by subtracting conventional tank weight and adding a hydrogen tank weight. (Since FAST's database does not include component weights, a regression on conventional architecture fuel tanks can be found in Raymer (1989) [23], while estimates for a hydrogen fuel tank weight can be found using fuel weight and an assumed gravimetric efficiency).
4) Build back up an OEW from the resultant airframe weight and estimates for any components relevant to the desired architecture. This includes gas turbines, fuel cells, and electric motors. Conventional OEW buildups

include only power and thrust sources, not energy sources. However, some sources may include batteries as part of an OEW buildup.

FAST uses this procedure for all aircraft sizing. Conventional architectures have their propulsive and structural OEW components split as well. This allows for easy handling in the OEW prediction in FAST. The user-intput architecture matrices, described in detail in Mokotoff et al. [7], tell FAST which propulsion models to run. These could be gas turbine, electric motor, fuel cell, etc. The weight estimations for these components are output if the model is called, otherwise their weight is kept as its initial value (zero) and it does not impact the sizing process.

# III. Regressions Used in FAST

FAST utilizes several regressions to initialize an aircraft, as well as during the sizing process. Figure 5 outlines FAST's workflow, and highlights the code blocks which call regressions in green, and can be further distinguished with an "R" character. The initialization block calls regressions to set values for unspecified variables which FAST requires to run. For example, if a user does not set an initial MTOW guess, FAST will use required inputs, such as payload weight and range, to predict these unknowns. The initialization regressions are not updated as the aircraft sizes. In contrast, the regressions in the weight buildup, which predict "airframe weight" (see Sec II.E) and engine weight. These regressions are run inside the sizing iteration, updating their predictions depending on the current iteration's MTOW, thrust, etc. The regression inputs and outputs are described in detail as follows:

**Initialization:** FAST parses user specified values from the following list of parameters:

- Cruise Speed (if unspecified in mission profile)
- Cruise Altitude (if unspecified in mission profile)
- MTOW (initial guess)
- Wing loading
- Lift-to-drag ratio (turbofans only; insufficient turboprop data)
- Thrust-to-weight ratio (turbofans) / Power-to-weight (turboprops)
- SLS thrust (turbofans) / SLS power (turboprops)

Then the known values, in combination with required inputs, payload weight and range, are used to predict the unspecified values.

**Airframe Weight:** Turbofan airframe weight is predicted using wing area, MTOW, SLS thrust, and entry-to-service year. There is insufficient data to perform a probabilistic regression for turboprops, instead a linear fit to airframe weight as a function of MTOW is fit using the sparse data.

**Engine Weight:** Predicted using aircraft thrust or power at takeoff, depending on whether the aircraft uses turbofan or turboprop engines.

FAST is an actively developed tool, and is constantly being improved. As more data is collected, the regressions FAST uses are updated. At present, FAST does not use input/output variables identical to those presented in Sec. IV for its regressions. As part of its development, the input/output variables listed above change as improvements are made. This section described the variables as they currently exist in FAST.

# IV. Results

## A. Historical Database

The historical database is available online, free of charge, in several formats. Large Excel spreadsheets are available for use with statistical software, in addition to MATLAB data structures which contain the same information. These data structures, which include both engine and aircraft data, are formatted to integrate seamlessly with FAST, allowing designers to create their own regressions. The database [6] can be accessed at https://github.com/ideas-um/FAST.

## B. Regressions

The FAST tool is capable of creating over $10^{29}$ unique regressions using the information in the database. Of these, the most useful regressions utilize aircraft performance metrics, such as design range and payload, to predict values that are not intuitive or known early in the conceptual design. This section provides three examples (OEW, engine weight, and aircraft length) which are used within FAST during the aircraft sizing process. The regression outputs from FAST are compared to reference regressions from contemporary literature, and their errors are evaluated. The data

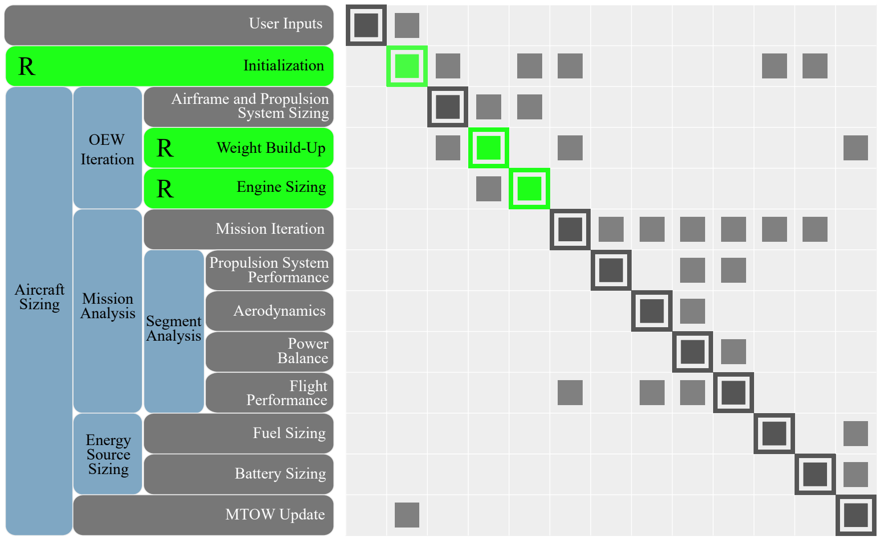


**Fig. 5 FAST's overarching workflow, from Mokotoff et al. [7], with regression-using blocks highlighted in green and marked with "R."**

used for the studies presented in this section come from commercial aircraft utilizing turbofan engines only, and is sourced directly from FAST's database. Appendix V.A provides a comprehensive tutorial for recreating the regressions presented in this section.

Similar to the study presented in Section II.D.4, each regression is evaluated using 90% of the data for training and 10% for testing. The data is then randomized to produce a new set of training and test data, where a new regression is created and evaluated, and this process is repeated 100 times. The results from these tests are shown as cyan circles in Figs. 6- 14, and is referred to as "FAST 1" in the following section.

In addition to the 90/10 splits, another FAST regression is performed using 100% of the data for both training and testing. This approach mirrors the potential overlap between training and test data in reference regressions, where the original data used to create these models is not explicitly known. As a result, some of the data in FAST's database may overlap with the data used in the creation of these reference regressions. To ensure a fair comparison, the results from this 100% data regression are also included and are shown as black circles in Figs. 6- 14, and is referred to as "FAST 2" in the following section. It is of note that because the selection of training versus validation data in this study is random, results may vary slightly if a different random number generation seed is used. For those interested in reproducing these results *exactly*, the seeds are documented and available upon request.

*1. Operating Empty Weight Regression*

OEW is an important parameter in conceptual aircraft design and its estimation as a function of other aircraft parameters is essential in FAST during the sizing process. Canonical aircraft design literature offers data-driven regressions to predict OEW such as Raymer [24], Jenkinson [25], and Roskam [26], as shown in Equations 8, 9, and 10 respectively. These reference regressions primarily predict OEW based on MTOW, and in one case, the number of engines.

In contrast, the FAST regression captures more nuanced fluctuations in OEW by incorporating additional parameters such as MTOW, payload, range, SLS thrust, and the number of engines. By allowing for more inputs, the FAST regression identifies trends in OEW that are not visible when only considering MTOW and the number of engines.

The outputs from the five regressions were compared to the actual OEW values in the database. The results of FAST, Raymer, Jenkinson, and Roskam are shown in Figs. 6, 7, 8, and 9 respectively, with a summary of errors provided in Table 3. The mean error indicates a tendency to over or underestimate a parameter value in general. Most regressions have mean errors close to zero, showing low bias overall. Raymer (-8.104%) has the highest negative bias, meaning it tends to underestimate. Median errors are also close to zero for most regressions, suggesting that the typical error is minimal. Raymer (-8.453%) again shows the most negative deviation from zero, with Roskam being the second most deviated, indicating some bias in these regressions. The standard deviation measures the variability of errors around

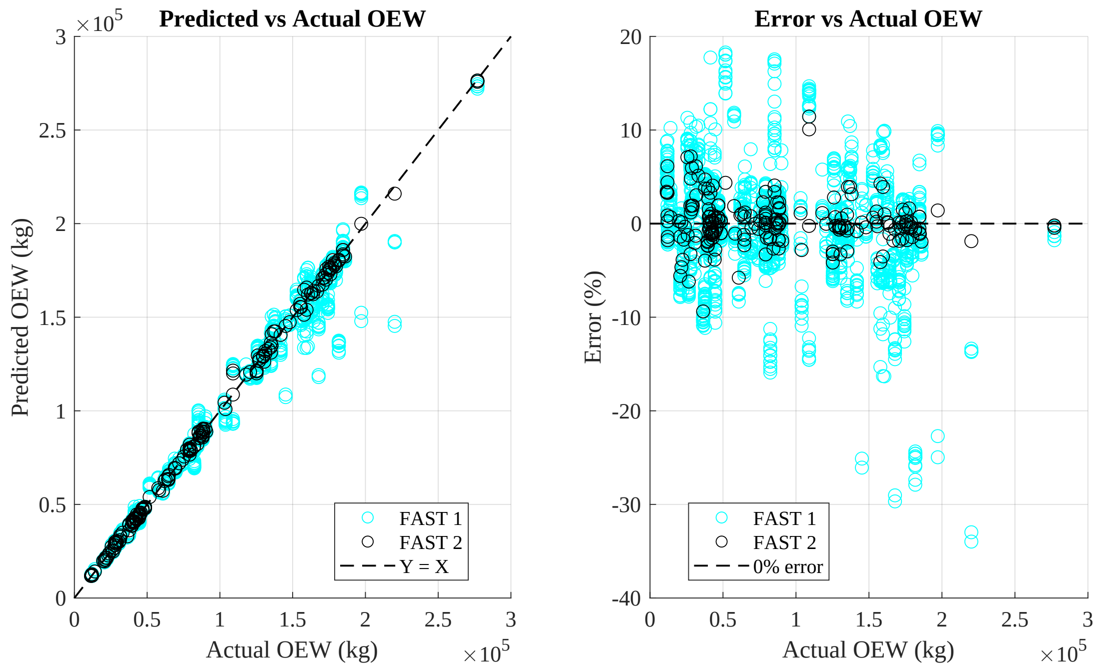


**Fig. 6 FAST OEW regression outputs: predicted vs. actual and error vs. actual**

the mean, with smaller values indicating less error deviation. Lower standard deviations are preferred as they indicate more predictable errors. All regressions show moderate to high variability, with Jenkinson having the highest standard deviation (18.71%), indicating less consistent predictions. FAST 2 has the lowest (7.653%), suggesting relatively stable predictions. Most regressions have skewness values close to zero, indicating a roughly symmetric error distribution. Raymer has slight negative skew, indicating a few larger positive errors, while FAST 1 and FAST 2 show slight positive skew. The kurtosis values suggest moderate to heavy-tailed distributions. FAST 2 shows the highest kurtosis (6.229), indicating a higher likelihood of extreme outliers. Roskam's kurtosis is closest to 3, suggesting a nearly normal-tailed distribution with fewer extreme outliers. Among the reference regressions, the Jenkinson regression [25] yields the lowest mean error. This is likely because it uses an average OEW-to-MTOW fraction. This averaging naturally reduces the mean error as more data is included. However, using an average OEW-to-MTOW fraction does not necessarily result in the lowest standard deviation or the lowest kurtosis error distribution compared to the other regressions.

Both of FAST's regressions outperform the reference regressions in terms of standard deviation and skewness of the error distribution. FAST 1, which represents predictions for aircraft *not* stored in the database, has a slightly negative bias of -0.2741%, but achieves approximately 40% lower standard deviation than the next best regression. Additionally, FAST 1 has the smallest difference between the mean and median of the error distribution, indicating the least skewed distribution among the regressions.

FAST 2 outperforms FAST 1 because it uses more training data and is predicting parameters for aircraft it is already familiar with. It has means and medians closer to zero with the smallest standard deviation of all the regressions.

$$W_{\text{empty}} = 1.02 W_{\text{takeoff}}^{0.0833} \tag{8}$$

$$W_{\text{empty}} = \begin{cases} 0.55 W_{\text{Takeoff}} & N_{\text{Engines}} = 2 \\ 0.47 W_{\text{Takeoff}} & N_{\text{Engines}} > 2 \end{cases} \tag{9}$$

$$W_{\text{empty}} = 10^{-0.0802} W_{\text{Takeoff}}^{-0.0369} \tag{10}$$

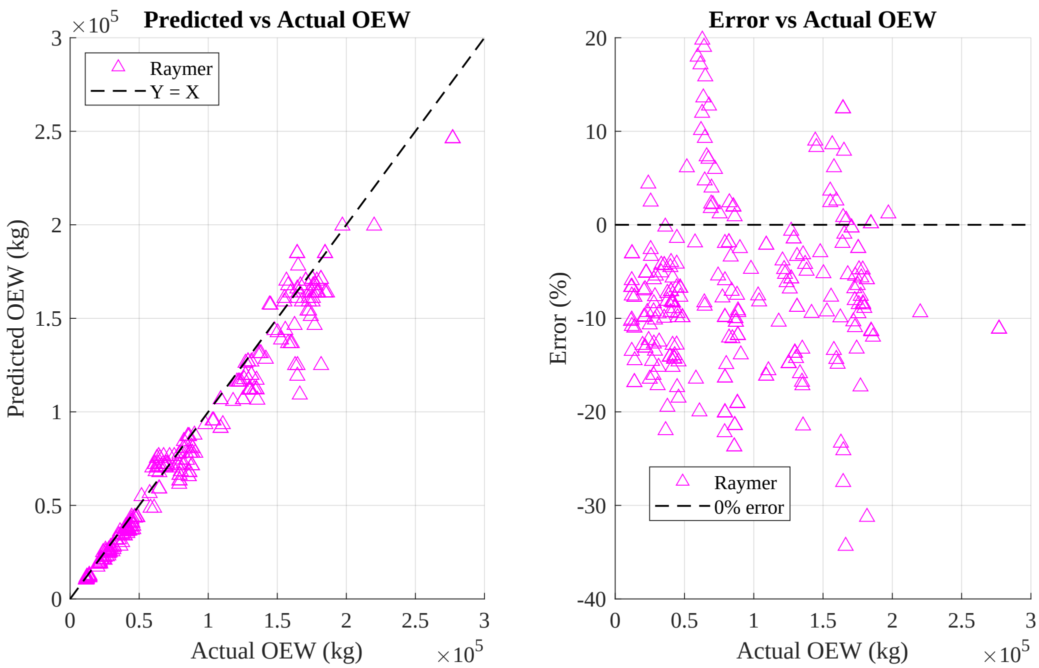


**Fig. 7 Raymer [24] OEW regression outputs: predicted vs. actual and error vs. actual**

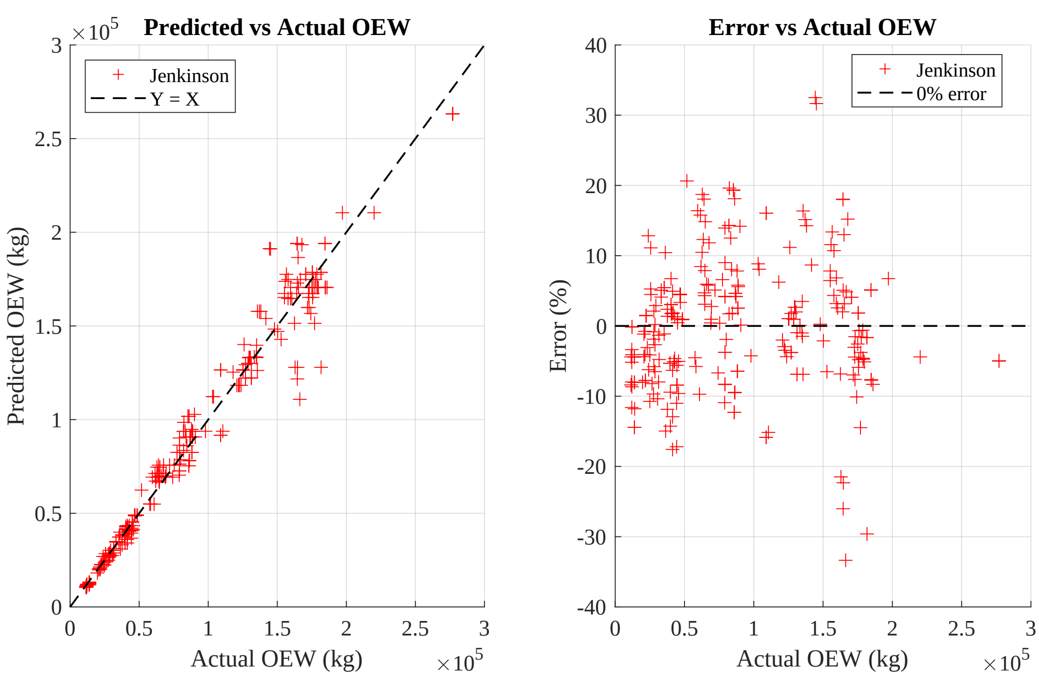


**Fig. 8 Jenkinson [25] OEW regression outputs: predicted vs. actual and error vs. actual**

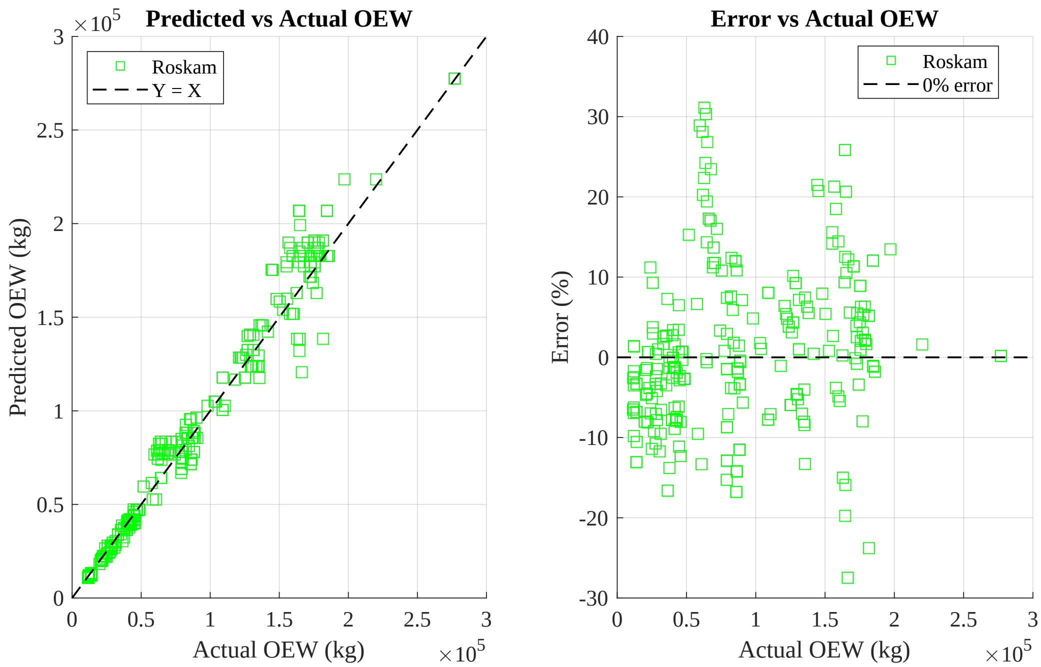


**Fig. 9 Roskam [26] OEW regression outputs: predicted vs. actual and error vs. actual**

**Table 3 Error metric summary for FAST and reference regressions on OEW.**

| Error Metric | Mean [%] | Median [%] | Standard Deviation [%] | Skewness | Kurtosis |
|---|---|---|---|---|---|
| FAST 1 | -0.2741 | -0.3064 | 5.047 | -0.8503 | 9.656 |
| FAST 2 | 0.1072 | -0.0260 | 2.398 | 0.7471 | 6.915 |
| Raymer [24] | -8.104 | -8.453 | 8.331 | 0.6310 | 4.370 |
| Jenkinson [25] | $-3.887 \times 10^{-4}$ | 0.9278 | 8.798 | 0.1359 | 4.572 |
| Roskam [26] | 0.2523 | -1.176 | 9.551 | 0.7269 | 3.964 |

*2. Engine Weight Regression*

As seen in Eq. 7, conventional engine weight is necessary to predict when estimating OEW for an alternative energy aircraft. In this section, the FAST engine weight regressions are compared to those seen in the canonical literature. Regressions from Raymer [27], Jenkinson [4] , and Svoboda [28] are shown in Equations 11, 12, and 13 respectively. The engine weight regressions for FAST, Raymer, Jenkinson, and Svoboda are shown in in Figs. 10, 11, 12, and 13 respectively. The error summary can be seen seen in Tab. 4. FAST regressions have low positive mean errors, indicating minimal bias in prediction, whereas Raymer and Jenkinson have significant negative means (-36.96% and -26.23%, respectively), showing substantial underestimation. The median is close to the mean for each model, reinforcing the general bias trends, with Raymer and Jenkinson having large negative medians. All regressions show moderate to high variability, with Jenkinson having the highest standard deviation (18.71%), indicating less consistent predictions. FAST 2 has the lowest (7.653%), suggesting relatively stable predictions. Most regressions have skewness close to zero, indicating a roughly symmetric error distribution. Raymer has slight negative skew, meaning a few larger positive errors, while FAST 1 and FAST 2 have slight positive skew. FAST 2 shows the highest kurtosis value indicating a heavy tailed distribution with some extreme outliers. As compared to other regressions, the large kurtosis and lower standard deviation of the FAST regressions suggests that generally, FAST will predict better with lower errors, however when it does make a mistake, the error is likely to be higher than that of other regressions.

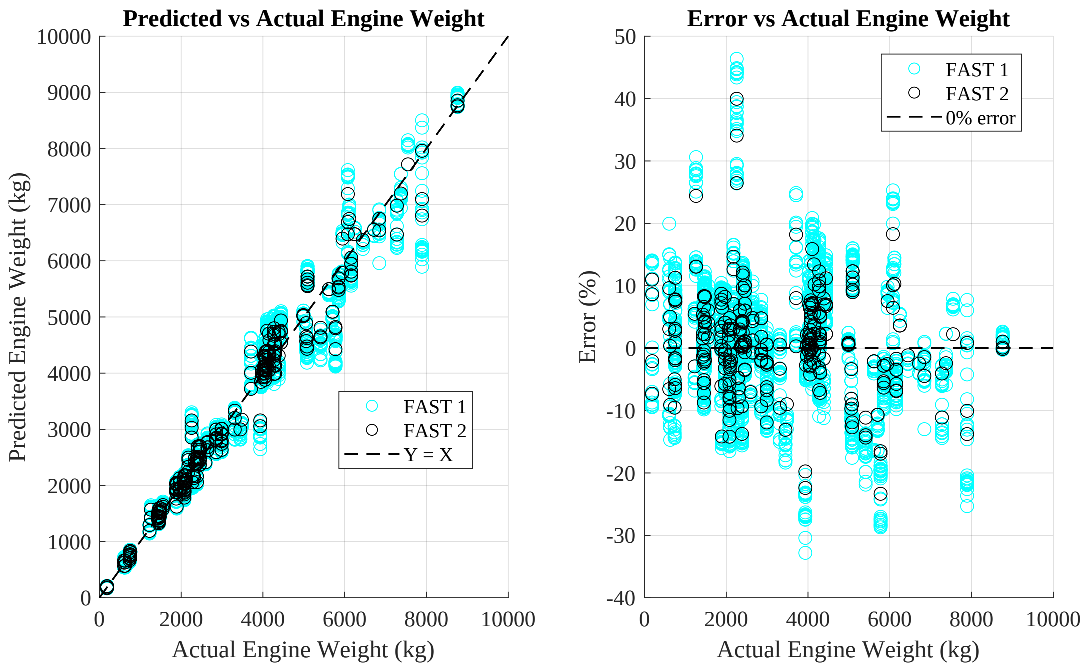


**Fig. 10 FAST OEW regression outputs: predicted vs. actual and error vs. actual**

$$W_{\text{Engine}}(\text{kg}) = 14.7 T_{\text{Takeoff}}(\text{kN}) 1.1 e^{-0.045BPR} \tag{11}$$

$$W_{\text{Engine}}(\text{kg}) = (8.7 + 1.14BPR) * T_{\text{Takeoff}}(\text{kN}) \tag{12}$$

$$W_{\text{Engine}}(\text{lbs}) = 250 + 0.175 T_{\text{Takeoff}}(\text{lbf}) \tag{13}$$

**Table 4 Error metric summary for FAST and reference regressions on engine weight.**

| Error Metric | Mean [%] | Median [%] | Standard Deviation [%] | Skewness | Kurtosis |
|---|---|---|---|---|---|
| FAST 1 | 0.66120 | 0.6006 | 9.185 | 0.2807 | 5.371 |
| FAST 2 | 0.7404 | 0.5805 | 7.653 | 0.5393 | 6.229 |
| Raymer [27] | -36.96 | -36.00 | 10.24 | -0.5207 | 3.215 |
| Jenkinson [4] | -26.23 | -25.13 | 18.71 | -0.0547 | 2.667 |
| Svoboda [28] | -7.150 | -6.539 | 11.64 | -0.1715 | 2.623 |

*3. Fuselage Length Regression*

Since FAST can generate thousands of different regressions, it has the ability to unveil relationships between various aircraft parameters. One application of this is FAST's visualization feature, which requires an accurate prediction of total aircraft length. A regression to predict aircraft length based on MTOW, range, payload, and wing area is shown in Fig. 14. This section does not include comparisons to historical references as it was not found in those references. The fuselage length regression is an example of a new regression that FAST can create based on data already collected in the FAST database.

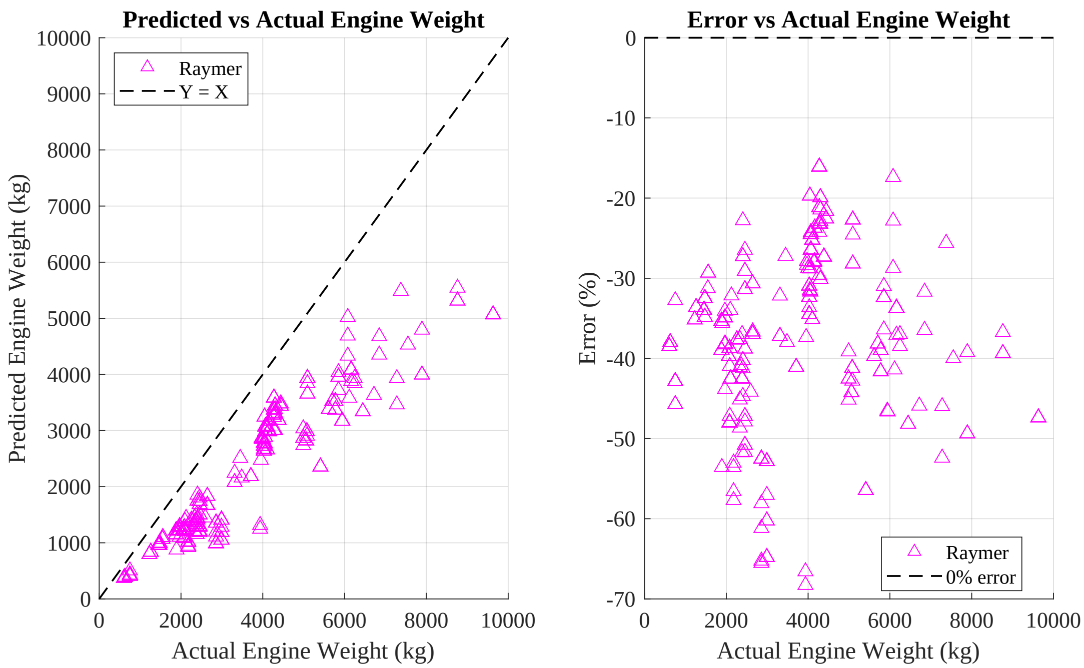


**Fig. 11 Raymer [27] engine weight regression outputs: predicted vs. actual and error vs. actual**

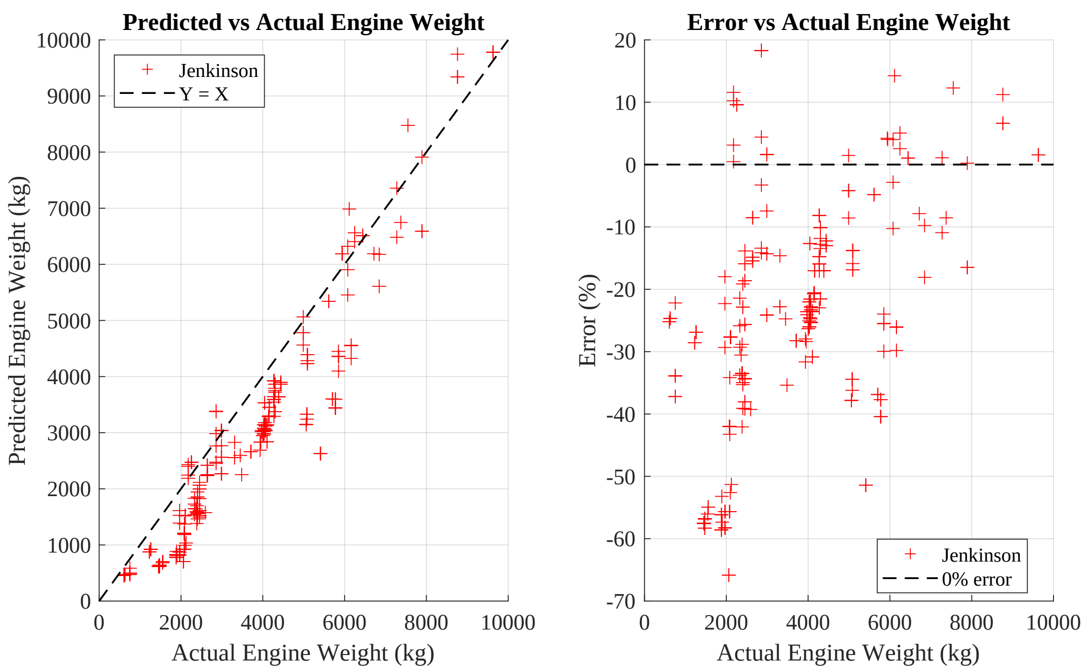


**Fig. 12 Jenkinson [4] engine weight regression outputs: predicted vs. actual and error vs. actual**

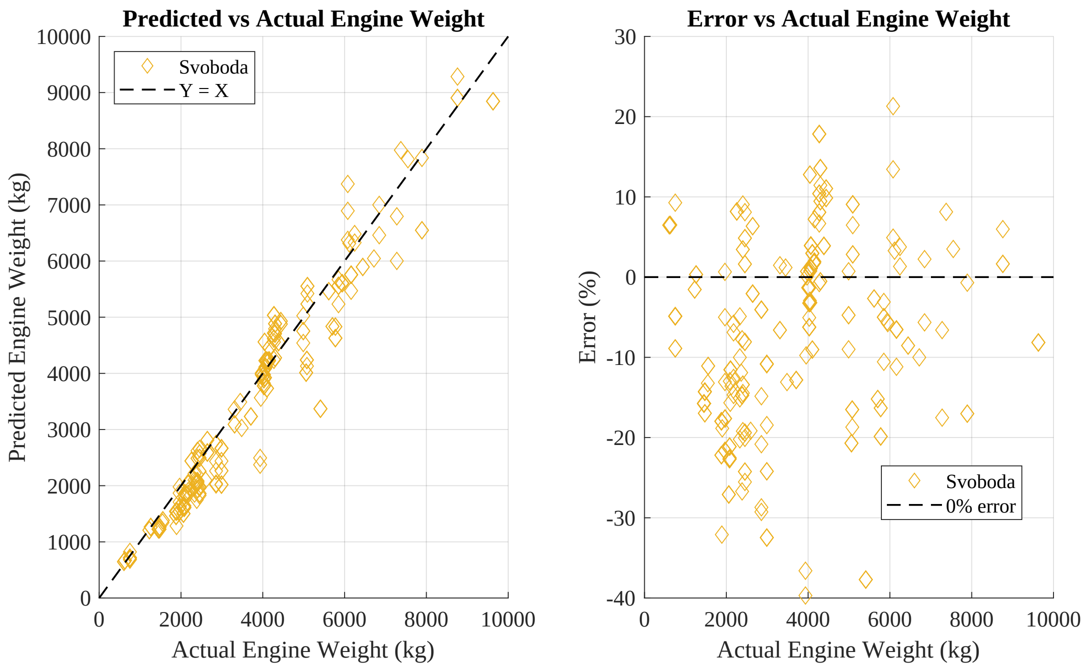


**Fig. 13 Svoboda [28] engine weight regression outputs: predicted vs. actual and error vs. actual**

One notable outlier in the dataset is the Bombardier Challenger 300, which consistently shows high errors in both training and test datasets. This is likely because it is the lightest aircraft in the database with a recorded length. The regression erroneously assumes that this aircraft must have a higher length since all the other aircraft are longer. Additional data on lighter aircraft, such as business jets, would likely resolve this issue by providing better representation across different weight categories.

Despite this outlier, the error metrics shown in Table 5 indicate that the predicted aircraft lengths, presented in Fig. 14, have error distributions similar to those observed in the OEW, shown in Figs. 6- 9, and engine weight regressions, shown in Figs. 10 - 13. Both FAST regressions have low mean errors (0.4938% for FAST 1 and 0.2555% for FAST 2), indicating minimal bias. The median values are close to zero, further indicating minimal bias in the central tendency of errors. FAST 2 has a lower standard deviation (5.167%) compared to FAST 1 (6.791%), suggesting more consistent error predictions. Both regressions show positive skew, with FAST 1 being slightly more skewed (0.5759), indicating occasional larger negative errors. High kurtosis is shown in both regressions, particularly FAST 1 (7.607), indicating a greater likelihood of outliers.

**Table 5 Error metric summary for FAST on aircraft length.**

| Error Metric | Mean [%] | Median [%] | Standard Deviation [%] | Skewness | Kurtosis |
|---|---|---|---|---|---|
| FAST 1 | 0.4938 | 0.02595 | 6.791 | 0.5759 | 7.607 |
| FAST 2 | 0.2555 | 0.03376 | 5.167 | 0.2669 | 4.213 |

## V. Conclusions and Future Work

This paper presents a novel approach to predicting key aircraft design parameters by leveraging Gaussian Process Regressions applied to a large, open-source historical database of aircraft and engine specifications. This database, comprising over 400 aircraft and 200 engines, is made freely available through the Future Aircraft Sizing Tool (FAST), offering the aviation design community an accessible, valuable resource for conceptual aircraft design.

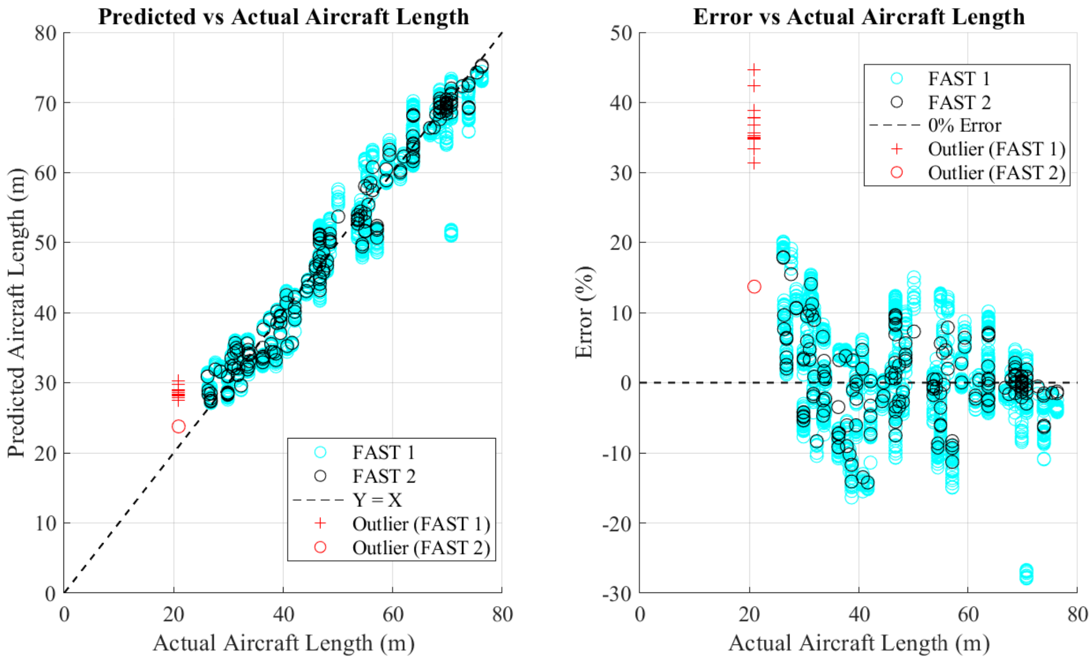


**Fig. 14 FAST aircraft length regression outputs: predicted vs. actual and error vs. actual. Outlier (Challenger 300) shown in red.**

The research demonstrates that the GPR models, integrated with FAST's database, provide significant improvements over reference regressions from canonical literature. These models, which incorporate a broader range of input parameters—such as MTOW, payload, range, and SLS thrust—enable more accurate predictions of critical parameters like OEW and engine weight. The results show that FAST's regression models not only reduce prediction errors but also deliver tighter error distributions and smaller standard deviations compared to existing methods, providing aircraft designers with more reliable estimates early in the conceptual design phase.

A key contribution of this work is the flexibility of FAST's non-parametric GPR framework, which allows for dynamic regression generation without predefined input constraints. This capability enables aircraft designers to explore a wide range of design configurations with precision, demonstrating the tool's adaptability across different aircraft and engine architectures. Additionally, this paper introduced a methodology to adapt conventional data for use in hybrid-electric and fully electric aircraft designs. By modifying conventional regressions, this approach ensures that historical data can still be leveraged to predict parameters for novel propulsion systems, enabling the design of sustainable, electrified aircraft concepts.

The development of the database and regression tools is part of an ongoing effort to continually refine and expand the FAST platform. Future work will focus on incorporating additional aircraft and engine data as they become available, and further improving the accuracy of the regression models as new user-contributed data enhances the database. Moreover, the flexibility of FAST allows it to be adapted for various aircraft configurations and propulsion technologies, including the integration of alternative fuels and hybrid-electric systems. As the tool evolves, it is expected to provide even more robust capabilities for the conceptual design of next-generation aircraft.

# Appendix

### A. FAST Regression Tutorial

Figure 15 shows a very simple example of a user specified regression call in FAST. This example illustrates a prediction for OEW (kg) based on a known MTOW (kg) and aircraft range (m). FAST aircraft are stored in a data structure, so each parameter on lines 2-4 are the paths through the structure to the location of the desired parameter. FAST will always treat the last entry of the Input/Output cell array as the output. Input order does not matter so long as the Target array matches the variable order, shown in line 8. Line 8 declares the input parameter values that are known by a user, and as mentioned the order must match that of the Input. In this example, since MTOW is declared as the first input while Range is declared as the second, and MTOW = 50,000 kg and Range = 3,000 km,

`Target = [5,000 , 3,000e3]` is correct, while

`Target = [3,000e3 , 5,000]` would be *incorrect*.

With the variables and their values declared, the next step is to load the FAST database, shown in line 11. The database stores 6 structures: turbofan aircraft, turbofan engines, turboprop aircraft, turboprop engines, fan reference, and propeller reference. The latter two mimic the structure of the others such that the format of the data structures can be explored and the variable paths (lines 2-4 in the example) can be known. The reference structures contain units in place of the numerical values that are stored in the database, so targets (line 8) can be entered in the correct units. The engine database contains all engines which are used by the aircraft in the aircraft database. The same information for each engine is stored within `Aircraft.Specs.Propulsion.Engine`. The engine databases are included separately because a user may wish to run a regression on an engine parameter (such as OPR as a function of EIS), and using the aircraft database would be biased towards engines which are used more frequently amongst different types of aircraft.

With all required input variables initialized, the regression can be run, as shown on line 15. Outputs of the regression function are the mean and variance of the normal distribution, which the output parameter is assumed to take. In FAST, the mean is always taken as the value the regression predicts while the variance is used as a measure of confidence the regression has in its prediction. Two additional inputs, weights and prior, can be included after a target in the regression call. Both are options to manually tune the regression and are not used in FAST. Weights adds a scaling factor to each input variable when summed in the covariance matrix as described in Sec. II.D. In this case, replacing line 15 with

`[Mu, Sigma2] = RegressionPkg.NLGPR(TurbofanAC, InputOutput, Target, [2 , 1]);`

would tell the regression that MTOW is twice as important as range when predicting OEW. Finally, a prior can be set which overwrites the FAST default of the database average. If a user believed the OEW in this example should be 20,000 kg, but wanted the database to modify that guess using historical data, they would replace line 15 with

`[Mu, Sigma2] = RegressionPkg.NLGPR(TurbofanAC, InputOutput, Target, [1 , 1], 20e3);`

Finally, the regression is capable of making multiple predictions in a single call. The only change to the inputs would be in Target and Prior (if Prior was set). Both of the variables would get extended to be multiple *rows* larger. For example, if a user wanted to additionally predict the OEW at an MTOW of 40,000 kg and 2,500 km, line 8 would be replaced with

`Target = [50e3, 3000e3; 40e3, 2500e3];`

whereupon the outputs would be $2 \times 1$ sized vectors as opposed to scalars that the example yields.

```
1  % Set Input/Output Variables (Cell array composed of string vectors)
2  InputOutput = {["Specs", "Weight",      "MTOW" ],...
3                 ["Specs", "Performance", "Range"],...
4                 ["Specs", "Weight",      "OEW"  ]};
5
6  % Set known input space values (double vector)
7  %          MTOW      Range
8  Target = [50e3,      3000e3];
9
10 % Load database (TurbofanAC variable will be loaded)
11 load('+DatabasePkg/IDEAS_DB.mat')
12
13 % Call the regression, predicting a mean and variance for the distribution
14 % of OEWs at the given Target vector
15 [Mu, Sigma2] = RegressionPkg.NLGPR(TurbofanAC, InputOutput, Target);
```

**Fig. 15 Example single regression call in FAST, formatted using Korn [29].**

# Funding Sources

This work is sponsored by the NASA Aeronautics Research Mission Directorate and Electrified Powertrain Flight Demonstration project, “Development of a Parametrically Driven Electrified Aircraft Design and Optimization Tool”, WO-0238. Karin Bozak serves as NASA’s Technical Monitor.

# Acknowledgments

The authors would like to thank Ralph Jansen, Amy Chicatelli, Andrew Meade, Karin Bozak, Noah Listgarten, Dennis Rohn, and Gaudy Bezos-O’Connor from NASA’s Electrified Powertrain Flight Demonstration project for supporting this work and providing valuable technical input and feedback throughout the duration of the project. The authors also thank Huseyin Acar, Benjamin Bates, Nawa Khailany, Paul Mokotoff, Janki Patel, Joaquin Rey, Jayda Shine, and Michael Tsai for their contributions to either FAST or the historical aircraft database that the regressions and projections rely on.

# References

[1] Roskam, J., *Airplane Design: Preliminary Sizing of Airplanes*, DAR Corporation, 1985.

[2] Raymer, D., *Aircraft Design: A Conceptual Approach, Sixth Edition*, American Institute of Aeronautics and Astronautics, Inc., 2018. https://doi.org/10.2514/4.104909.

[3] International Civil Aviation Organization, “Report on the Feasibility of a Long-Term Aspirational Goal (LTAG) for International Civil Aviation $CO_2$ Emission Reductions,” *ICAO Environment*, 2024. URL https://www.icao.int/environmental-protection/LTAG/Pages/LTAGreport.aspx.

[4] Jenkinson, L., Simpkin, P., and Rhodes, D., *Civil Jet Aircraft Design*, American Institute of Aeronautics and Astronautics, Inc., 1999. https://doi.org/10.2514/4.473500.

[5] Cinar, G., Cai, Y., Bendarkar, M. V., Burrell, A. I., Denney, R. K., and Mavris, D. N., “System Analysis and Design Space Exploration of Regional Aircraft with Electrified Powertrains,” *Journal of Aircraft*, Vol. 60, No. 2, 2023, p. 382–409. https://doi.org/10.2514/1.c036919.

[6] IDEAS Laboratory, “Future Aircraft Sizing Tool (FAST),” *Github*, 2024. URL https://github.com/ideas-um/FAST.

[7] Mokotoff, P., Arnson, M., and Cinar, G., “FAST: A Future Aircraft Sizing Tool for Electrified Aircraft Design,” AIAA SciTech Forum, 2025.

[8] Services, J. I., *Janes All the World’s Aircraft: In Service*, 2023.

[9] Roux, E., *Turbofan and Turbojet Engines: Database Handbook*, Éditions Élodie Roux, 2007.

[10] Roux, E., *Turboshaft, Turboprop and Propfan: Database Handbook*, Éditions Élodie Roux, 2011.

[11] Federal Aviation Administration, “Aircraft Characteristics Data,” retrieved from FAA, https://www.faa.gov/airports/engineering/aircraft_char_database/data, 2023.

[12] Eurocontrol, “Aircraft Performance Database,” https://contentzone.eurocontrol.int/aircraftperformance/default.aspx, n.d.

[13] Jenkinson, L., Simpkin, P., and Rhodes, D., “Civil Jet Aircraft Design Data Sets,” , 2001.

[14] Meier, N., “Jet Engine Specification Database,” retrived from https://www.jet-engine.net/, Sep 2021.

[15] Federal Aviation Administration, “Type Certificate Data Sheet A16WE, Revision 66,” , 2020.

[16] Boeing, “737-100 through 737-500 Airplane Characteristics for Airport Planning, Revision D,” , 2023.

[17] “Gas Turbine Engines,” *Aviation Week and Space Technology*, 2008.

[18] Schulz, E., Speekenbrink, M., and Krause, A., “A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions,” *Journal of Mathematical Psychology*, Vol. 85, 2018, pp. 1–16. https://doi.org/https://doi.org/10.1016/j.jmp.2018.03.001.

[19] Shawe-Taylor, J., and Cristianini, N., *Basic kernels and kernel types*, Cambridge University Press, 2004, p. 291–326.

[20] MacKay, D. J. C., "Introduction to Gaussian processes," 1998. URL https://api.semanticscholar.org/CorpusID:116281095.

[21] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning series, MIT Press, London, England, 2005.

[22] Arnson, M. G., Cinar, G., Waddington, E., Ansell, P., Clarke, M., de Vries, R., Salucci, F., Prabhakar, N., Gladin, J. C., Shi, M., Lovelace, E., and de Bock, P., "System-Level Energy Pack Requirements for Sustainable Commercial Aviation," *AIAA AVIATION FORUM AND ASCEND 2024*, American Institute of Aeronautics and Astronautics, 2024, p. 3828. https://doi.org/10.2514/6.2024-3828.

[23] Raymer, D., of Aeronautics, A. I., and Astronautics, *Aircraft Design: A Conceptual Approach*, AIAA Textbook Series, American Institute of Aeronautics and Astronautics, 1989. URL https://books.google.com/books?id=Q9QeAQAAIAAJ.

[24] Raymer, D., *Aircraft Design: A Conceptual Approach*, No. v. 1 in AIAA education series, American Institute of Aeronautics and Astronautics, 2006. URL https://books.google.com/books?id=Uf2yQgAACAAJ.

[25] Jenkinson, L. R., and Marchman, J. F., *Aircraft Design Projects for Engineering Students*, Butterworth-Heinemann, 2003.

[26] Roskam, J., *Airplane Design: Preliminary Sizing of Airplanes*, DAR Corporation, 1985.

[27] Raymer, D., of Aeronautics, A. I., and Astronautics, *Aircraft design: a conceptual approach, 3rd ed.*, 3rd ed., AIAA Textbook Series, American Institute of Aeronautics and Astronautics, 1999. URL https://discovery.hw.ac.uk/primo-explore/fulldisplay/44hwa_alma2128621720003206/44HWA_V1.

[28] Svoboda, C., "Turbofan engine database as a preliminary design tool," *Aircraft Design*, Vol. 3, No. 1, 2000, pp. 17–31. https://doi.org/https://doi.org/10.1016/S1369-8869(99)00021-X.

[29] Korn, F., "mcode.sty Demo," https://www.overleaf.com/latex/templates/highlighting-matlab-code-in-latex-with-mcode/nhtksndnsmmx.pdf, 2014.