Evaluation of Response Surface Experiment Designs for Distributed Propulsion Aircraft Aero-Propulsive Modeling

Benjamin M. Simmons*
NASA Langley Research Center, Hampton, Virginia 23681

Modern distributed hybrid and electric propulsion aircraft, including vertical, short, and conventional takeoff and landing configurations, exhibit significant aero-propulsive complexity and a large number of interacting test factors. This paper presents the development and evaluation of experiment designs for aero-propulsive characterization of distributed propulsion aircraft. Five different foundational response surface designs are evaluated to inform the development of two sequential design approaches tailored to complex aircraft aerodynamic characterization experiments. The first approach, which builds on sequential face-centered central composite designs, has been used previously to develop aero-propulsive models for complex aircraft using wind tunnel testing. The second approach is a new design strategy leveraging a regular I-optimal and nested I-optimal design that was developed for this study. The two sequential design strategies are compared for experiments with a large number of test factors using pre-experiment design evaluation metrics, as well as modeling results obtained from simulated wind tunnel data for the NASA LA-8 aircraft. The design evaluation metrics show that the sequential I-optimal base design has higher statistical power, lower correlation among candidate regressors, lower prediction variance, and more precise parameter estimates. The simulated wind tunnel experiments conducted using each design reveal that the sequential I-optimal base design has better predictive capability with fewer test points. The experiment design and evaluation procedures are described in detail to inform future aerodynamic characterization experiments for complex aircraft.

I. Introduction

Many complex distributed hybrid and electric propulsion aircraft concepts have emerged to enable future Advanced Air Mobility (AAM) transportation missions [1–6]. There are many design concepts with a variety of mission profiles, which include vertical takeoff and landing (VTOL), short takeoff and landing (STOL), and conventional takeoff and landing (CTOL) configurations. Common characteristics of these modern aircraft include the use of many control surfaces and propulsors, as well as significant propulsion-airframe interactions. There has been a desire to study these aircraft in recent years which has led NASA Langley Research Center to develop the GL-10 [7] and LA-8 [8] aircraft as the first in a series of complex electric VTOL (eVTOL) aircraft enabling advancements in many areas including rapid vehicle development [9], computational predictions [10], flight controls [11, 12], wind tunnel testing [13–15], and flight testing [16–19]. One of the applications of these aircraft is to study the complex aero-propulsive characteristics associated with eVTOL aircraft configurations [20–22].

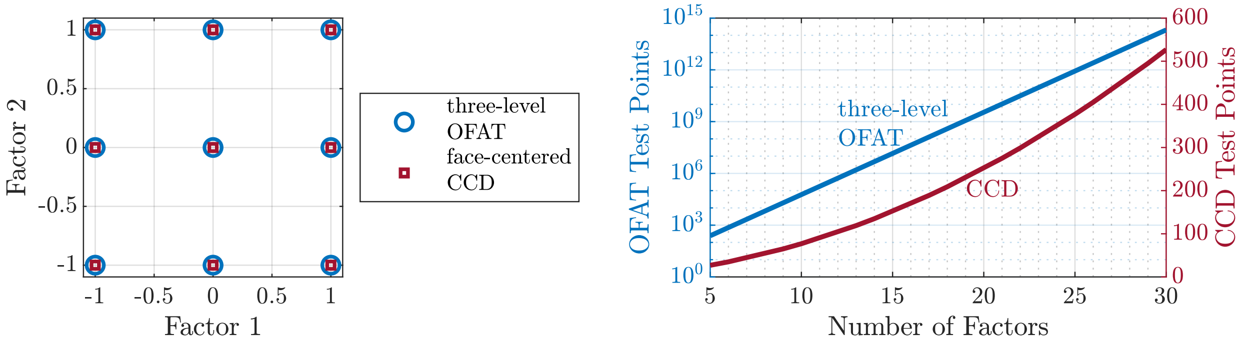
Traditional aircraft aerodynamic characterization methods generally involve using one-factor-at-a-time (OFAT) experiments, where testing is commonly conducted by sweeping one variable with the other variables held at a constant setting. This approach has been used successfully for many years in aerospace testing and yields suitable results for conventional aerospace vehicles. However, traditional static OFAT testing is not practical for developing models describing the complex nonlinear aerodynamics and interactions present with distributed propulsion aircraft due to the large number of test factors. Fortunately, experiments planned using design of experiments (DOE) [23] and response surface methodology (RSM) [24] theory can efficiently scale to a large number of factors, allowing tests to be completed in a reasonable amount of time while also supporting identification of interaction effects. DOE/RSM-based testing fundamentally provides a statistically-rigorous experiment design approach supplying rich information content in a relatively compact data set. The model development process also benefits from additional design properties of orthogonality, randomization, replication, blocking, and sequential testing [23].

As an example of the efficiency gains realized using DOE/RSM techniques, consider a complex aircraft with 22 test factors—the number of factors independently varied in static LA-8 wind tunnel experiments. OFAT testing

---

[*] Research Engineer, Flight Dynamics Branch, MS 308, Member AIAA.

covering all possible combinations of each test factor at three different levels (a low, medium, and high value) requires 31,381,059,609 test points. Alternatively, a minimum run resolution V, face-centered central composite design (CCD) with one center point, which also tests three levels of each factor and allows characterization of interactions among all pairs of test factors as well as quadratic response variation with each test factor, requires only 299 test points. A two-dimensional slice of the OFAT and face-centered CCD factor space is shown in Fig. 1a. This example comparing the number of test points for a three-level OFAT and CCD test matrix is expanded to between 5 and 30 test factors in Fig. 1b. Clearly, application of the CCD response surface design has substantially reduced the number of required test points, while still allowing characterization of nonlinear, interactional features.



(a) Two-dimensional design space slice

(b) Number of test points

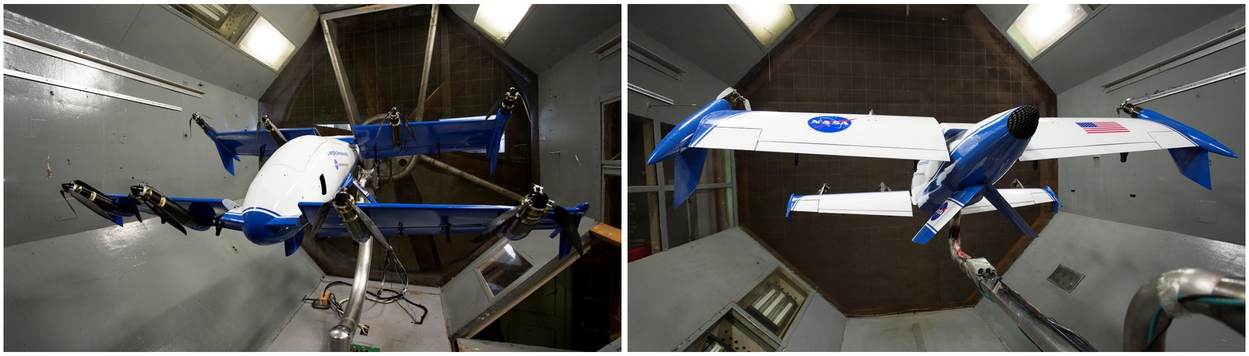
Fig. 1 Comparison of three-level OFAT and minimum run resolution V, face-centered CCD experiments.

This paper presents an assessment of multiple static experiment designs and evaluates their relative utility for application to experiments characterizing complex distributed propulsion aircraft. A new DOE/RSM test matrix design approach is formulated and compared to standard Rapid Aero Modeling (RAM) test blocks [21, 25–27] using design evaluation metrics and prediction capability assessment from simulated LA-8 wind tunnel experiments. The paper is organized as follows: Section II introduces the LA-8 aircraft, provides an overview of its wind tunnel tests, and describes the LA-8 wind tunnel simulation applied in this paper. Section III compares five foundational response surface designs using pre-experiment design evaluation metrics. This analysis informs presentation of a new block design approach developed alongside the RAM block design approach in Sec. IV. A comparison of pre-experiment design evaluation metrics for the two sequential design approaches is shown in Sec. V. An overview of the employed model identification methods is given in Sec. VI, followed by modeling results for simulated wind tunnel experiments presented in Sec. VII. Overall conclusions are summarized in Sec. VIII.

II. LA-8 Aircraft

The static experiment designs presented in this paper were applied to develop an aero-propulsive model for the Langley Aerodrome No. 8 (LA-8) vehicle [8]. The LA-8, pictured in Fig. 2, is a subscale, tandem tilt-wing, distributed electric propulsion aircraft configuration built as a testbed for eVTOL aircraft technology. The LA-8 was developed at NASA Langley Research Center as one of several eVTOL research aircraft intended to explore their unique flight characteristics and resolve implementation challenges to help bring similar full-scale vehicles into mainstream operation. The LA-8 is equipped with 20 control effectors, including two tilting wings, four elevons, four flaps, two ruddervators, and eight electric motors/propellers. Figure 3 is a photo of the LA-8 with annotations showing the vehicle propulsor and control surface definitions.

A series of LA-8 vehicle wind tunnel experiments [14, 15, 21, 22] have been performed in the NASA Langley 12-Foot Low-Speed Tunnel [28] to develop a high-fidelity aero-propulsive model of the aircraft, as well as investigate wind tunnel test techniques and aero-propulsive characteristics for eVTOL aircraft. The experiments have included OFAT testing for envelope expansion and DOE/RSM testing for aero-propulsive model development. The experimental factors specified for testing at several different tunnel dynamic pressure settings were angle of attack α, angle of sideslip β, wing angles δ_{w_1}, δ_{w_2}, elevon deflection angles δ_{e_1}, δ_{e_2}, δ_{e_3}, δ_{e_4}, flap deflection angles δ_{f_1}, δ_{f_2}, δ_{f_3}, δ_{f_4}, ruddervator deflection angles δ_{r_1}, δ_{r_2}, and motor pulse width modulation (PWM) commands η_1, η_2,..., η_8, resulting in 22 independently varied factors. Although the test matrices were defined using experimental factors of airflow angles



(a) LA-8 front view **(b) LA-8 rear view**

Fig. 2 LA-8 mounted in the NASA Langley 12-Foot Low-Speed Tunnel.

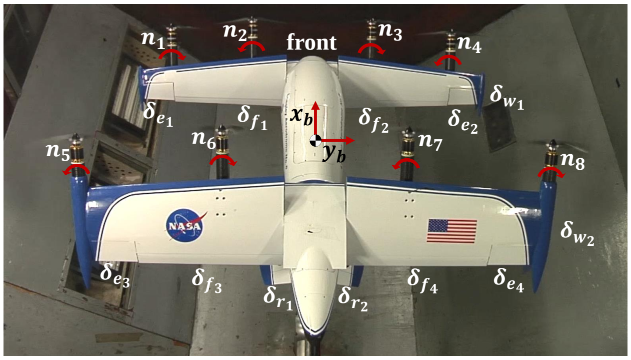


Fig. 3 LA-8 propulsor and control surface definitions.

α, β and motor PWM commands η_1, η_2,..., η_8 for operational convenience, modeling was performed using body-axis velocity components v, w and measured propeller rotational speeds n_1, n_2,..., n_8. Note that forward body-axis velocity u is not included as an explanatory variable, but its effect is implicitly described by the set of aero-propulsive models identified at different dynamic pressure settings. The LA-8 wind tunnel-based aero-propulsive model development approach is described further in Ref. [21].

The experiment designs compared in this paper were executed in an LA-8 simulation modeling the primary aero-propulsive characteristics of tilt-wing aircraft. The aircraft components included in the model were the fuselage, two rotating wings, inverted v-tail, four elevons, four flaps, two ruddervators, and eight wing-mounted propellers with geometry and placement consistent with the LA-8 vehicle. Other smaller components, such as winglets and propeller support struts, were not included in the model. The semi-empirical aircraft model utilizes strip theory-based tilt-wing aerodynamic predictions [29] combined with high-fidelity LA-8 propeller models identified from isolated propeller wind tunnel testing [30]. For this study, force and moment predictions were corrupted using zero-mean, Gaussian, white noise with measurement noise standard deviations representative of values seen in the LA-8 wind tunnel tests. The simulation was intended to be representative of LA-8 wind tunnel test results to inform future wind tunnel experiments and allow rapid, inexpensive experimentation in a controlled, repeatable test environment. The simulated experiments were executed at a dynamic pressure of $\bar{q}$ = 3.5 psf (freestream airspeed of V = 54.3 ft/s at standard sea-level conditions), with the test factor ranges shown in Table 1. This condition represents a high-speed transition phase of flight for the LA-8 aircraft. Multiple dynamic pressure settings need to be tested to develop a full-envelope aero-propulsive model [21], but these experiments provide a suitable data set for evaluation and comparison of experiment designs.

Table 1 LA-8 simulated wind tunnel experiment test factor ranges at $\bar{q}$ = 3.5 psf (V = 54.3 ft/s)

Factor(s)	Units	Minimum	Maximum
α	deg	−6	+6
β	deg	−5	+5
$\delta_{w_1}, \delta_{w_2}$	deg	0	+25
$\delta_{e_1}, \delta_{e_2}, \delta_{e_3}, \delta_{e_4}$	deg	−25	+25
$\delta_{f_1}, \delta_{f_2}, \delta_{f_3}, \delta_{f_4}$	deg	0	+20
$\delta_{r_1}, \delta_{r_2}$	deg	−30	+30
$n_1, n_2, ..., n_8$	rpm	3700	6100

III. Evaluation of Foundational Response Surface Experiment Designs

Within the body of DOE/RSM theory, there are multiple response surface design types that can yield an adequate model for experiments with a large number of test factors. This section compares five different 22-factor, cuboidal, completely randomized response surface experiment designs supporting identification of a full quadratic design model (all possible linear, pure quadratic, and two-factor interaction model terms). All the experiment designs presented in this paper were created using Design-Expert®, a commercially available statistical software package [31]. The designs include: (1) a minimum run resolution V, face-centered central composite design (FCCCD); (2) an I-optimal design; (3) an A-optimal design; (4) a D-optimal design; and (5) a distance-based optimal design. The FCCCD is composed of a two-level fractional factorial design, six center points, and a low and high axial point for each test factor, for a total of 304 test points. Minimum run resolution V fractional factorial designs contain the minimum number of test points to support estimation of linear and two-factor interaction model terms. The axial test points augment the fractional factorial design to allow estimation of pure quadratic model terms. The center points aid in stabilization of the prediction variance within the experimental region and allow estimation of pure error [23]. The other designs considered are I-, A-, D-, and distance-based optimal designs for a full quadratic design model, each with the same number of test points and center points as the FCCCD to allow direct comparisons of the designs. It has also been observed in related research that using the same number of test points as a minimum run resolution V FCCCD for a given number of test factors generally yields good design evaluation metrics for optimal designs. I-optimal designs minimize the integrated prediction variance for a predefined model order over the range of factors [23, 24, 32, 33]. Alternatively, D- and A-optimal designs focus on optimizing the design for precise estimation of model parameters for a predefined model order. D-optimal designs minimize the confidence ellipsoid volume of the model parameters and A-optimal designs minimize the sum of the variances of model parameters. The distance-based design is a Maximin design, or a design where the minimum distance between design points is maximized [24], that is modified to ensure that model terms are not aliased [31]. The designs were optimized using 298 free design points and one center point. After design optimization had completed, 5 additional center points were added to the design and the test matrix was re-randomized. This was done because it has been noted that including multiple replicate points in the center of a design during the optimization process can repel the optimized points away from the interior of the design space.

Figure 4 shows two-dimensional slices of the 22-factor space for each design. The figure shows that the I- and A-optimal designs cover the broadest number of individual factor settings in two-dimensions, although, the test points are mostly concentrated around the design space boundaries. The D-optimal and distance-based designs are more heavily concentrated around the design boundaries, with very few points in the interior of the displayed factor space. The FCCCD design only tests the low, high, and center value of each test factor. Favoring the design space boundaries can result in significant bias error in the interior of the design space when there are nonlinearities not described by the assumed model [24]. The remainder of this section presents a comparison of the design evaluation metrics of the five candidate base designs. These metrics provide insight into the design quality prior to conducing an experiment. Evaluation of the base designs is then used to justify the choice of a new sequential design aimed at mitigating against bias errors in the interior of the design space.

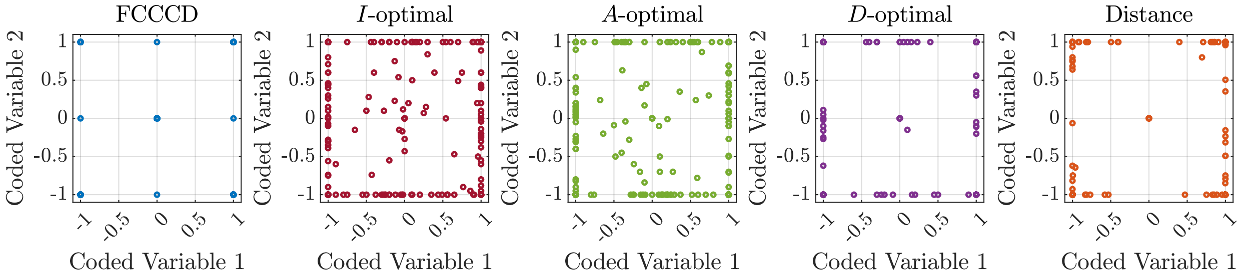


Fig. 4 Two-dimensional slices of the 22-factor space for each base design.

A. Statistical Power

Statistical power is the probability of detecting a significant effect in a model [23, 32, 34]. In other words, power is the probability of including a particular model term that is significant to the model. Power is a function of the number of design points, the placement of the design points, the significance level α_p, and the ratio of the effect size δ and noise level σ. A general rule of thumb for experiment design is to strive for statistical power of at least 80% [34].

The lowest power among all the linear (L), interaction (I), and quadratic (Q) model terms for each base design is shown graphically in Fig. 5 and numerically in Table 2. Figure 5a and Table 2a show the power for each design with $\alpha_p = 0.05$ and $\delta/\sigma = 2$, which are common values used when assessing power. All designs have nearly 100% power for the linear and interaction model terms. The power for quadratic model terms is just below 80% for the FCCCD and distance-based design, whereas the I-, A-, and D-optimal designs have close to 100% power for quadratic model terms. Figure 5b and Table 2b show the power for each design with $\alpha_p = 0.0001$ and $\delta/\sigma = 1$, which are much more conservative numbers to use in the power calculation and, accordingly, the power is lower. The FCCCD design has the lowest power for each classification of model terms, followed by the distance-based design. The highest power is generally observed for the I- and A-optimal designs. This suggests that models estimated from the I- and A-optimal designs have a lower probability of failing to include model terms that are significant to the model.

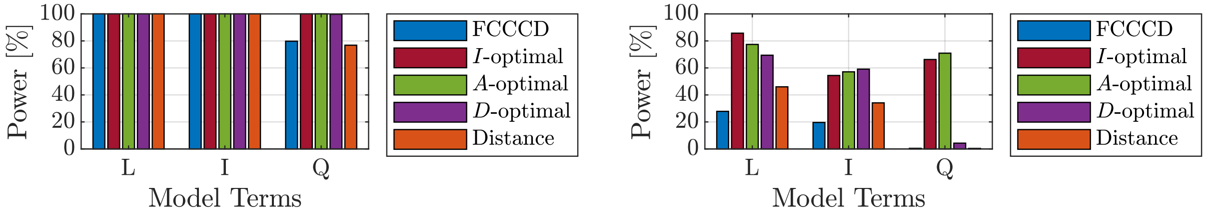
Table 2 Base design power comparison (expressed as a percentage)

(a) Power calculation with $\alpha_p = 0.05$ and $\delta/\sigma = 2$

Model Terms	Power for FCCCD	Power for I-optimal design	Power for A-optimal design	Power for D-optimal design	Power for distance-based design
L	99.999996	100.0000000000	100.0000000000	99.9999999999	99.999999995
I	99.99991	99.9999999997	99.9999999999	99.9999999999	99.9999996
Q	79.8	99.9999999999	99.9999999999	99.8	76.8

(b) Power calculation with $\alpha_p = 0.0001$ and $\delta/\sigma = 1$

Model Terms	Power for FCCCD	Power for I-optimal design	Power for A-optimal design	Power for D-optimal design	Power for distance-based design
L	27.8	85.7	77.4	69.4	46.0
I	19.6	54.4	57.1	59.1	34.1
Q	0.4	66.2	70.9	4.3	0.4



(a) Power calculation with $\alpha_p = 0.05$ and $\delta/\sigma = 2$

(b) Power calculation with $\alpha_p = 0.0001$ and $\delta/\sigma = 1$

Fig. 5 Base design power comparison.

B. Correlation Metrics

Data collinearity is defined as a correlation between regressors high enough to cause corrupted model identification [35]. Data collinearity will cause difficulty in both model structure determination and parameter estimation because the effects of certain regressors on the response cannot be distinguished. Model structure identification is corrupted by candidate regressor correlation, particularly for the large number of candidate model terms associated with modeling eVTOL aircraft, because an algorithm is more inclined to include model terms that lack physical reality or exclude model terms describing significant aerodynamic phenomena [21]. Parameter estimation algorithms cannot differentiate between highly correlated model terms, resulting in inaccurate parameter estimates and uncertainties from the poorly conditioned estimation problem [35]. For these reasons, it is important to develop an experiment design with low correlation among candidate regressors.

Correlation between two candidate regressors can be assessed using the pairwise correlation coefficient

$$r_{ij} = \frac{(\boldsymbol{x}_i - \bar{x}_i)^T (\boldsymbol{x}_j - \bar{x}_j)}{\sqrt{(\boldsymbol{x}_i - \bar{x}_i)^T (\boldsymbol{x}_i - \bar{x}_i)}\sqrt{(\boldsymbol{x}_j - \bar{x}_j)^T (\boldsymbol{x}_j - \bar{x}_j)}} \tag{1}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are two regressor measurement histories, with means denoted $\bar{x}_i$ and $\bar{x}_j$, respectively. A correlation coefficient value of zero means the regressors are uncorrelated, or orthogonal, and an absolute correlation coefficient of one indicates that the regressors are completely correlated. A correlation coefficient between regressors with magnitude greater than 0.9 indicates that data collinearity problems may be encountered [35, 36]. Another metric that can be used to assess candidate regressor correlation is the variance inflation factor (VIF). For the regressors $\boldsymbol{x}_j$, the respective VIF is

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \tag{2}$$

where R_j^2 is the coefficient of determination obtained through creating a regression model of $\boldsymbol{x}_j$ as a function of all other regressors. A VIF value greater than 10 suggests that data collinearity may be present [33, 35, 37]. The r_{ij}, VIF_j, and R_j^2 metrics only quantify correlation between pairs of regressors and, thus, cannot diagnose collinearity among more than two regressors [35, 38].

An alternative method that can be used to assess multiple correlation between more than two inputs is analysis of the eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$, where $\boldsymbol{X}$ is a matrix composed of column vectors of the regressors in the regression model. The inverse of the $\boldsymbol{X}^T\boldsymbol{X}$ matrix is required to compute the ordinary least-squares regression solution. The ratio of the maximum eigenvalue and minimum eigenvalue, $\kappa = \lambda_{\max}/\lambda_{\min}$, is the condition number of the $\boldsymbol{X}^T\boldsymbol{X}$ matrix. A value of κ close to one indicates low multiple correlation whereas a large value of κ indicates an ill-conditioned estimation problem due to data collinearity. Values of κ indicating adverse effects from data collinearity range anywhere from 100 to 100,000 depending on the particular data set [35–38].

The pairwise and multiple correlation metrics for each design evaluated for a full quadratic model are shown in Figs. 6-9. This represents a conservative analysis performed prior to the experiment that assumes all candidate regressors are included in the model. Many candidate model terms considered here for the full quadratic model were expected to be excluded through model structure determination after data collection.

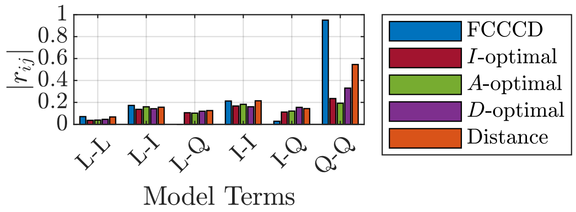


Fig. 6 Maximum absolute r_{ij} values among candidate regressors in a quadratic model for each base design.

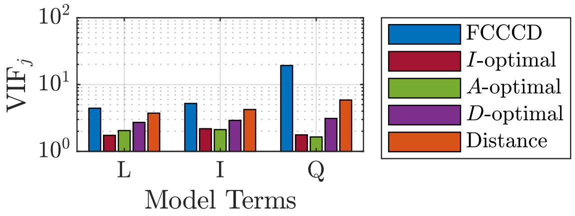


Fig. 7 Maximum VIF$_j$ for candidate regressors in a quadratic evaluation model for each base design.

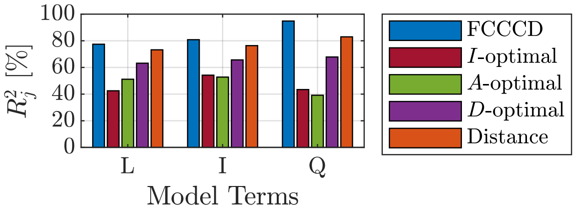


Fig. 8 Maximum R_j^2 for candidate regressors in a quadratic evaluation model for each base design.

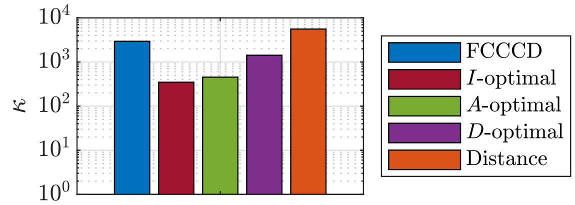


Fig. 9 Condition number of $X^T X$ in a full quadratic evaluation model for each base design.

Figure 6 shows the maximum absolute pairwise correlation value, $|r_{ij}|$, among pairs of linear regressors (L-L), linear-interaction regressors (L-I), linear-quadratic regressors (L-Q), interaction regressors (I-I), interaction-quadratic regressors (I-Q), and quadratic regressors (Q-Q). The maximum VIF$_j$ and R_j^2 for linear (L), two-factor interaction (I), and quadratic (Q) model terms are shown in Figs. 7-8. Figure 9 shows the condition number κ of $\boldsymbol{X}^T\boldsymbol{X}$ for a full quadratic model. The overall takeaway from these figures is that the correlation metric values are generally lowest for the I- and A-optimal designs. The correlation metrics are generally highest for the FCCCD and distance-based designs, with the correlation metrics associated with the quadratic model terms being the highest, particularly for the FCCCD, where the pairwise correlation and VIF for the quadratic regressors are above 0.95, and 19, respectively. This analysis supports the recommendation given in Ref. [39] against using a face-centered central composite design for a large number of test factors because of the high correlation among the quadratic model terms.

C. Prediction Variance

Assessment of the prediction variance of a response surface design for a given model structure provides insight into its precision of prediction and allows comparison of different response surface designs. The variance of the predicted response is

$$\mathrm{Var}[\hat{y}(\boldsymbol{x}_0)] = \sigma^2 \boldsymbol{x}_0^T \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \boldsymbol{x}_0 \tag{3}$$

where $\hat{y}(\boldsymbol{x}_0)$ is the predicted response evaluated at the design space location $\boldsymbol{x}_0$ expanded to the form of the model structure, $\boldsymbol{X}$ is a matrix composed of the designed test points in the form of the model structure, and σ^2 is the measurement error variance [24]. From Eq. (3), the prediction variance is a function of the experiment design, the model structure, the location in the design space, and the measurement facility error variance. The unscaled prediction variance (UPV), defined as

$$\mathrm{UPV} = \frac{\mathrm{Var}[\hat{y}(\boldsymbol{x}_0)]}{\sigma^2} = \boldsymbol{x}_0^T \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \boldsymbol{x}_0 \tag{4}$$

removes the dependence on σ^2 and, thus, can be used to compare experiment designs when σ^2 is unknown prior to experimentation.

Graphical presentation of the distribution of prediction variance throughout the design space is an effective way to assess experiment designs. Fraction of design space (FDS) plots, introduced by Ref. [40], depict the prediction variance distribution over the design space in a concise manner, where the prediction variance metrics are plotted against the FDS encompassing a prediction variance less than or equal to a particular value. It is also useful to consider the FDS including a particular model precision, quantified by the confidence interval half-width δ [33, 41, 42]. The model precision δ normalized by the response standard deviation σ plotted against FDS provides further insight into the prediction capability of the model developed from a particular experiment design, prior to conducting the experiment. For this study, a design was deemed to be adequate for fitting a particular model complexity if δ/σ was less than two for greater than 95% of the design space. The prediction variance threshold PV* used to determine the FDS within a given model precision level is

$$\mathrm{PV}^* = \left(\frac{\delta/\sigma}{t_{\alpha_p/2,N-p}}\right)^2 \tag{5}$$

where N is the number of test points, p is the number of parameters in the model, and α_p is the significance level chosen as $\alpha_p = 0.05$.

Figure 10 shows the UPV and δ/σ threshold values against FDS for each base experiment design using a quadratic evaluation model order. The average UPV for each design is also shown. The FCCCD and distance-based designs have the highest UPV and δ/σ threshold across the design space; the I- and A-optimal designs have the lowest UPV and δ/σ threshold across the design space. The I-optimal design has the lowest average UPV, which is expected because the design objective for the I-optimal designs is to minimize the average prediction variance across the design space; however, the prediction variance distribution for the A-optimal design is very close to the I-optimal design.

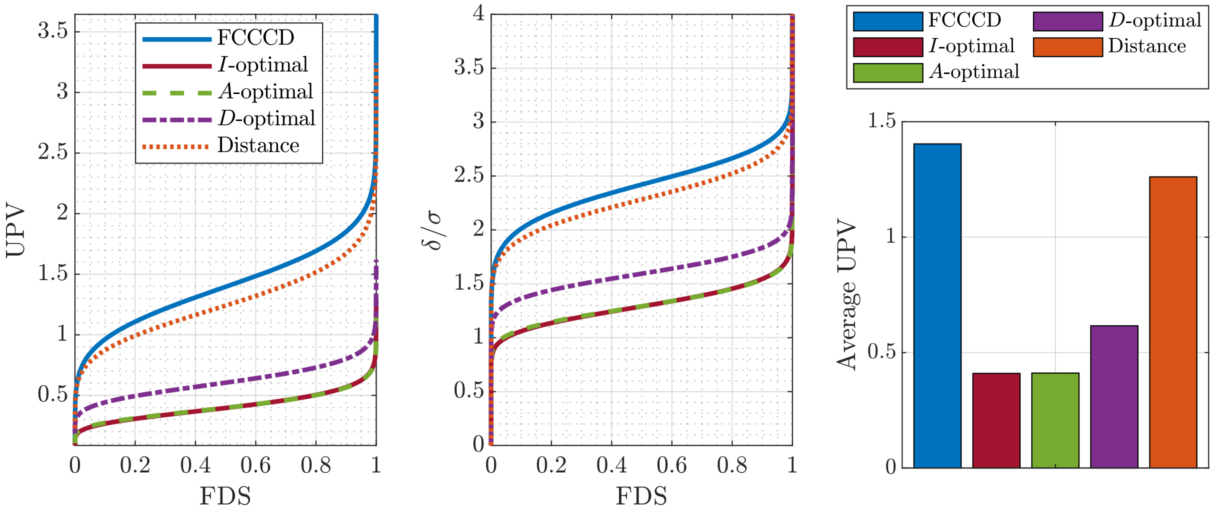


Fig. 10 Prediction variance plots for a quadratic evaluation model for each base design.

Table 3 lists the FDS with $\delta/\sigma \leq 1$, $\delta/\sigma \leq 1.5$, and $\delta/\sigma \leq 2$ for each design using a full quadratic evaluation model. An adequate FDS (FDS ≥ 0.95) for a normalized model precision $\delta/\sigma \leq 2$ is only obtained for the I-, D-, and A-optimal designs. Evaluation with $\delta/\sigma \leq 1$ and $\delta/\sigma \leq 1.5$ shows that, the I- and A-optimal designs have the highest FDS meeting these δ/σ thresholds.

Table 3 Prediction variance threshold FDS values using a quadratic evaluation model for each base design

Design	FDS with $\delta/\sigma \leq 1$	FDS with $\delta/\sigma \leq 1.5$	FDS with $\delta/\sigma \leq 2$
FCCCD	0.000	0.003	0.093
I-optimal	0.048	0.853	0.999
A-optimal	0.037	0.858	0.999
D-optimal	0.000	0.297	0.982
distance-based	0.000	0.003	0.161

D. Model Parameter Precision

The precision of the estimated model parameters can be assessed using the standard error of individual model parameters or properties of the $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ matrix. Figure 11 shows the highest standard error for linear (L), interaction (I), and quadratic (Q) model terms for each design. The standard errors are normalized by assuming unity measurement noise variance and are denoted SE*. The standard error of the model terms estimated using the FCCCD and distance-based designs are the largest, particularly for the quadratic model terms, as a consequence of the higher correlation associated with the quadratic model terms. The I- and A-optimal designs generally have the lowest standard error values. Figure 12 shows the scaled D-optimality criterion [39]

$$D^* = N\left(\det\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right]\right)^{1/p} \tag{6}$$

and the trace of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ which reflects the A-optimality [24]. The D-optimal design has the lowest D^* and the A-optimal design has the lowest value of $\mathrm{tr}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}]$, as expected. The FCCCD design has the highest D^* and $\mathrm{tr}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}]$ values.

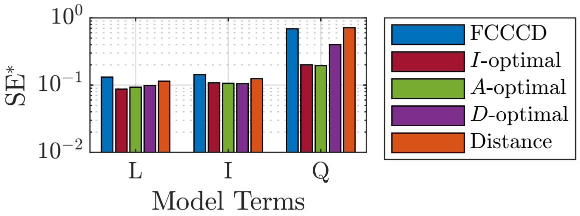
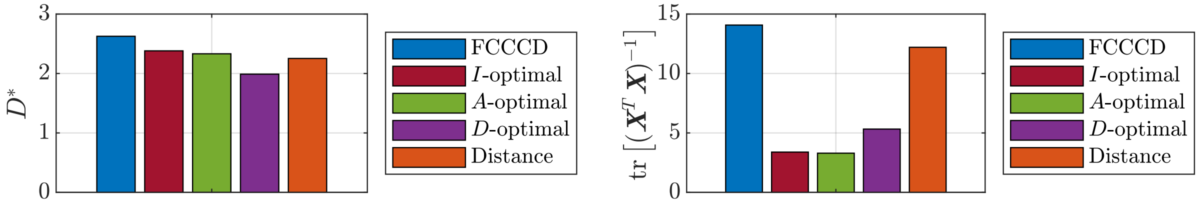


Fig. 11 Normalized standard error of the model parameters in a full quadratic model for each base design.



(a) Scaled D-optimality metric

(b) A-optimality metric

Fig. 12 Overall model parameter precision metrics for a full quadratic evaluation model for each base design.

E. Discussion

The analysis in the previous subsections has shown that the I- and A-optimal designs generally have the highest statistical power for candidate model terms, the lowest correlation among candidate model terms, the lowest prediction variance across the design space, and the lowest model parameter standard errors. Based on these metrics, the I- and A-optimal designs are expected to yield the most appropriate model structure, provide the most precise estimates of model parameters, and have the best predictive performance. The FCCCD and distance-based designs generally have the worst design evaluation metrics, and the D-optimal design generally has moderate values between the most and least favorable designs. Based on this analysis, the I- and A-optimal designs are recommended for experiments with a large number of test factors in a cuboidal test region, as is needed to develop an aero-propulsive model for a distributed propulsion aircraft. For this study, the decision was made to proceed with the I-optimal over the A-optimal design because of the slightly better precision of prediction. An I-optimal design will be expanded in the next section to address the scarcity of design points in the interior of the design space.

IV. Sequential Response Surface Experiment Designs

Two separate sequential, cuboidal, completely randomized, response surface block designs were executed and compared to investigate possible improvement of modeling results and reduction of the number of test points needed for future aircraft characterization experiments. The designs will be referred to as the face-centered base (FCB) design and a *I*-optimal base (IOB) design. The FCB design has been used previously in wind tunnel and computational testing for eVTOL aircraft [20, 21, 43] and is the set of test blocks used for the RAM process [25–27]. The IOB design is a new block design process introduced in this work building off of previous work which developed the FCB design. The *I*-optimal design was selected for this study based on the design evaluation analysis presented in Sec. III. The block design approach for the FCB and IOB designs are described in the following subsections. For each design, a series of five test blocks was designed to acquire the data necessary to incrementally develop more accurate aero-propulsive models.

A. Face-Centered Base (FCB) Design

The sequential FCB design blocks for the LA-8 22-factor experiments are as follows [21]:

- **Block 1** is the same minimum run resolution V, face-centered central composite design (FCCCD) described in Sec. III.
- **Block 2** is a nested FCCCD. The nested FCCCD was originally developed in Ref. [44], where it was shown to be an effective design strategy for a 5-factor fixed-wing aircraft aerodynamic characterization experiment. The nested design augments the conventional FCCCD to allow estimation of pure cubic model terms and provides additional information in the interior region of the design space. The nested FCCCD applied for this work emulates the ordinary FCCCD, except that the factorial and axial points are all located at half the distance from the center of the design space. The nested FCCCD is designed separately from the Block 1 FCCCD (i.e., knowledge of the data points within Block 1 do not influence the choice of data points in Block 2).
- **Block 3** is an *I*-optimal design for a full quadratic design model with the number of model points being five times the number of test factors [20]. Five additional lack-of-fit points, which are selected using the Maximin distance-based criterion, as well as five additional non-center replicates are also included in the design [39]. Inclusion of five lack-of-fit points and five non-center replicates are the default settings in Design-Expert® for base optimal response surface designs to permit lack-of-fit testing [31]. The *I*-optimal design augments the designs from previous blocks to sequentially improve the model and avoid duplicating previously tested combinations of factor settings. In other words, the previously tested FCCCD and nested FCCCD design points are factored into the optimization algorithm to minimize the prediction variance.
- **Block 4** is another augmented *I*-optimal design for a quadratic design model, which follows the same augmented design procedure from Block 3. With fewer test factors, the design model could be increased to a higher order, for example, including cubic model terms [26, 27].
- **Block 5** is another augmented *I*-optimal design for a quadratic design model, or the highest design model complexity from Blocks 1-4, which is used for model validation. The block contains 75 validation test points, which has been found to provide a good estimate of prediction error while remaining a modest number of test points. No lack-of-fit, replicate, or center points are included. Because the validation data are *I*-optimal test points designed to augment the data used for modeling to optimally reduce prediction error, but are not used to fit the models, the data provide a rigorous prediction test for the designed model complexity. The factor settings for these validation data, however, are dependent on the designs used for model identification and tend to favor the boundaries of the design space.

B. *I*-Optimal Base (IOB) Design

The sequential IOB design blocks for the LA-8 22-factor experiments are as follows:

- **Block 1** is the same *I*-optimal design for a quadratic design model described in Sec. III. The design includes the same total number of test points and the same number of center points as the Block 1 FCB design to provide a direct comparison. Lack-of-fit points and non-center replicates are not included in the design.
- **Block 2** is a nested *I*-optimal design for a quadratic design model, which was inspired by the concept of the nested FCCCD proposed in Ref. [44]. The nested *I*-optimal emulates the same design process as Block 1, except that the design points are all located at half the distance from the center of the design space. The nested *I*-optimal design was created separately from the Block 1 *I*-optimal design (i.e., knowledge of the Block 1 design does not influence the Block 2 design). The design provides more broad coverage of the interior portion of the design space, which aids in reducing model bias [24].
- **Block 3** is an *I*-optimal design for a quadratic design model with the number of model points being five times

the number of experimental factors, following Ref. [20]. Five additional lack-of-fit points and five additional non-center replicates are also included in the design, following the Block 3 FCB design approach. The I-optimal design augments the designs from previous blocks to sequentially improve the model and avoid duplicating previously tested combinations of factor settings.

- **Block 4** is another augmented I-optimal design for a quadratic design model following the same augmented design procedure from the previous block, and also following the Block 4 FCB design approach.
- **Block 5** contains 75 additional validation test points selected using a random number generator. This block provides validation data that are agnostic to the model development experiment design and the model complexity it was designed for. Since this block is design agnostic, it will be used to directly compare the IOB design with the FCB design later in the paper. Note that for the simulated LA-8 wind tunnel experiments executed for this study, the validation block was increased to 304 test points (the same number of test points as Block 1 and Block 2) selected using a random number generator to provide a greater validation sample size. Increasing the number of validation points for these simulated experiments is justified because of the low computational expense of executing each test point, but acquiring a large number of validation test points would not be practical for an expensive and/or time-consuming experiment.

C. Block Design Summary

The FCB and IOB designs are summarized in Table 4. The designs intentionally have the same number of test points in each block so that the design qualities can be compared directly. The difference between the sets of test blocks is the type of design used in Block 1, Block 2, and Block 5. It is worth noting that the FCCCD designs are available nearly instantaneously, whereas each I-optimal design blocks can take roughly 1-2 hours to create due to the computational expense of the employed coordinate exchange optimization algorithm [24, 32]. Figures 13-14 show two-dimensional slices of the 22-factor space for each design. Each block is plotted sequentially with points from the previous blocks to show how the higher complexity designs fill the design space. The figure shows that the IOB design covers a broader number of individual factor settings compared to the FCB design. The next section compares the pre-experiment design evaluation metrics for the FCB and IOB designs.

Table 4 FCB and IOB test block design summary

(a) FCB design

Block	Design Type	Block Points	Cumulative Points
1	FCCCD	304	304
2	nested FCCCD	304	608
3	I-optimal	120	728
4	I-optimal	120	848
5	I-optimal (validation)	75	923

(b) IOB design

Block	Design Type	Block Points	Cumulative Points
1	I-optimal	304	304
2	nested I-optimal	304	608
3	I-optimal	120	728
4	I-optimal	120	848
5	random (validation)	75	923

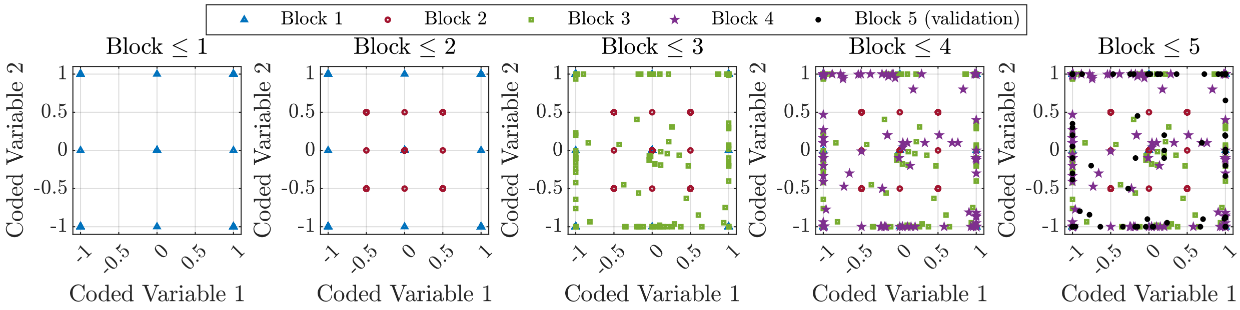


Fig. 13 Sequential two-dimensional slices of the 22-factor space for each FCB test block.

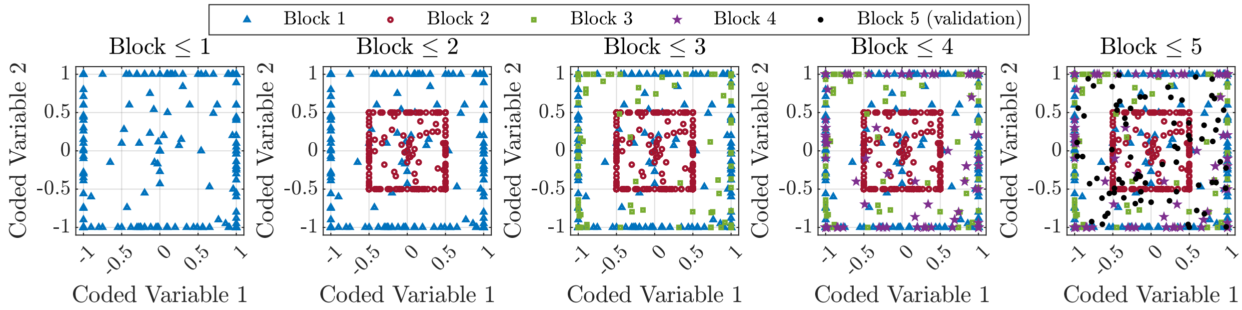


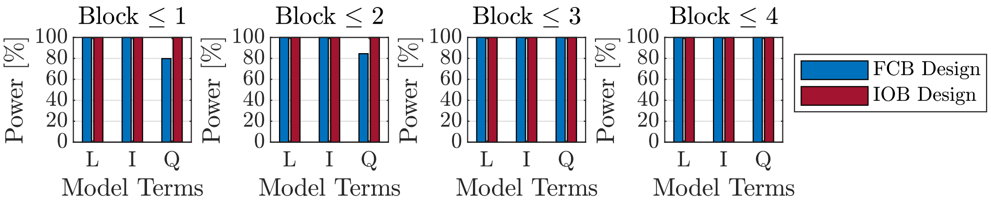
Fig. 14 Sequential two-dimensional slices of the 22-factor space for each IOB test block.

V. Sequential Experiment Design Evaluation

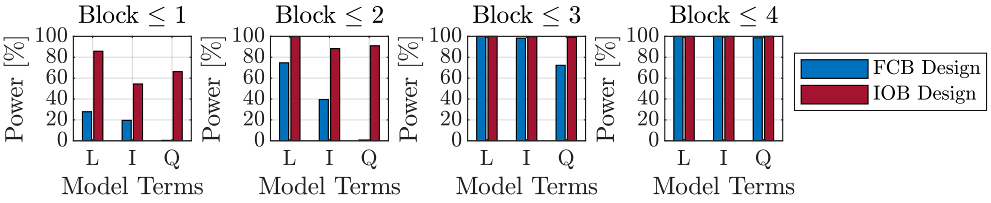
This section compares the sequential FCB and IOB experiment design approaches using the pre-experiment design evaluation metrics described in Sec. III. The designs are assessed sequentially at each block, meaning that the analysis for each block includes the design points in the current block and all previous blocks.

A. Statistical Power

The lowest statistical power among all the linear (L), interaction (I), and quadratic (Q) model terms for each design and block number are shown in Fig. 15. The power values are also shown numerically in Table 5. Figure 15a and Table 5a show the power for each design with common values of $\alpha_p = 0.05$ and $\delta/\sigma = 2$. The FCB design has nearly 100% power for the linear and interaction model terms in all blocks. The power for quadratic model terms is near 80% for Blocks 1-2 and near 100% for Blocks 3-4. The IOB design has nearly 100% power for all model terms in all blocks. Figure 15b and Table 5b show the power for each design with conservative values of $\alpha_p = 0.0001$ and $\delta/\sigma = 1$, resulting in lower power. For Block 1, the lowest power among the groups of model terms is between 54% and 86% for the IOB design and between 0% and 28% for the FCB design. For Block 2, each group of model terms has a power value below 75% for the FCB design, whereas all model terms have a power above 88% for the IOB design. For Block 3, the lowest power for the quadratic model terms is 72.3% for the FCB design, but above 99% for the IOB design; the power for linear and interaction model terms are above 98% for both designs, but the power is higher for the IOB design. For Block 4, all model terms in each design have a high power, but the power for the IOB design is higher. The overall takeaway is that, for the same number of design points, the statistical power for the IOB design is always higher than the FCB design, particularly for the quadratic model terms. This means that for the same significance level, models



(a) Power calculation with $\alpha_p = 0.05$ and $\delta/\sigma = 2$



(b) Power calculation with $\alpha_p = 0.0001$ and $\delta/\sigma = 1$

Fig. 15 FCB and IOB design power comparison.

estimated from the IOB design have a lower probability of failing to include model terms that are significant to the model. The power is also seen to increase as the block number increases, as would be expected when increasing the total number of available data points.

Table 5 FCB and IOB design power comparison (expressed as a percentage)

(a) Power calculation with α_p = 0.05 and δ/σ = 2

Block	Model Terms	Power for FCB Design	Power for IOB Design
1	L	99.999996	100.0000000000
1	I	99.99991	99.9999999997
1	Q	79.8	99.9999999999
2	L	99.9999999999	100.0000000000
2	I	99.999996	99.9999999999
2	Q	84.5	100.0000000000
3	L	100.0000000000	100.0000000000
3	I	100.0000000000	100.0000000000
3	Q	99.9999999999	100.0000000000
4	L	100.0000000000	100.0000000000
4	I	100.0000000000	100.0000000000
4	Q	100.0000000000	100.0000000000

(b) Power calculation with α_p = 0.0001 and δ/σ = 1

Block	Model Terms	Power for FCB Design	Power for IOB Design
1	L	27.8	85.7
1	I	19.6	54.4
1	Q	0.4	66.2
2	L	74.6	99.9
2	I	39.5	88.2
2	Q	0.8	91.0
3	L	99.96	99.9996
3	I	98.3	99.6
3	Q	72.3	99.2
4	L	99.999997	99.9999998
4	I	99.990	99.998
4	Q	98.7	99.98

B. Correlation Metrics

The pairwise and multiple correlation metrics for each design evaluated for a full quadratic model are shown in Figs. 16-19. As mentioned in Sec. III.B, this is a conservative analysis that assumes all candidate regressors are included in the model and many candidate model terms were expected to be excluded through model structure determination after data collection. The important takeaway from these figures is that the correlation metric values are generally lower for the IOB design compared to the FCB design. For the FCB design, the correlation metrics associated with the quadratic model terms are the highest, particularly for the Block 1 FCCCD and Block 2 nested FCCCD, where the pairwise correlation and VIF for the quadratic regressors are above 0.95, and 19, respectively. The condition number for each FCB design block is also higher compared to the corresponding IOB block, indicating that adverse effects from multiple correlation are greater, particularly for the Block 1 FCCCD and Block 2 nested FCCCD designs. Again, this analysis assumes a full quadratic model; if a subset of the model terms is selected for the model, then the condition number will be reduced.

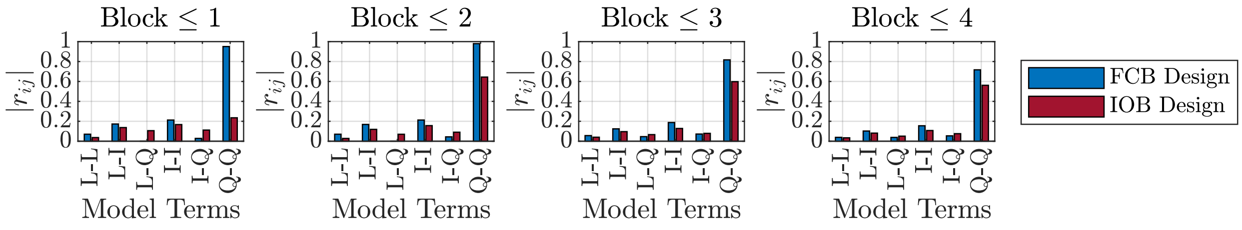


Fig. 16 Maximum absolute r_{ij} values among candidate regressors in a quadratic model for FCB and IOB designs.

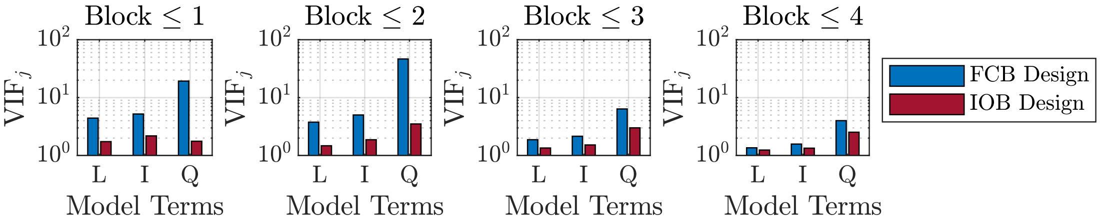


Fig. 17 Maximum VIF_j for candidate regressors in a quadratic evaluation model for FCB and IOB designs.

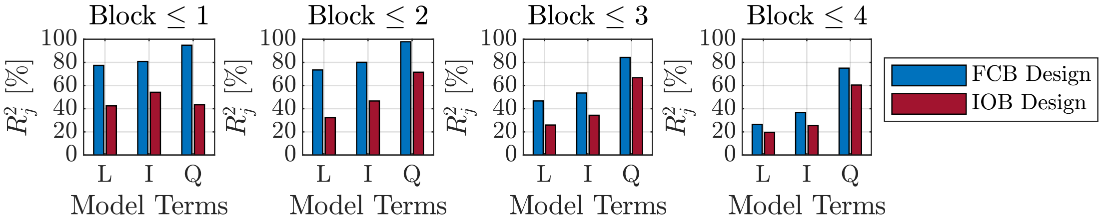


Fig. 18 Maximum R_j^2 for candidate regressors in a quadratic evaluation model for FCB and IOB designs.

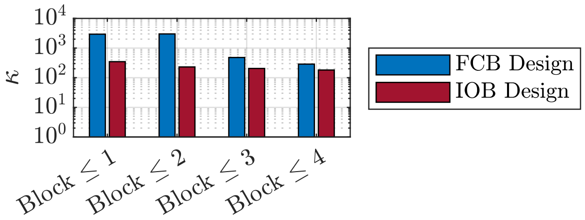


Fig. 19 Condition number of $X^T X$ in a full quadratic evaluation model for FCB and IOB designs.

C. Prediction Variance

Figure 20 shows the UPV and δ/σ threshold values against FDS, as well as the average UPV, for each experiment design using a quadratic evaluation model order. The UPV and δ/σ threshold curve decreases in value and becomes more uniform (flat) as the block number increases. Block 1 and Block 2 for the FCB design (a FCCCD and nested FCCCD) have a significantly higher UPV across the design space compared to Block 1 and Block 2 for the IOB design (an I-optimal and nested I-optimal design). This is expected because the design objective for the I-optimal designs is to minimize the average prediction variance across the design space. The respective Block 3 and Block 4 designs have a more similar prediction variance distribution and average value because both design approaches employ augmented I-optimal designs for these blocks, but the IOB design blocks still have a lower UPV value across the design space.

Table 6 lists the FDS with $\delta/\sigma \leq 1$ and $\delta/\sigma \leq 2$ for each design using a full quadratic evaluation model. An adequate FDS (FDS ≥ 0.95) for a normalized model precision $\delta/\sigma \leq 2$ is obtained for each IOB block, whereas this threshold is only achieved in Block 3 and Block 4 for the FCB design. Evaluation with $\delta/\sigma \leq 1$ shows that, for each block, a larger FDS meets this more stringent requirement for the IOB design compared to the FCB design.

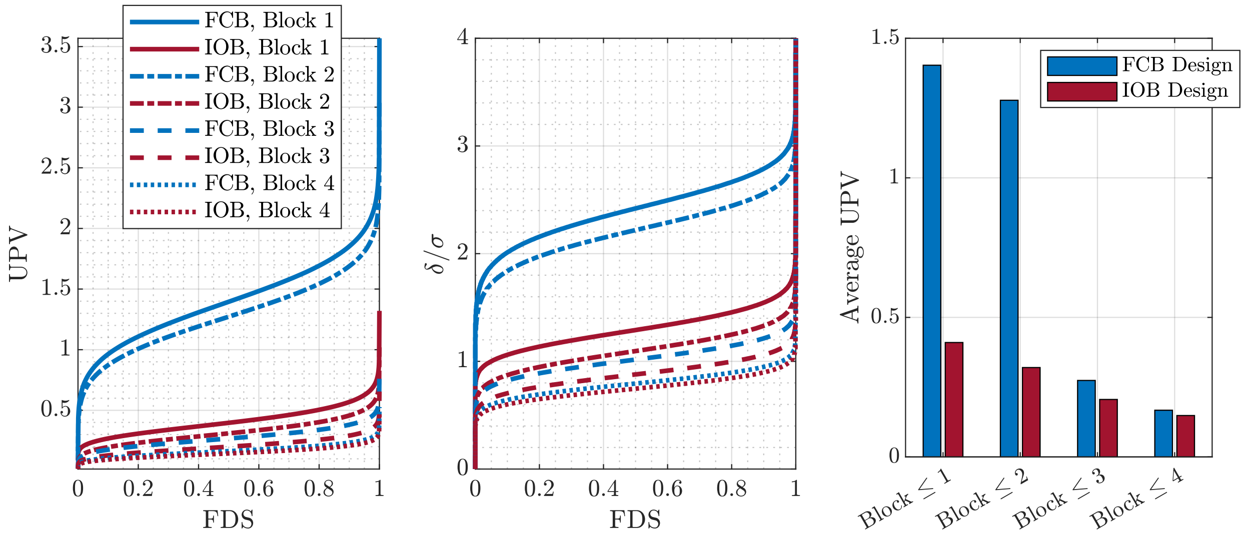


Fig. 20 Prediction variance plots for a quadratic evaluation model for FCB and IOB designs.

Table 6 Prediction variance threshold FDS values using a quadratic evaluation model for FCB and IOB designs

Block	FCB Design FDS with $\delta/\sigma \leq 1$	IOB Design FDS with $\delta/\sigma \leq 1$	FCB Design FDS with $\delta/\sigma \leq 2$	IOB Design FDS with $\delta/\sigma \leq 2$
1	0.000	0.048	0.093	0.999
2	0.000	0.288	0.222	1.000
3	0.448	0.799	1.000	1.000
4	0.953	0.982	1.000	1.000

D. Model Parameter Precision

Figure 21 shows the highest standard error for each group of model terms for each design and block number. The standard error of the model terms estimated using the FCB design are greater, particularly for the quadratic model terms, as a consequence of the higher correlation associated with the quadratic model terms. Figure 22 shows the scaled D-optimality criterion [39] and the trace of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ (representing the A-optimality [24]). For Block 1 and Block 2, the IOB design has lower D- and A-optimality metrics compared to the FCB designs. For Block 3 and Block 4, the D- and A-optimality metrics are similar for each design.

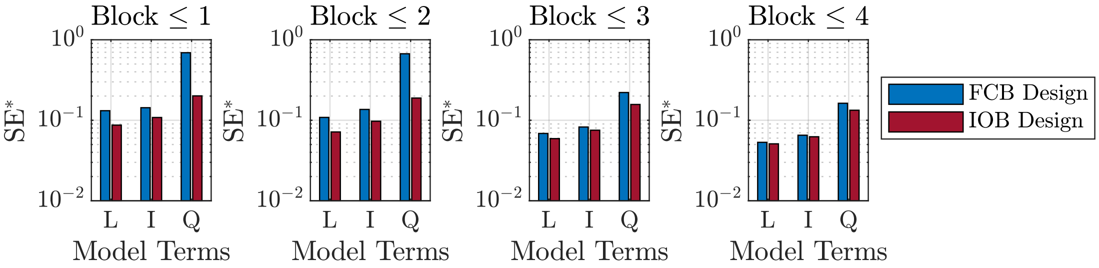
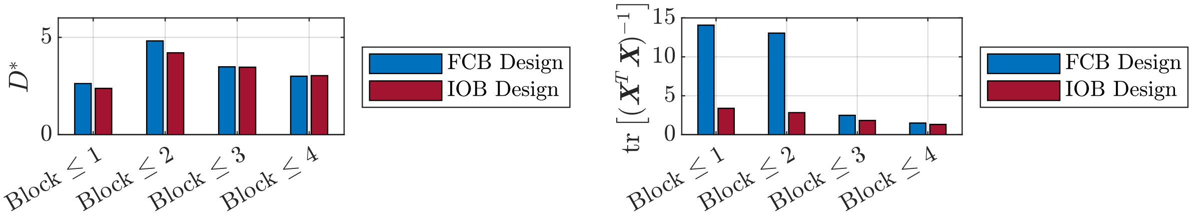


Fig. 21 Normalized standard error of the model parameters in a full quadratic model for FCB and IOB designs.



(a) Scaled D-optimality metric

(b) A-optimality metric

Fig. 22 Overall model parameter precision metrics for a full quadratic evaluation model for FCB and IOB designs.

E. Discussion

This section presented several pre-experiment design metrics and direct comparisons were made between the FCB and IOB design approaches assuming a full quadratic model structure. The analysis indicated that the IOB design has improved statistical power, correlation among model terms, prediction variance, and precision of estimated model parameters compared to the FCB design. The following sections further investigate the two experiment design approaches by comparing results from simulated LA-8 wind tunnel experiments.

VI. Model Identification Approach

Aero-propulsive modeling for this effort focuses on developing a polynomial representation of the aero-propulsive forces and moments as a function of vehicle state and control variables. Factors under test, or close variants (as discussed in Sec. II and Ref. [21]), are defined as explanatory variables, and a model is identified from the data collected using the experiment designs described in Sec. IV. Aero-propulsive modeling for tilt-wing, distributed electric propulsion aircraft requires a different approach compared to conventional fixed-wing and rotary-wing aircraft modeling approaches. Transitioning eVTOL vehicles can be considered a fixed-wing/rotary-wing hybrid, suggesting that a combination of modeling approaches will facilitate suitable model development. Accordingly, the modeling variable formulation described in Ref. [21], and applied here, merges appropriate fixed-wing and rotary-wing modeling attributes to model the LA-8 and other similar vehicles.

The modeled responses are the dimensional body-axis aero-propulsive forces X, Y, Z in lbf and moments L, M, N in ft·lbf. The explanatory variables are defined as the body-axis velocity components v, w in ft/s, propeller speeds $n_1, n_2, \ldots, n_8$ in revolutions per second, wing angles $\delta_{w_1}, \delta_{w_2}$ in radians, elevon deflections $\delta_{e_1}, \delta_{e_2}, \delta_{e_3}, \delta_{e_4}$ in radians, flap deflections $\delta_{f_1}, \delta_{f_2}, \delta_{f_3}, \delta_{f_4}$ in radians, and ruddervator deflections $\delta_{r_1}, \delta_{r_2}$ in radians. Note that it is important to perform modeling with explanatory variables expressed in coded units, or in engineering units with the explanatory variables centered on a reference value, to maintain low correlation among candidate regressors.

The model structure identification and parameter estimation methods used for this work were adapted from the System IDentification Programs for AirCraft (SIDPAC) software toolbox [35, 45]. The model structure was developed using the stepwise regression algorithm described in Ref. [46] and the model parameters were estimated using ordinary least-squares regression. The stepwise regression algorithm is a combination of forward selection and backwards elimination of candidate regressors where a single regressor is either added to or removed from the model at each iteration. If the partial F-statistic for a term included in the model falls below a cutoff threshold $F(\alpha_p, 1, N-p)$, the term is removed from the model. Otherwise, the excluded candidate model terms with the highest correlation to the unmodeled portion of the response variable is added into the model. The algorithm was run automatically until the remaining excluded model terms did not surpass the partial F-statistic cutoff value when added to the model. The significance level α_p is commonly chosen as $\alpha_p = 0.05$, or 95% confidence that a model term is significant; however, this threshold has been noted to admit a large number of model terms that lack physical justification for eVTOL aircraft aero-propulsive modeling problems. Over-parameterizing a model is undesirable because the model can yield unrealistic response predictions and unnecessary curvature. For the data analyzed in this work, a significance level from $\alpha_p = 0.01$ to $\alpha_p = 0.0001$ (99% confidence to 99.99% confidence, respectively) appeared to be a good choice to obtain a parsimonious model with good prediction capability.

After identifying the model structure and parameter estimates, model adequacy was examined using data withheld from the model development process. Model fit metrics and modeling residuals alone do not provide information about the model predictive capability. Assessment of model performance using validation data not used for modeling provides a more reliable estimate of model prediction accuracy. Validation assessment can be performed by analyzing the prediction residuals ($\boldsymbol{e} = \boldsymbol{z} - \hat{\boldsymbol{y}}$) between the measured response $\boldsymbol{z}$ and predicted response $\hat{\boldsymbol{y}}$. Comparison of modeling and prediction residuals is useful because a significant increase in the spread of prediction residuals compared to modeling residuals is a way of diagnosing an inadequate model. Residuals and their statistical properties can be given further interpretability by normalization. A good error normalization metric is the range of response variable measurements used to develop the model, $\text{range}(\boldsymbol{z}) = z_{\text{max}} - z_{\text{min}}$. Range normalization provides a fair comparison between prediction error metrics for different response variables used for aircraft modeling because longitudinal responses are generally biased above or below zero and lateral-directional responses are generally centered about zero. The normalized residual vector is defined as:

$$\boldsymbol{e}^* = \frac{\boldsymbol{z} - \hat{\boldsymbol{y}}}{\text{range}(\boldsymbol{z})} \tag{7}$$

Similarly, the normalized root-mean-square modeling error (NRMSE) is defined as:

$$\text{NRMSE} = \frac{1}{\text{range}(\boldsymbol{z})} \sqrt{\frac{(\boldsymbol{z} - \hat{\boldsymbol{y}})^T (\boldsymbol{z} - \hat{\boldsymbol{y}})}{N}} \tag{8}$$

The range-normalized NRMSE is a good metric for model validation because it:

1) succinctly presents a quantitative assessment and comparison of the model fit and prediction performance,
2) allows comparisons of the prediction capability for responses with different character and units, and
3) is straightforward to interpret as a percent error quantity.

VII. Modeling Results

Separate aero-propulsive models were identified using the data collected from the FCB and IOB simulated wind tunnel experiments for each sequential test block. Figure 23 shows a comparison of the modeling NRMSE (NRMSE_m) and validation NRMSE (NRMSE_v) for the sequential models developed using the FCB and IOB designs. Histories of normalized modeling residuals and normalized validation residuals, as well as validation NRMSE metric bounds ($\pm 2\,\text{NRMSE}_v$), for each block are shown in Figs. 24-27 in the Appendix to compare residual values and verify that the residuals are independent. The y-axis limits for each response across each block and design type are identical to aid in visual comparisons of residual character. Additional omitted residual diagnostic plots were viewed for each design to verify that the residuals were normally distributed and had constant variance.

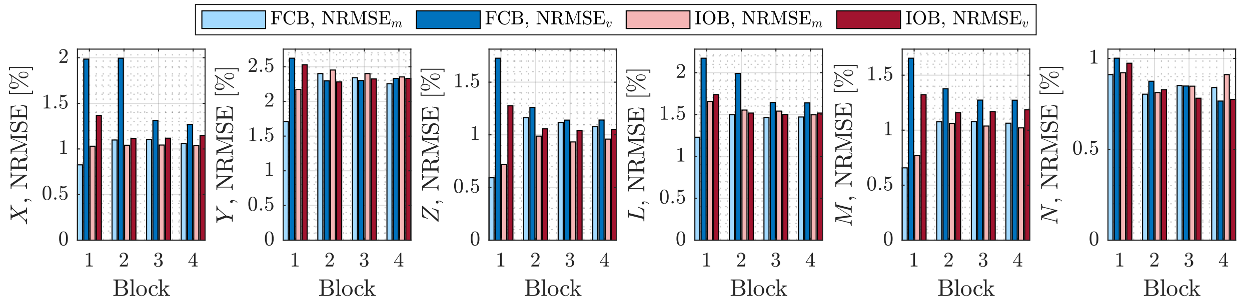


Fig. 23 Modeling and validation NRMSE for models developed at each test block for LA-8 simulated wind tunnel experiments.

As can be seen in Fig. 23, most respective IOB-design NRMSE values calculated using the modeling and validation data for each response are similar and low-valued signifying that a high-quality model has generally been identified at each block; however, for the first IOB design block, the NRMSE_v values for Z and M are notably larger than the respective NRMSE_m values suggesting that the Block 1 model for these responses may require improvement to obtain an adequate model. There is a decrease between the NRMSE_v values between Block 1 and Block 2 for all responses indicating that augmentation with the nested I-optimal design has improved the prediction capability. The NRMSE_v values for Block 3 and Block 4 are similar to those obtained for Block 2 indicating that the additional data collected in

Block 3 and Block 4 provide little additional benefit in terms of prediction capability. These results suggests that only the first two IOB test blocks need to be executed for this experiment.

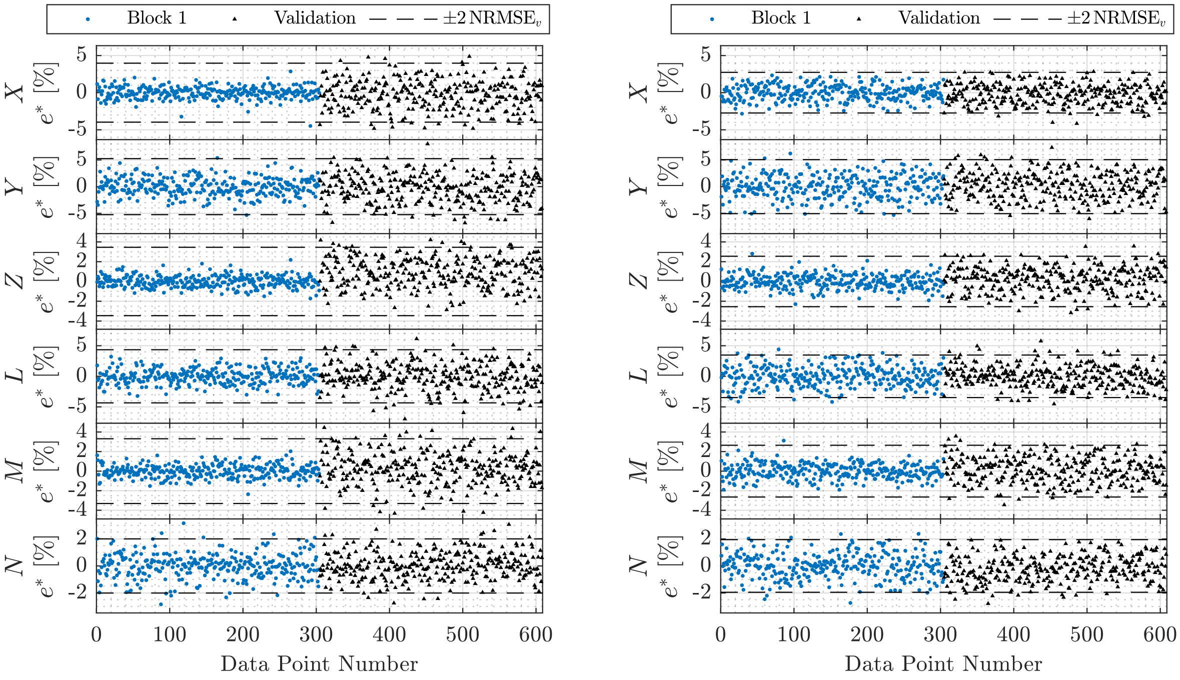
For the first FCB design block, the NRMSE_v values for all responses except N are significantly larger than the respective NRMSE_m values suggesting that the model may be deficient, which could be attributed to the sparsity of interior data points. This residual character is also clearly reflected in Fig. 24a, where the magnitude of normalized validation residuals for most responses is significantly larger than the magnitude of the respective normalized modeling residuals. Although model adequacy reservations were noted for the Block 1 IOB design for some responses, the disparity between the respective modeling and validation NRMSE values is more prominent for the FCB design. Execution of the second FCB block decreases the NRMSE_v for most responses and results in the NRMSE_v values for each response being closer to the respective NRMSE_m values; however, the NRMSE_v values are still notably larger compared to the NRMSE_m values for certain responses. For the Block 3 and Block 4 FCB models, the NRMSE_m and NRMSE_v values for each response are similar in value, providing confidence that model development was successful. The NRMSE_v values obtained for the models identified from the Block 3 and Block 4 FCB designs are also similar in value to the corresponding NRMSE_v values obtained for the Block 3 and Block 4 IOB models, but lower prediction error is still generally obtained for the models identified using the IOB block designs. I-optimal designs are optimized to reduce prediction error for the identified models, so it makes sense that the IOB design blocks result in lower prediction error values. Future studies are recommended to further investigate the utility of the presented response surface experiment design approaches in wind tunnel experimentation and other complex modeling problems with a large number of test factors.

VIII. Conclusions

Distributed propulsion aircraft configurations present new challenges for aero-propulsive modeling and are currently an important area of research. The mathematical models developed to characterize aero-propulsive phenomena for these complex aircraft must readily describe many control effectors and interactions, while also being amendable to drastically changing aerodynamics at numerous different flight conditions across a wide flight envelope. The result is a large modeling problem with significant nonlinearities and many interaction effects that requires an efficient experiment design within cost and time constraints. Design of experiments and response surface methodology theory provide efficient, statistically-rigorous experiment designs that enable accurate characterization of complex distributed propulsion aircraft.

Several experiment designs were described and evaluated for modeling the aerodynamics of a tandem tilt-wing, distributed electric propulsion aircraft in the form of a response surface model including linear, two-factor interaction, and pure quadratic model terms. Evaluation of single block foundational experiment designs revealed that I-optimal and A-optimal designs generally have the highest statistical power, lowest correlation among candidate model terms, lowest prediction variance across the design space, and yield the most precise parameter estimates. However, the single block designs were noted to place most of the design points near the exterior of the design space, which can result in model bias errors in the interior of the design space. To resolve this shortcoming, two sequential experiment designs containing nested designs that provide more coverage of the interior of the design space were investigated. The first sequential design approach used ordinary and nested minimum run resolution V, face-centered central composite designs as its base and represents a legacy design approach that has been applied previously to characterize complex aircraft. The second sequential design approach used a regular I-optimal design and a nested I-optimal design as its base with the same number of test points as the central composite designs. The nested I-optimal design was a new experiment design approach proposed in this paper for complex aircraft aero-propulsive characterization. Design evaluation metrics available prior to conducting an experiment were used to compare the two design approaches and revealed that the I-optimal base design had better statistical power for the candidate model terms, lower correlation among candidate model terms, and was expected to yield models with better prediction capability and more precise parameter estimates. Each set of sequential test matrices was applied in simulated LA-8 wind tunnel experiments to develop a model of the dimensional forces and moments exerted on the aircraft at a reference airspeed condition. Modeling results obtained from the simulated wind tunnel data suggested that the I-optimal based design provides improved prediction capability for data not used in the modeling process compared to the design blocks built off of the face-centered central composite design. The results also indicated that an adequate model was obtained earlier in the block design sequence for the I-optimal base design. Based on the design evaluation metrics and modeling results presented in this paper, the nested I-optimal design approach is recommended for future aero-propulsive characterization experiments for complex distributed propulsion aircraft.

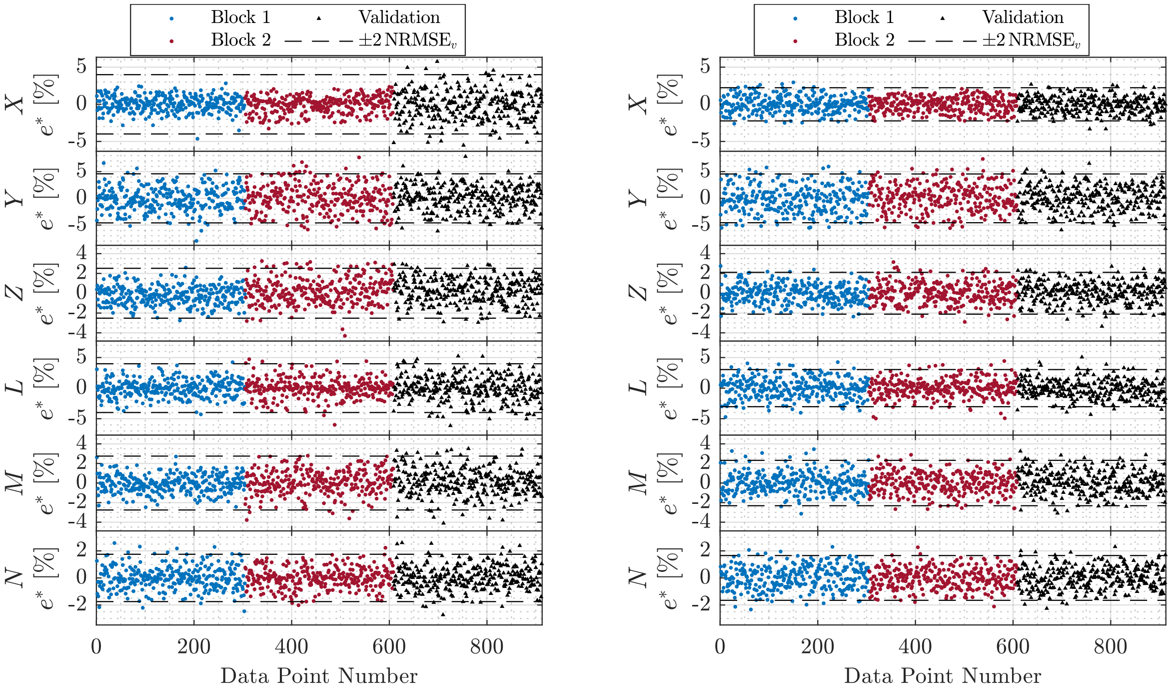
Appendix: Residual Plots



(a) FCB test matrix

(b) IOB test matrix

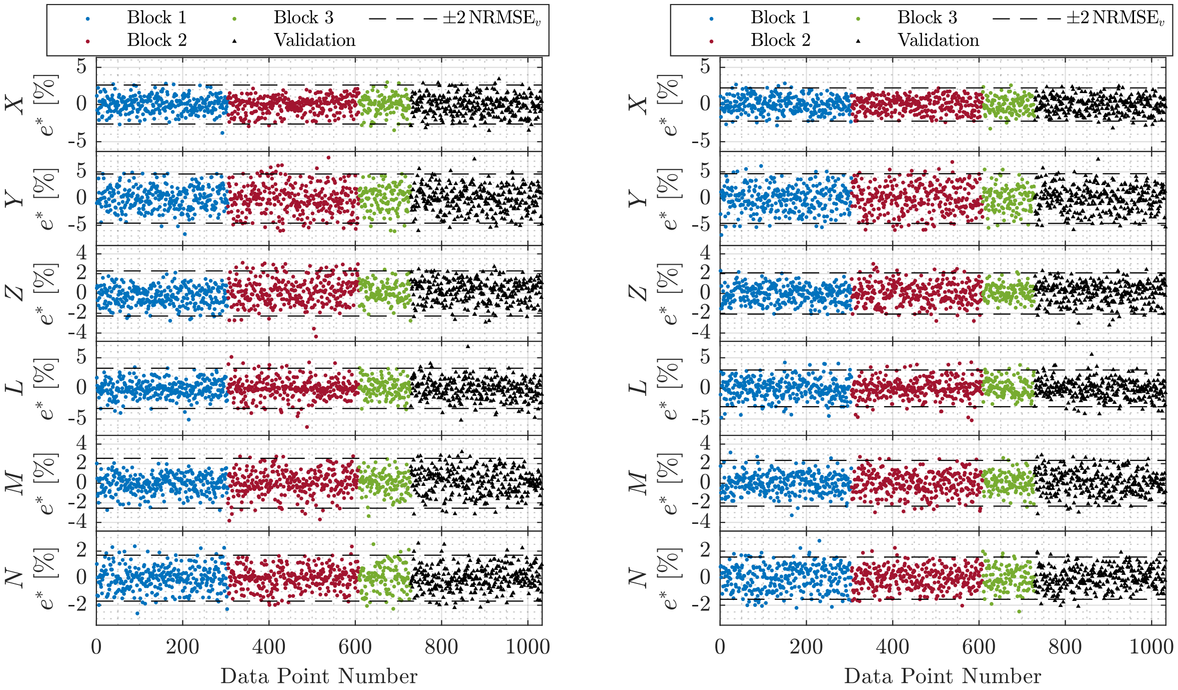
Fig. 24 Normalized modeling and prediction residuals (1 modeling block).



(a) FCB test matrix

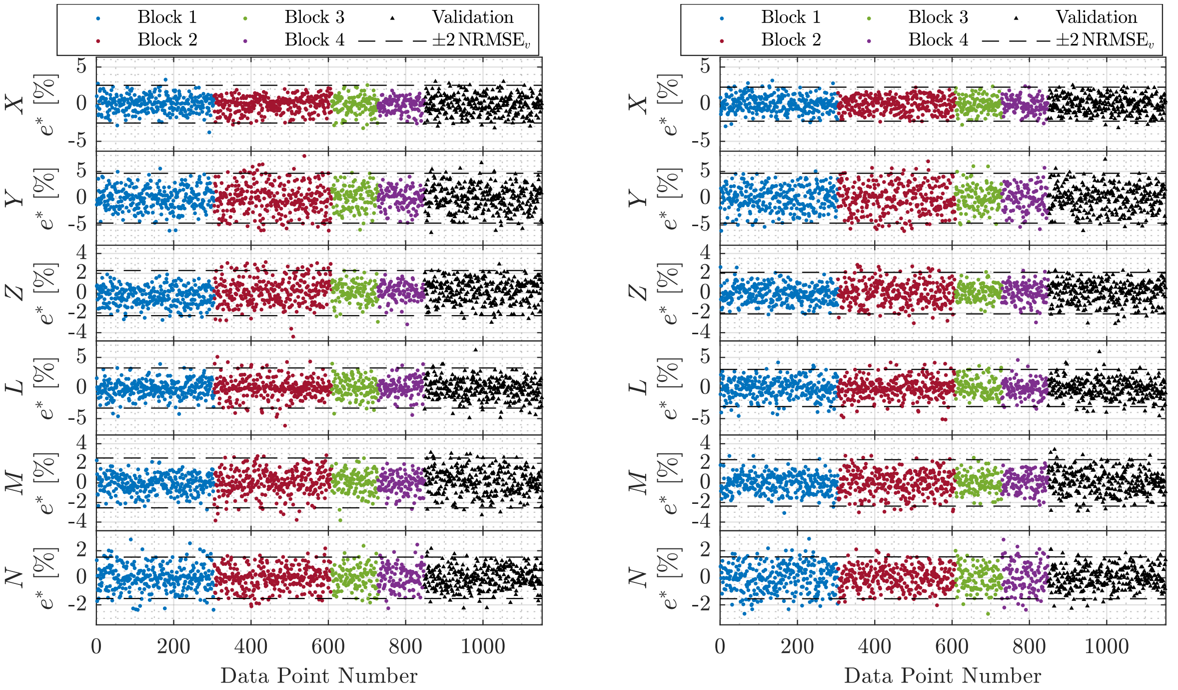
(b) IOB test matrix

Fig. 25 Normalized modeling and prediction residuals (2 modeling blocks).



(a) FCB test matrix **(b) IOB test matrix**

Fig. 26 Normalized modeling and prediction residuals (3 modeling blocks).



(a) FCB test matrix **(b) IOB test matrix**

Fig. 27 Normalized modeling and prediction residuals (4 modeling blocks).

Acknowledgments

This research was funded by the NASA Aeronautics Research Mission Directorate (ARMD) Transformational Tools and Technologies (TTT) project. The RAM block design process and LA-8 face-centered base design strategy were developed by Patrick Murphy. The LA-8 strip theory model used to conduct simulated experiments was developed by Jacob Cook. LA-8 photography support was provided by Lee Pollard.

References

[1] Johnson, W., Silva, C., and Solis, E., "Concept Vehicles for VTOL Air Taxi Operations," *AHS Technical Conference on Aeromechanics Design for Transformative Vertical Flight*, Jan. 2018.

[2] Silva, C., Johnson, W., Antcliff, K. R., and Patterson, M. D., "VTOL Urban Air Mobility Concept Vehicles for Technology Development," *2018 Aviation Technology, Integration, and Operations Conference*, AIAA Paper 2018-3847, Jun. 2018. https://doi.org/10.2514/6.2018-3847.

[3] Saeed, A. S., Younes, A. B., Cai, C., and Cai, G., "A survey of hybrid Unmanned Aerial Vehicles," *Progress in Aerospace Sciences*, Vol. 98, 2018, pp. 91–105. https://doi.org/10.1016/j.paerosci.2018.03.007.

[4] Kim, H. D., Perry, A. T., and Ansell, P. J., "A Review of Distributed Electric Propulsion Concepts for Air Vehicle Technology," *2018 AIAA/IEEE Electric Aircraft Technologies Symposium*, AIAA Paper 2018-4998, Jul. 2018. https://doi.org/10.2514/6.2018-4998.

[5] Johnson, W., and Silva, C., "NASA Concept Vehicles and the Engineering of Advanced Air Mobility Aircraft," *The Aeronautical Journal*, Vol. 126, No. 1295, 2022, pp. 59–91. https://doi.org/10.1017/aer.2021.92.

[6] "eVTOL Aircraft Directory," *Electric VTOL News™*, https://evtol.news/aircraft, Accessed 29 October 2021.

[7] Rothhaar, P. M., Murphy, P. C., Bacon, B. J., Gregory, I. M., Grauer, J. A., Busan, R. C., and Croom, M. A., "NASA Langley Distributed Propulsion VTOL Tilt-Wing Aircraft Testing, Modeling, Simulation, Control, and Flight Test Development," *14th AIAA Aviation Technology, Integration, and Operations Conference*, AIAA Paper 2014-2999, Jun. 2014. https://doi.org/10.2514/6.2014-2999.

[8] North, D. D., Busan, R. C., and Howland, G., "Design and Fabrication of the Langley Aerodrome No. 8 Distributed Electric Propulsion VTOL Testbed," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1188, Jan. 2021. https://doi.org/10.2514/6.2021-1188.

[9] McSwain, R. G., Geuther, S. C., Howland, G., Patterson, M. D., Whiteside, S. K., and North, D. D., "An Experimental Approach to a Rapid Propulsion and Aeronautics Concepts Testbed," NASA TM–2020-220437, Jan. 2020.

[10] Geuther, S. C., and Fei, X., "LA-8 Computational Analysis and Validation Studies Using FlightStream," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1191, Jan. 2021. https://doi.org/10.2514/6.2021-1191.

[11] Cooper, J. R., Ackerman, K. A., Rothhaar, P. M., and Gregory, I. M., "Autonomous Path-Following for a Tilt-Wing, Distributed Electric Propulsion, Vertical Take-Off and Landing Unmanned Aerial System in Hover Mode," NASA TM–2018–220109, Nov. 2018.

[12] Cook, J., and Gregory, I., "A Robust Uniform Control Approach for VTOL Aircraft," *VFS Autonomous VTOL Technical Meeting and Electric VTOL Symposium*, Jan. 2021.

[13] Busan, R. C., Rothhaar, P. M., Croom, M. A., Murphy, P. C., Grafton, S. B., and O'Neal, A. W., "Enabling Advanced Wind-Tunnel Research Methods Using the NASA Langley 12-Foot Low Speed Tunnel," *14th AIAA Aviation Technology, Integration, and Operations Conference*, AIAA Paper 2014-3000, Jun. 2014. https://doi.org/10.2514/6.2014-3000.

[14] Geuther, S. C., North, D. D., and Busan, R. C., "Investigation of a Tandem Tilt-wing VTOL Aircraft in the NASA Langley 12-Foot Low-Speed Tunnel," NASA TM–2020–5003178, Jun. 2020.

[15] Busan, R. C., Murphy, P. C., Hatke, D. B., and Simmons, B. M., "Wind Tunnel Testing Techniques for a Tandem Tilt-Wing, Distributed Electric Propulsion VTOL Aircraft," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1189, Jan. 2021. https://doi.org/10.2514/6.2021-1189.

[16] McSwain, R. G., Glaab, L. J., and Theodore, C. R., "Greased Lightning (GL-10) Performance Flight Research – Flight Data Report," NASA TM–2017–219794, Nov. 2017.

[17] Fredericks, W. J., McSwain, R. G., Beaton, B. F., Klassman, D. W., and Theodore, C. R., "Greased Lightning (GL-10) Flight Testing Campaign," NASA TM–2017–219643, Jul. 2017.

[18] North, D. D., "Flight Testing of a Scale Urban Air Mobility Technology Testbed," *AIAA SciTech 2021 Forum*, AIAA Presentation, Jan. 2021. https://ntrs.nasa.gov/citations/20205010998, Accessed 17 October 2022.

[19] Simmons, B. M., "System Identification for eVTOL Aircraft Using Simulated Flight Data," *AIAA SciTech 2022 Forum*, AIAA Paper 2022-2409, Jan. 2022. https://doi.org/10.2514/6.2022-2409.

[20] Murphy, P. C., and Landman, D., "Experiment Design for Complex VTOL Aircraft with Distributed Propulsion and Tilt Wing," *AIAA Atmospheric Flight Mechanics Conference*, AIAA Paper 2015-0017, Jan. 2015. https://doi.org/10.2514/6.2015-0017.

[21] Simmons, B. M., and Murphy, P. C., "Aero-Propulsive Modeling for Tilt-Wing, Distributed Propulsion Aircraft Using Wind Tunnel Data," *Journal of Aircraft*, Vol. 59, No. 5, 2022, pp. 1162–1178. https://doi.org/10.2514/1.C036351.

[22] Simmons, B. M., Morelli, E. A., Busan, R. C., Hatke, D. B., and O'Neal, A. W., "Aero-Propulsive Modeling for eVTOL Aircraft Using Wind Tunnel Testing with Multisine Inputs," *AIAA AVIATION 2022 Forum*, AIAA Paper 2022-3603, Jun. 2022. https://doi.org/10.2514/6.2022-3603.

[23] Montgomery, D. C., *Design And Analysis of Experiments*, 8th ed., John Wiley & Sons, Inc., Hoboken, NJ, 2013.

[24] Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 4th ed., John Wiley & Sons, 2016.

[25] Murphy, P. C., Hatke, D. B., Aubuchon, V. V., Weinstein, R., and Busan, R. C., "Preliminary Steps in Developing Rapid Aero Modeling Technology," *AIAA SciTech 2020 Forum*, AIAA Paper 2020-0764, Jan. 2020. https://doi.org/10.2514/6.2020-0764.

[26] Murphy, P. C., Buning, P. G., and Simmons, B. M., "Rapid Aero Modeling for Urban Air Mobility Aircraft in Computational Experiments," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1002, Jan. 2021. https://doi.org/10.2514/6.2021-1002.

[27] Murphy, P. C., Simmons, B. M., Hatke, D. B., and Busan, R. C., "Rapid Aero Modeling for Urban Air Mobility Aircraft in Wind-Tunnel Tests," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1644, Jan. 2021. https://doi.org/10.2514/6.2021-1644.

[28] "NASA Langley 12-Foot Low-Speed Tunnel," https://researchdirectorate.larc.nasa.gov/12-foot-low-speed-tunnel-12-ft-lst/, Accessed 15 October 2022.

[29] Cook, J. W., and Hauser, J., "A Strip Theory Approach to Dynamic Modeling of eVTOL Aircraft," *AIAA SciTech 2021 Forum*, AIAA Paper 2021-1720, Jan. 2021. https://doi.org/10.2514/6.2021-1720.

[30] Simmons, B. M., "System Identification for Propellers at High Incidence Angles," *Journal of Aircraft*, Vol. 58, No. 6, 2021, pp. 1336–1350. https://doi.org/10.2514/1.C036329.

[31] "Design-Expert®," Version 13, Stat-Ease, Inc., https://www.statease.com/software/design-expert/, Accessed 11 May 2022.

[32] Goos, P., and Jones, B., *Optimal Design of Experiments*, John Wiley & Sons, West Sussex, United Kingdom, 2011.

[33] Anderson, M. J., and Whitcomb, P. J., *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*, 2nd ed., CRC Press, Boca Raton, FL, 2017. https://doi.org/10.1201/9781315382326.

[34] Anderson, M. J., and Whitcomb, P. J., *DOE Simplified: Practical Tools for Effective Experimentation*, 3rd ed., CRC Press, Boca Raton, FL, 2015. https://doi.org/10.1201/b18479.

[35] Morelli, E. A., and Klein, V., *Aircraft System Identification: Theory and Practice*, 2nd ed., Sunflyte Enterprises, Williamsburg, VA, 2016.

[36] Jategaonkar, R. V., *Flight Vehicle System Identification: A Time-Domain Methodology*, 2nd ed., American Institute of Aeronautics and Astronautics, Reston, VA, 2015. https://doi.org/10.2514/4.102790.

[37] Montgomery, D. C., Peck, E. A., and Vining, G. G., *Introduction to Linear Regression Analysis*, 5th ed., John Wiley & Sons, Hoboken, New Jersey, 2012.

[38] Belsley, D. A., Kuh, E., and Welsch, R. E., *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, Hoboken, NJ, 2004.

[39] "Stat-Ease Handbook for Experimenters," Version 13.00, Stat-Ease, Inc., Minneapolis, MN, Dec. 2020.

[40] Zahran, A., Anderson-Cook, C. M., and Myers, R. H., "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs," *Journal of Quality Technology*, Vol. 35, No. 4, 2003, pp. 377–386. https://doi.org/10.1080/00224065.2003.11980235.

[41] Anderson, M. J., Adams, W. F., and Whitcomb, P. J., "How to Properly Size Response Surface Method Experiment (RSM) Designs for System Optimization," 2016. https://cdn.statease.com/media/public/documents/how_to_properly_size_experiments_aimed_at_system_optimization.pdf, Accessed 17 October 2022.

[42] Whitcomb, P., "FDS—A Power Tool for Designers of Optimization Experiments," *Stat-Teaser Newsletter*, Stat-Ease, Inc., Sep. 2008, pp. 1–3. https://cdn.statease.com/media/public/documents/statteaser-0908.pdf, Accessed 17 October 2022.

[43] Simmons, B. M., Buning, P. G., and Murphy, P. C., "Full-Envelope Aero-Propulsive Model Identification for Lift+Cruise Aircraft Using Computational Experiments," *AIAA AVIATION 2021 Forum*, AIAA Paper 2021-3170, Aug. 2021. https://doi.org/10.2514/6.2021-3170.

[44] Landman, D., Simpson, J., Mariani, R., Ortiz, F., and Britcher, C., "Hybrid Design for Aircraft Wind-Tunnel Testing Using Response Surface Methodologies," *Journal of Aircraft*, Vol. 44, No. 4, 2007, pp. 1214–1221. https://doi.org/10.2514/1.25914.

[45] "System IDentification Programs for AirCraft (SIDPAC)," *NASA Technology Transfer Program*, https://software.nasa.gov/software/LAR-16100-1, Accessed 03 September 2022.

[46] Klein, V., Batterson, J. G., and Murphy, P. C., "Determination of Airplane Model Structure from Flight Data by Using Modified Stepwise Regression," NASA TP-1916, Oct. 1981.