

MULTIDISCIPLINARY PHYSICS-CONSTRAINED GENERATIVE DESIGN OF DISTRIBUTED ELECTRIC PROPULSION AIRCRAFT

Wenyu Wei[1,2], Yiyuan Ma[1,2,*] & Zhonghua Han[1,2]

[1]Institute of Aerodynamic and Multidisciplinary Design Optimization, School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072 Shaanxi, People's Republic of China
[2]National Key Laboratory of Aircraft Configuration Design, Xi'an 710072, People's Republic of China

Abstract

Early-stage distributed electric propulsion (DEP) aircraft design requires simultaneous treatment of geometry, propulsion layout, battery topology, and mission variables while retaining diverse candidate families for later assessment. This paper formulates DEP parameter synthesis as a mixed-variable, physics-constrained distribution-learning problem. An analysis-supervised differentiable physical-response model is embedded into a continuous/categorical adversarial generator through engineering constraints, soft performance preferences, and diversity regularization. On grouped validation and held-out test splits, the selected three-member physical-response ensemble achieves R^2 ranges of 0.9345–0.9891 and 0.9174–0.9884, respectively. In a 10000-sample audit, the selected generator produces 9956 designs satisfying the model-based feasibility screen, with a mean constraint loss of 1.97×10^{-3} and broad discrete support, including propeller blade counts from 2 to 10. Representative generated families show compact, dense-propulsion, long-range, low-noise, high-altitude, and alternate-configuration characteristics. The results suggest that differentiable physical guidance and explicit distributional regularization can support DEP concept generation while exposing the trade-off between predicted performance and design diversity.

Keywords: Distributed electric propulsion; generative design; physics-constrained generative model; multidisciplinary optimization; design diversity

1. Introduction

Distributed electric propulsion (DEP) aircraft integrate propulsion, lift augmentation, energy management, and noise control within a single aircraft-level design problem. Flight-scale DEP studies indicate that distributed propulsors should be treated as elements of an integrated aircraft concept rather than as isolated propulsion components [1]. Propulsor placement and operating condition can also act as aerodynamic design variables through aero-propulsive interaction, especially in high-lift configurations [2]. DEP design is therefore more than propulsion sizing: it involves continuous and discrete decisions across aircraft geometry, propulsion layout, energy storage, and mission definition. Many combinations that satisfy simple bounds can still become low-performing or infeasible after multidisciplinary evaluation.

The system-level nature of this challenge is also evident in electric and hybrid-electric aircraft studies. Architecture selection, component modeling, propulsion-airframe coupling, and mission analysis are tightly connected in fixed-wing electric-propulsion concepts [3]. Energy storage, electrical conversion, thermal management, and motor sizing further couple the aircraft-level performance and feasibility margins [4]. Battery specific energy and pack-level mass remain central constraints on mission range and reserve for all-electric aircraft [5], while propulsor number and arrangement influence aerodynamic, structural, reliability, and energy behavior [6]. These factors suggest that early-stage

DEP synthesis is better formulated as a mixed-variable system-design problem than as a sequence of separate aerodynamic, propulsion, and battery-sizing tasks.

Multidisciplinary design analysis and optimization (MDAO) provides an established way to organize such coupled aircraft-design problems. For unconventional civil-aircraft concepts, recurring challenges include system-level integration, early geometry definition, and credible low-to-high-fidelity analysis [7]. Collaborative MDAO workflows further indicate that distributed disciplinary execution and consistent data exchange are important for coupled wing and nacelle design tasks [8]. In aerodynamic shape optimization, data-driven response models can reduce repeated evaluation cost, although accuracy and extrapolation remain important concerns [9]. For DEP aircraft, these concerns become more pronounced because categorical choices such as propulsor count, blade count, airfoil family, and battery topology interact with continuous geometry and mission variables.

Generative modeling provides a complementary perspective to conventional optimization because it seeks a distribution of candidate designs rather than a single solution. Adversarial generative modeling formulates distribution learning through coupled generator–discriminator training [10]. Wasserstein objectives can improve the stability of adversarial training [11], and gradient-penalty regularization further constrains the critic during this process [12]. In aerodynamic design, Xie et al. [13] developed parametric generative schemes with geometric constraints for airfoil synthesis. Wang et al. [14] proposed a conditional generative inverse-design method for supercritical airfoils. Jin et al. [15] applied conditional generative modeling to aerodynamic and stealth-oriented airfoil design. Wada et al. [16] introduced physics-guided training for airfoil synthesis.

Recent work has begun to extend generative methods beyond isolated two-dimensional shapes. Shu et al. [17] investigated three-dimensional generative design with physics-based validation. Nobari et al. [18] studied range-constrained generative synthesis for bounded design spaces. Sisk and Du [19] developed physics-constrained generative AI for aircraft takeoff trajectory design. However, DEP aircraft parameter synthesis poses a broader mixed-variable requirement: a generator must coordinate aircraft geometry, propulsion categories, battery topology, and mission variables while remaining sensitive to multidisciplinary responses. This requirement is only partly covered by methods that focus mainly on a single geometric object, a prescribed design family, or a narrow operating task.

The above literature motivates two methodological requirements for DEP generative design. First, feasibility should be evaluated through response quantities such as lift-to-drag ratio, trim behavior, battery reserve, thermal response, and noise, rather than only through variable bounds. Second, physical guidance should be balanced with distributional diversity, because a generator that collapses to one predicted high-performing family provides limited value for early-stage concept exploration. The central question considered in this paper is therefore whether a generator can learn a broad DEP design distribution while being guided toward physically reasonable regions of the design space.

This study develops a workflow for DEP aircraft parameter synthesis around this question. A differentiable multi-task physical-response model is learned from curated analysis data and embedded into the generative objective, so that candidate designs receive multidisciplinary guidance during distribution learning. The generator, critic, constraint module, and physical objective are coupled during training, while generated designs are examined through parameter decoding, response prediction, and geometric assessment. The intent is not to replace final independent verification, but to provide a structured way to generate, screen, and interpret candidate families before expensive downstream refinement.

The contributions of the work are summarized as follows.

- A mixed-variable DEP representation that jointly describes aircraft geometry, propulsion layout, energy storage, and mission variables under explicit engineering constraints.
- An analysis-supervised differentiable physical-response model that provides aerodynamic, trim, battery, acoustic, and constraint guidance inside generative training.
- A diversity-aware generative strategy that combines categorical distribution matching, entropy regularization, model-based screening, and geometric consistency assessment to examine the performance-diversity trade-off.

2. Method

2.1 Problem formulation and design representation

The method is organized around three requirements that are specific to DEP aircraft synthesis. First, the design representation should contain heterogeneous decisions: continuous geometry variables, discrete propulsion and battery choices, airfoil categories, and mission variables. Second, the generator benefits from physical guidance during training, because simple variable bounds cannot express whether a coupled aircraft-propulsion-energy configuration is useful. Third, the learned distribution should remain diverse, because a generative model that collapses to one high-scoring family provides limited value for early-stage aircraft design.

The workflow, summarized in Figure 1, follows these requirements in five ordered steps. First, a mixed-variable DEP design schema defines the admissible aircraft, propulsion, battery, and mission parameters. Second, multidisciplinary analyses provide supervised response labels for feasible samples, so that the learning problem is tied to physically evaluated aircraft rather than to geometry alone. Third, an end-to-end physical-response model approximates the response map in differentiable form. Fourth, a mixed continuous/categorical adversarial generator uses this response model, together with explicit constraint and diversity losses, to learn a design distribution. Finally, generated samples are decoded and assessed with a fixed evaluation protocol. This ordering makes four roles explicit: the design schema defines what may be generated, the physical-response model supplies differentiable guidance, the generator defines a candidate distribution, and the downstream audit checks the decoded designs.

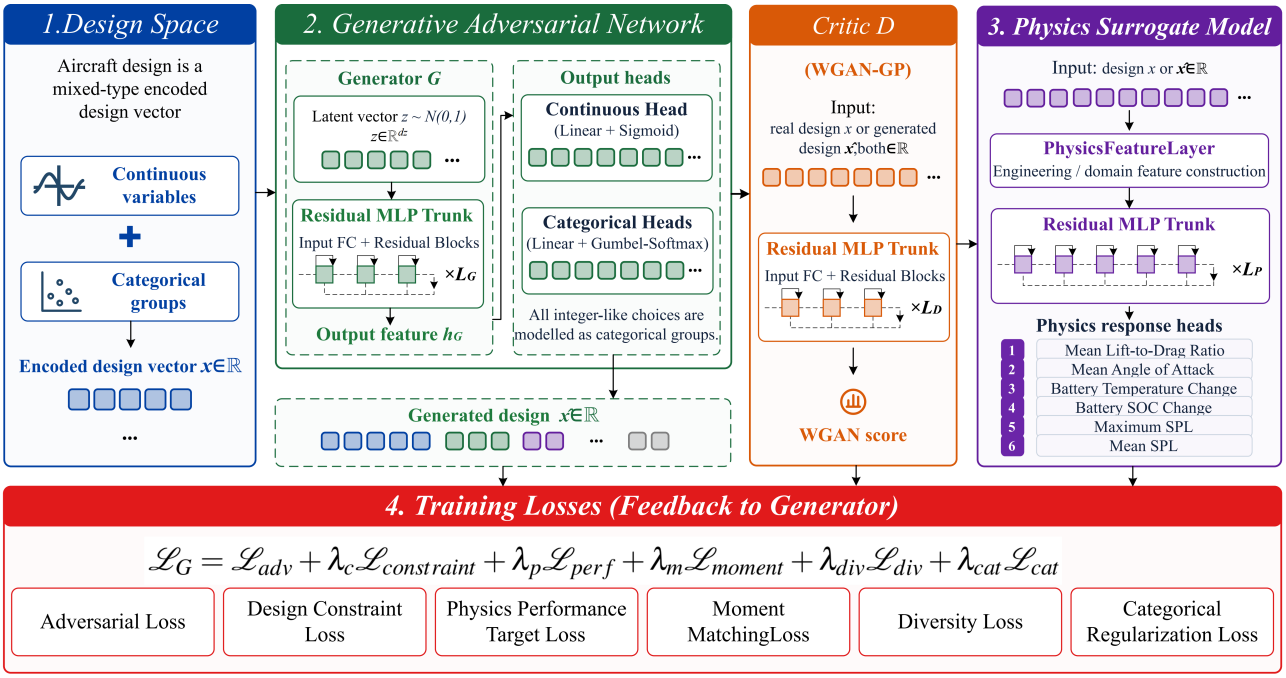


Figure 1 – Physics-constrained generative-design workflow and model architecture.

Each sample is represented as a structured parameter record. Throughout the paper, bold lowercase symbols denote vectors, calligraphic symbols denote sets or loss families, hats denote model predictions, and tildes denote generated encoded samples. The design record is partitioned into continuous aircraft and propulsion variables $\mathbf{x}_c$, discrete categories $\mathbf{x}_d$, and explicit mission variables $\mathbf{m}$. Including mission variables in the design record is important for DEP aircraft because the same geometry can have substantially different battery and acoustic responses when range, altitude, or cruise speed changes. The mission variables are treated as part of the generated design condition rather than as hidden operating assumptions.

Table 1 lists the independent design variables used in the present schema. The schema is intentionally mixed-variable. Propulsor count, blade count, airfoil identity, and battery topology are kept as discrete choices because rounding them after continuous generation would create ambiguous or

physically unintended designs. In contrast, wing shape, propulsor position, diameter, thrust setting, and mission variables are represented as bounded continuous quantities. In the table, $\mathscr{A}_i$ and $\mathscr{A}_t$ denote screened internal- and tip-propeller airfoil subsets drawn from the UIUC Airfoil Coordinates Database. The screening removes records that do not satisfy the propeller-section requirements used in this study, including complete coordinate availability, non-degenerate geometry, and compatibility with the propeller geometry generator. Propeller airfoil names are categorical variables; their CST coefficients are deterministic attributes of the selected airfoil records and are written to the analysis files, but they are not counted as separate independent variables. Categorical variables are encoded as one-hot groups during training and decoded back to valid engineering categories for design realization.

Table 1 – Independent design variables.

Group	Variables	Admissible domain
Propulsion layout	Configuration mode c_m; number of propulsors per side N_s	$c_m \in \{0,1\}$; $N_s \in \{2,3,4,5\}$
Internal propulsors	Longitudinal, lateral, and vertical locations x_i, y_i, z_i; lateral spacing Δy_i; diameter D_i; design thrust T_i; hub radius $r_{h,i}$; design angular velocity ω_i; design lift coefficient $C_{l,i}^{des}$; blade count B_i; airfoil category a_i	$x_i \in [1.5, 2.8]$ m, $y_i \in [0.2, 1.5]$ m, $\Delta y_i \in [0.6, 1.8]$ m, $z_i \in [0.6, 1.3]$ m, $D_i \in [0.4, 2.0]$ m, $T_i \in [200, 1800]$ N, $r_{h,i} \in [0.12, 0.30]$ m, $\omega_i \in [250, 350]$ rad/s, $C_{l,i}^{des} \in [0.5, 0.9]$, $B_i \in \{2, \ldots, 10\}$, $a_i \in \mathscr{A}_i \subset \mathscr{A}_{UIUC}$
Tip propulsors	Longitudinal, lateral, and vertical locations x_t, y_t, z_t; diameter D_t; design thrust T_t; hub radius $r_{h,t}$; design angular velocity ω_t; design lift coefficient $C_{l,t}^{des}$; blade count B_t; airfoil category a_t	$x_t \in [2.0, 2.5]$ m, $y_t \in [4.0, 5.2]$ m, $z_t \in [0.9, 1.2]$ m, $D_t \in [0.4, 1.7]$ m, $T_t \in [500, 750]$ N, $r_{h,t} \in [0.10, 0.25]$ m, $\omega_t \in [280, 380]$ rad/s, $C_{l,t}^{des} \in [0.5, 0.9]$, $B_t \in \{2, \ldots, 10\}$, $a_t \in \mathscr{A}_t \subset \mathscr{A}_{UIUC}$
Wing planform	Projected span b; root and tip chords c_r, c_t; inboard twist, dihedral, and quarter-chord sweep $\theta_{in}, \Gamma_{in}, \Lambda_{in}$; outboard span station, twist, root-chord ratio, dihedral, and sweep $\eta_o, \theta_o, \kappa_o, \Gamma_o, \Lambda_o$; winglet span station, twist, root-chord ratio, dihedral, and sweep $\eta_w, \theta_w, \kappa_w, \Gamma_w, \Lambda_w$; tip twist and root-chord ratio $\theta_{tip}, \kappa_{tip}$	$b \in [10.8, 11.8]$ m, $c_r \in [1.35, 1.55]$ m, $c_t \in [0.85, 1.00]$ m; $\theta_{in} \in [1.0, 4.0]^\circ$, $\Gamma_{in} \in [-1.0, 1.0]^\circ$, $\Lambda_{in} \in [-2.0, 2.0]^\circ$; $\eta_o \in [0.45, 0.65]$, $\theta_o \in [0.5, 3.5]^\circ$, $\kappa_o \in [0.70, 0.95]$, $\Gamma_o \in [-1.0, 1.0]^\circ$, $\Lambda_o \in [-2.0, 2.0]^\circ$; $\eta_w \in [0.96, 0.99]$, $\theta_w \in [-1.0, 2.0]^\circ$, $\kappa_w \in [0.50, 0.75]$, $\Gamma_w \in [65, 85]^\circ$, $\Lambda_w \in [25, 40]^\circ$; $\theta_{tip} \in [-1.0, 1.0]^\circ$, $\kappa_{tip} \in [0.08, 0.20]$
Wing airfoils	CST upper and lower coefficients $A_{s,k}^u, A_{s,k}^l$ for $s \in \{\text{inboard}, \text{outboard}, \text{winglet}\}$ and $k = 0, \ldots, 4$ [20]	Upper: 0.1463, 0.2361, 0.1602, 0.3201, 0.0649; lower: -0.1198, -0.0818, -0.1610, -0.0359, 0.0384; all ± 0.12
Battery topology	Battery module count M_b; series count N_{ser}; parallel count N_{par}	$M_b \in \{6, \ldots, 12\}$, $N_{ser} \in \{12, \ldots, 24\}$, $N_{par} \in \{25, \ldots, 50\}$
Mission	Cruise distance R; cruise altitude h; cruise speed V	$R \in [8, 200]$ nmi, $h \in [5000, 12000]$ ft, $V \in [110, 135]$ mph

In Table 1, the configuration mode c_m is a binary propulsion-layout topology variable. When $c_m = 0$, all N_s propulsors on each side are treated as internal distributed propulsors. When $c_m = 1$, each side contains $N_s - 1$ internal propulsors and one tip propulsor, so the tip-propulsor position, diameter, blade, and airfoil variables become active in the decoded layout.

Rather than solving a deterministic single-design optimization problem, the method learns a feasible design distribution. The engineering-feasible set, denoted by Ω, is defined over the full mixed-variable domain in Table 1 and includes geometric compatibility, propulsion operating limits, battery-pack limits, and mission-consistency requirements. Constraint residuals are written so that non-positive values indicate satisfaction, and their positive parts are aggregated into differentiable penalties during training. This design choice makes the constraints active during distribution learning, instead of using them only to discard samples after generation.

The multidisciplinary quality measure combines lift-to-drag ratio, trim behavior, battery temperature rise, final state of charge (SOC), and acoustic response. The final SOC is computed from the normalized initial value and the predicted change in SOC; thus a negative ΔSOC represents battery depletion, and SOC_f denotes the remaining normalized charge at the end of the mission. The desired generated distribution is encouraged to remain close to the analysis-labeled design distribution, shift probability mass toward favorable feasible regions, and retain sufficient entropy to avoid collapse to a single design family. In practice, this means that the method does not seek one numerically top-

ranked aircraft. It seeks a population of aircraft that remains plausible under the design schema while being biased toward useful multidisciplinary behavior.

For each accepted multidisciplinary analysis case, the supervised response vector contains six labels: mission-averaged lift-to-drag ratio $\overline{L/D}$, average angle of attack $\overline{\alpha}$, battery temperature change ΔT_b, battery SOC change ΔSOC, maximum sound pressure level $SPL_{\max}$, and average sound pressure level $\overline{SPL}$. These labels cover the main couplings needed in the present study: aerodynamic performance, trim consistency, energy usage, thermal response, and noise. They are used only as response labels for model training and evaluation, not as input design variables.

The workflow distinguishes training-time differentiable guidance from evaluation-time assessment. During training, constraints and physical-response losses provide gradients to the generator. During evaluation, generated records are decoded from the declared schema and passed through the same analysis and geometry-realization loaders. This separation is important because a differentiable model can guide sampling efficiently, whereas final interpretation still requires decoded engineering records and independent audit quantities.

2.2 Differentiable physical-response model

The physical-response model provides the differentiable link between a generated design and its expected multidisciplinary behavior. Direct analysis inside every generator update would be computationally inefficient and non-differentiable with respect to the generator parameters. The model learns the analysis-label-scale response map from curated samples and is then embedded as a fixed evaluator during generator training. Its role is guidance and screening, not replacement of final engineering verification.

The model maps the encoded design vector $\mathbf{e}$ to the six predicted responses denoted by $\hat{\mathbf{y}}$. Continuous variables are normalized using schema bounds. Categorical variables are represented by one-hot vectors for supervised training and by soft Gumbel-Softmax vectors when the model is called inside the generator [21, 22]. The architecture uses a differentiable engineering-feature layer, layer normalization, residual multi-layer perceptron blocks, and multiple response heads. The multiple-head structure allows shared low-level representation while preserving target-specific response behavior. When ensemble guidance is used, F_{ens} denotes the average prediction of K independently initialized physical-response models; the value of K is reported with the model results.

The residual tabular blocks use layer normalization, expansion to twice the hidden width, GELU activation, dropout, projection back to the original width, and residual addition. The engineering-feature layer includes differentiable proxies for propeller disk loading, tip Mach number, advance ratio, blade-passing frequency, battery voltage, energy capacity, shaft-power demand, C-rate, wing loading, and acoustic loading. For example, disk loading is derived from thrust divided by propeller disk area, tip Mach number is derived from angular velocity, diameter, and local speed of sound, and pack-level energy and voltage are derived from module, series, and parallel counts. These quantities are not substituted for analysis outputs. They act as differentiable inductive biases that help the neural model learn the analysis response mapping.

The physics model is trained only against analysis labels. Numerically sensitive outputs are transformed before applying the weighted multi-task regression loss:

$$\mathscr{L}_{\text{phys}} = \sum_{j=1}^{6} w_j \left\| \frac{\hat{t}_j - t_j}{s_j + \varepsilon} \right\|_2^2 + \lambda_{\text{smooth}} \|\nabla_{\mathbf{e}} \hat{\mathbf{y}}\|_2^2 + \lambda_{\text{mono}} \mathscr{L}_{\text{mono}}, \tag{1}$$

where j indexes the six response targets, w_j is the target weight, $t_j = g_j(y_j)$ is the transformed target, $\hat{t}_j$ is the corresponding transformed prediction, g_j is a target-specific numerical transform, s_j is a scale factor, ε prevents division by zero, $\nabla_{\mathbf{e}} \hat{\mathbf{y}}$ is the response Jacobian with respect to the encoded input, and λ_{smooth} and λ_{mono} weight the smoothness and monotonicity penalties. The first term fits the supervised labels, the smoothness term discourages unrealistically sharp local response changes, and $\mathscr{L}_{\text{mono}}$ weakly encourages physically plausible monotonic trends for battery and acoustic quantities. The reported R^2 values are computed in original physical units.

2.3 Physics-constrained generative model

The generative model is constructed as a constrained distribution learner. The generator maps Gaussian latent vectors to encoded mixed-variable design records. Continuous variables are decoded through bounded sigmoid mappings using the schema limits in Table 1. Categorical variables use a Gumbel-Softmax relaxation during training and are decoded to valid discrete categories during design realization. This construction keeps the generator differentiable while ensuring that exported records remain within the declared schema.

Training combines an adversarial distribution-matching term with engineering and physical objectives. The adversarial component is the Wasserstein generative adversarial network with gradient penalty (WGAN-GP) [12]:

$$\mathscr{L}_D = \mathbb{E}_{\tilde{\mathbf{e}}\sim p_G}[D_\psi(\tilde{\mathbf{e}})] - \mathbb{E}_{\mathbf{e}\sim p_{\text{data}}}[D_\psi(\mathbf{e})] + \lambda_{\text{gp}}\mathbb{E}_{\hat{\mathbf{e}}}\left[(\|\nabla_{\hat{\mathbf{e}}} D_\psi(\hat{\mathbf{e}})\|_2 - 1)^2\right]. \tag{2}$$

Here $\mathscr{L}_D$ is the critic loss, D_ψ is the critic with trainable parameters ψ, p_G is the generated encoded-design distribution, p_{data} is the encoded training-data distribution, $\hat{\mathbf{e}}$ is sampled along straight-line interpolations between real and generated encoded designs, and λ_{gp} is the gradient-penalty weight. This term encourages generated samples to remain statistically close to the analysis-labeled design population.

The constraint module then restricts the adversarial sampler to engineering-plausible regions. It imposes differentiable penalties on aircraft geometry, propeller operating point, battery pack, and mission consistency. For example, propeller disk loading and battery constraints are written as

$$\mathscr{L}_{DL} = \left[\frac{DL_{\min} - DL}{s_{DL}}\right]_+^2 + \left[\frac{DL - DL_{\max}}{s_{DL}}\right]_+^2, \tag{3}$$

$$\mathscr{L}_{\text{bat}} = \ell(U_{\text{pack}}; U_{\min}, U_{\max}) + \ell(E_{\text{cap}}; E_{\min}, E_{\max}) + \ell(\rho_E; \rho_{\min}, \rho_{\max}) + \ell(C; C_{\min}, C_{\max}), \tag{4}$$

where $[x]_+ = \max(x, 0)$, $DL_{\min}$ and $DL_{\max}$ are lower and upper disk-loading limits, and s_{DL} is a disk-loading scale. The function $\ell(q; q_{\min}, q_{\max}) = [(q_{\min} - q)/s_q]_+^2 + [(q - q_{\max})/s_q]_+^2$ is a squared interval penalty with scale s_q. In Equation (4), $U_{\min}$ and $U_{\max}$ bound pack voltage, $E_{\min}$ and $E_{\max}$ bound pack energy capacity, ρ_E is a mission energy ratio with limits $\rho_{\min}$ and $\rho_{\max}$, and C is a C-rate proxy with limits $C_{\min}$ and $C_{\max}$. Blade counts are allowed in the range 2–10. These penalties give the generator continuous feedback near constraint boundaries, which is more informative than a binary accept-or-reject rule.

The generator evaluates the fixed analysis-supervised physical-response ensemble inside the generator loss. Thus $\hat{\mathbf{y}}$ in the generator objective is the differentiable ensemble prediction of the generated design response, not a newly evaluated independent-analysis label. The physical-performance term combines feasibility and soft optimization,

$$\mathscr{L}_{\text{perf}} = \sum_k \omega_k [\max(0, g_k(\hat{\mathbf{y}}))]^2 - \eta\, \mathscr{J}(\hat{\mathbf{y}}), \tag{5}$$

where $g_k(\hat{\mathbf{y}}) \le 0$ denotes satisfaction of response-level feasibility condition k, ω_k is the corresponding penalty weight, η controls the strength of quality-oriented optimization inside the feasible region, and $\mathscr{J}$ is a scalar utility that encourages higher lift-to-drag ratio L/D, reasonable angle of attack, lower battery temperature rise, appropriate final SOC, and lower acoustic levels. The utility term is deliberately soft: it biases the distribution toward better predicted aircraft while avoiding a hard selection of only the numerically top-ranked samples.

Because large physical-objective weights can compress a generator toward a narrow family, diversity is treated as an explicit part of the method. To discourage categorical mode collapse, the generator includes a categorical marginal and entropy loss,

$$\mathscr{L}_{cat} = \frac{1}{K_c}\sum_{k=1}^{K_c}\left[\|\bar{p}_{G,k} - \bar{p}_{data,k}\|_2^2 + 1 - \frac{H(\bar{p}_{G,k}^{\Omega})}{\log|\Omega_k|}\right], \tag{6}$$

where $\mathscr{L}_{cat}$ is the categorical regularization loss, K_c is the number of categorical groups, $\bar{p}_{G,k}$ and $\bar{p}_{data,k}$ are the generated and training-batch categorical marginals for group k, $\|\cdot\|_2$ is the Euclidean

norm, Ω_k is the support observed in training, $\tilde{p}_{G,k}^{\Omega}$ is the generated marginal restricted to that support, $H(\cdot)$ is Shannon entropy, and $\log|\Omega_k|$ normalizes entropy by the maximum possible value for group k. The marginal-matching term keeps generated discrete choices aligned with the labeled design population, while the entropy term prevents one category from dominating a group when several categories remain plausible.

The final generator objective is

$$\mathscr{L}_G = \mathscr{L}_{adv} + \lambda_c \mathscr{L}_{constraint} + \lambda_p \mathscr{L}_{perf} + \lambda_m \mathscr{L}_{moment} + \lambda_{div} \mathscr{L}_{div} + \lambda_{cat} \mathscr{L}_{cat}. \tag{7}$$

Here $\mathscr{L}_G$ is the generator loss, $\mathscr{L}_{adv}$ is the adversarial term derived from the critic, $\mathscr{L}_{constraint}$ collects differentiable engineering-constraint penalties, $\mathscr{L}_{moment}$ matches selected first- and second-order moments of the generated and training distributions, $\mathscr{L}_{div}$ promotes sample diversity, and λ_c, λ_p, λ_m, λ_{div}, and λ_{cat} weight the corresponding modules. The objective combines four complementary signals: distributional realism from the critic, local engineering feasibility from the constraint module, multidisciplinary preference from the physical-response model, and design-family coverage from the diversity terms. The resulting objective treats generation as constrained distribution learning rather than sequential sample generation followed by filtering.

3. Results

3.1 Experimental setup

The reported experiments use multidisciplinary response labels generated with an aircraft-analysis workflow. The final cleaned physics-model dataset contains 24590 analysis-labeled rows after metric-domain filtering. Rows with missing responses or physically invalid metric ranges are excluded before splitting, and no response metric is used as an input feature. The grouped split used for the selected ensemble contains 17969 training rows, 3001 validation rows, and 3620 held-out test rows. The encoded design dimension is 323, consisting of 68 continuous dimensions and categorical one-hot groups. The grouped split reduces the likelihood that near-duplicate variants appear in both training and validation/test subsets.

The aircraft-analysis workflow is implemented with RCAIDE, the Research Community Aircraft Interdisciplinary Design Environment. RCAIDE provides an open, object-oriented aircraft analysis framework for conceptual design and flight simulation, with coupled vehicle, mission, aerodynamics, propulsion, energy, and noise analyses suitable for multidisciplinary studies [23]. In this work, RCAIDE is used to generate independent response labels for sampled DEP configurations; these labels supervise the differentiable physical-response model but are not exposed as generator input variables. Each configuration is evaluated with its decoded cruise range, altitude, and speed, and the recorded responses are reduced to mission-averaged aerodynamic, trim, battery, and acoustic quantities. Acoustic labels are obtained from a frequency-domain buildup calculation on a reduced hemispherical microphone grid with 20×28 observer points, six time samples, and a five-point stencil. Representative decoded geometries are visualized with OpenVSP in the geometric inspection stage.

The selected physics model is a three-member ensemble. Its network structure, together with the generator, critic, and differentiable loss coupling, is summarized in Fig. 1. The physical-response model uses a shared residual trunk followed by six target heads. The generator uses a residual sampler with separate continuous and categorical output heads, and the critic operates in the same encoded design space as the generator output. The selected implementation uses a 96-dimensional latent input for the generator, a 1024-wide physical-response trunk with five residual blocks, and 512-wide generator and critic trunks with four residual blocks.

For visual and geometric inspection, a grouped gallery is selected from the generated audit set. The audit set is partitioned into design families using propulsor count, internal/tip propeller diameter, blade-count category, mission range, acoustic level, mission altitude, and configuration mode. Two representative samples from each family are materialized geometrically and accompanied by their decoded parameters and physics-model audit values.

3.2 Generated aircraft families

The family gallery in Figure 2 and Tables 2–3 makes the generated distribution inspectable at the aircraft level. The eight families are post-training analysis groups formed from the generated audit set,

not additional training labels. They separate compact low-blade propulsors, large tip-propeller layouts, dense distributed propulsion, high-blade electric propulsion, long-range high-L/D candidates, low-noise candidates, high-altitude mission variants, and an alternate configuration mode. Two samples are drawn from each family, giving 16 representative designs. In Table 2, N_s is the number of propulsors per side, B_i/B_t and D_i/D_t denote internal/tip blade counts and diameters, and battery topology is reported as module/series/parallel count. In Table 3, hatted quantities are predictions from the physical-response ensemble and the constraint column is the differentiable aggregate constraint loss.



Figure 2 – Representative generated DEP aircraft families.

The parameter spread in Table 2 indicates that the generated set is not limited to small perturbations of one baseline aircraft. The selected family samples cover two configuration modes, propulsor counts from two to five per side, blade-count combinations from $2/2$ to $10/7$ and $9/10$, internal diameters from 0.77 m to 1.13 m, tip diameters from 0.48 m to 1.60 m, and mission ranges from 91.3 nmi to 149.7 nmi. Battery topology also varies: the module count ranges from 6 to 12, the series count from 14 to 24, and the parallel count from 26 to 50.

The compact low-blade family uses two propulsors per side and low blade counts. Its representative samples maintain moderate diameters and compact ranges near 104–110 nmi, with predicted L/D values of 17.19–17.45 and final SOC values of 0.539–0.565. This family can be interpreted as a balanced short-to-medium-range region: it is not the highest-efficiency group in the audit set, but it preserves battery reserve while keeping the propulsion layout relatively simple.

The large-tip family isolates the effect of large tip propellers. The tip diameter reaches approximately 1.60 m, while the internal diameter remains near 0.89–1.07 m. The two samples have different blade-count patterns: one remains at $2/2$, whereas the other uses $10/7$. The $10/7$ case increases predicted L/D from 16.43 to 17.59 but also raises both acoustic indicators. This comparison suggests a local trade-off: larger and more heavily bladed tip propulsion can improve predicted aerodynamic

Table 2 – Representative family parameters.

Family	Sample	Mode	N_s	B_i/B_t	D_i/D_t (m)	Span (m)	Battery M/S/P	Range (nmi)	Alt. (kft)	Speed (mph)
Compact	gdr_07345	0	2	2/2	1.05/0.84	11.30	11/17/26	104.0	11.3	126.6
Compact	gdr_08565	0	2	3/2	0.88/0.92	11.15	6/19/47	109.5	7.1	119.4
Large-tip	gdr_01998	0	2	2/2	1.07/1.60	11.69	12/15/28	102.9	7.5	119.4
Large-tip	gdr_02359	0	2	10/7	0.89/1.59	11.19	8/24/36	117.6	7.5	125.3
Dense	gdr_07007	0	5	7/5	0.77/1.04	11.27	12/15/40	106.3	8.7	121.2
Dense	gdr_06344	0	4	3/3	0.84/1.02	11.39	11/14/44	116.9	7.6	122.7
High-blade	gdr_04370	0	2	9/10	0.92/1.28	11.64	6/22/39	106.4	10.0	112.2
High-blade	gdr_04916	0	3	6/8	1.06/0.48	11.72	10/23/29	117.6	7.7	116.9
Long-range	gdr_06537	0	2	2/2	1.13/1.52	11.54	12/21/34	137.4	6.6	121.5
Long-range	gdr_05934	0	2	3/9	1.00/1.04	11.78	7/23/45	149.7	6.5	117.3
Low-noise	gdr_01834	0	3	3/2	0.94/1.44	11.43	9/15/37	101.0	7.3	115.7
Low-noise	gdr_05741	0	3	3/2	0.97/1.57	11.44	9/15/37	115.0	7.4	119.1
High-alt.	gdr_08955	1	2	2/2	1.04/1.09	11.63	7/20/39	91.3	11.5	116.1
High-alt.	gdr_01804	1	2	2/2	1.04/1.10	11.61	7/20/39	92.9	9.9	121.4
Alt-config	gdr_00455	1	2	3/2	1.02/1.31	11.41	6/17/50	91.5	8.2	121.5
Alt-config	gdr_05107	1	2	2/2	1.04/1.12	11.60	7/20/39	92.7	8.9	125.6

Table 3 – Representative family performance.

Family	Sample	$\widehat{L/D}$	$\widehat{SPL}_{\max}$	$\widehat{\overline{SPL}}$	SOC_f	Constraint
Compact	gdr_07345	17.45	69.0	64.8	0.565	2.05e-03
Compact	gdr_08565	17.19	68.8	65.4	0.539	2.08e-03
Large-tip	gdr_01998	16.43	69.4	63.1	0.504	1.49e-04
Large-tip	gdr_02359	17.59	71.6	68.4	0.529	2.58e-03
Dense	gdr_07007	18.09	69.6	66.7	0.560	2.58e-04
Dense	gdr_06344	17.40	66.9	63.8	0.561	1.13e-03
High-blade	gdr_04370	17.65	72.1	68.6	0.461	1.48e-03
High-blade	gdr_04916	18.35	69.6	66.4	0.512	1.05e-03
Long-range	gdr_06537	18.00	70.5	64.1	0.547	2.87e-04
Long-range	gdr_05934	18.74	74.5	70.5	0.523	9.06e-04
Low-noise	gdr_01834	17.29	65.3	62.3	0.521	7.18e-04
Low-noise	gdr_05741	17.37	64.8	61.9	0.491	1.29e-03
High-alt.	gdr_08955	17.85	69.1	64.1	0.533	1.66e-08
High-alt.	gdr_01804	17.06	69.2	64.0	0.518	4.12e-05
Alt-config	gdr_00455	16.90	66.3	62.6	0.545	3.99e-06
Alt-config	gdr_05107	16.51	69.5	64.3	0.515	1.50e-08

performance, while acoustic penalties become more visible.

The dense distributed family increases the number of propulsors per side to four or five. Compared with the compact family, it uses smaller internal propellers and maintains tip diameters near 1.02–1.04 m. One dense sample reaches $\widehat{L/D} = 18.09$ with low constraint loss, suggesting that plausible distributed layouts remain available rather than being suppressed by the minimum propulsor-count option. The high-blade family further indicates that blade counts across the full 2–10 range are active. However, the high-blade sample with $9/10$ blades has $SOC_f = 0.461$ and relatively high predicted noise, suggesting that high solidity should be interpreted as a feasible but more constrained design direction rather than a generally preferable solution.

The long-range family illustrates the mission-level shift most directly. Its range rises to 137.4–149.7 nmi and the predicted L/D reaches 18.00–18.74, while final SOC remains near 0.523–0.547. This indicates that longer-range samples in the gallery are not obtained simply by depleting the battery reserve. The associated penalty is acoustic: the 149.7 nmi case has the highest listed $\widehat{SPL}_{\max}$ and $\widehat{\overline{SPL}}$, which is consistent with the expected coupling between propulsive loading, mission requirement, and noise.

The low-noise family occupies a different part of the design distribution. Both samples use three propulsors per side, moderate internal diameters, and large tip propellers. They have the lowest acoustic values in the family set, with $\widehat{SPL}_{\max} = 64.8$–65.3 dBA and $\widehat{\overline{SPL}} = 61.9$–62.3 dBA, while re-

taining $\widehat{L/D} \approx 17.3$ and acceptable final SOC. This region is relevant for concept screening because it sacrifices part of the long-range performance while remaining aerodynamically and energetically reasonable.

The high-altitude and alternate-configuration families both activate the second configuration mode. Their mission ranges are shorter, approximately 91–93 nmi, but their constraint losses are nearly zero and final SOC remains within 0.515–0.545. These samples indicate that a secondary design mode is retained in the generated distribution. The lower mean predicted L/D of the alternate-configuration samples suggests that configuration diversity is feasible, but family-specific optimization may be needed before this mode can match the higher-performing mode-0 candidates.

Across the 16 representative samples, the designs meet the model-based feasibility criterion with constraint losses below 2.6×10^{-3}. The final SOC range of 0.461–0.565 avoids both excessive reserve and excessive depletion in this audit, while the predicted L/D range of 16.43–18.74 and acoustic range of 64.8–74.5 dBA show observable performance variation. The generated set should be interpreted as a family of design alternatives rather than as a single optimized point.

3.3 Aggregate audit and physics-model accuracy

Table 4 reports the 10000-sample audit of the selected generator. The generator retains a large physics-model-usable rate, maintains a moderate spread in predicted aerodynamic performance, and keeps the mean constraint loss below 2.0×10^{-3}. The final SOC distribution remains within the intended reserve region on average, while the acoustic quantities stay below the screening thresholds used in the differentiable physics audit. The audit should be read as a generated-population screening result, with the independent physical-response accuracy reported separately in Table 5.

Table 4 – Selected-generator audit summary.

Metric	Selected generator
Physics-model-usable samples	9956/10000
Mean constraint loss	1.974×10^{-3}
Mean $\widehat{L/D}$	17.5608
Std. $\widehat{L/D}$	0.6792
Mean SOC_f	0.5293
Mean $\widehat{SPL}_{\max}$ (dBA)	69.6622
Mean $\widehat{SPL}$ (dBA)	65.6523
Range std. (nmi)	18.2853

The categorical marginal loss is intended to keep discrete design choices active during distribution learning. In the selected 10000-sample audit set, the generator covers both configuration modes, four propulsor-per-side categories, blade counts from 2 to 10 for both internal and tip propellers, 95 internal-propeller airfoils, 90 tip-propeller airfoils, 13 series-count values, 26 parallel-count values, and seven battery-module counts. Continuous geometry also remains broad: internal propeller diameter ranges from 0.7107 m to 1.1992 m, tip propeller diameter from 0.4300 m to 1.6935 m, and projected span from 10.8175 m to 11.7956 m.

Table 5 reports the selected three-member physics ensemble. The validation minimum coefficient of determination (R^2) is 0.9345, and the held-out test minimum R^2 is 0.9174. The aerodynamic, angle-of-attack, temperature, and SOC targets exceed 0.967 on the held-out test split, while the two acoustic targets have lower held-out accuracy than the other responses. The generated-sample acoustic values are interpreted as screening indicators within the same evidence level as the embedded physical-response model.

4. Discussion

The results are consistent with the proposed workflow, while also defining its present boundaries. The method should be interpreted as a distribution-learning framework for generating and screening DEP design candidates, not as a replacement for final independent verification. The discussion focuses on two central questions: which model components are responsible for the observed behavior, and

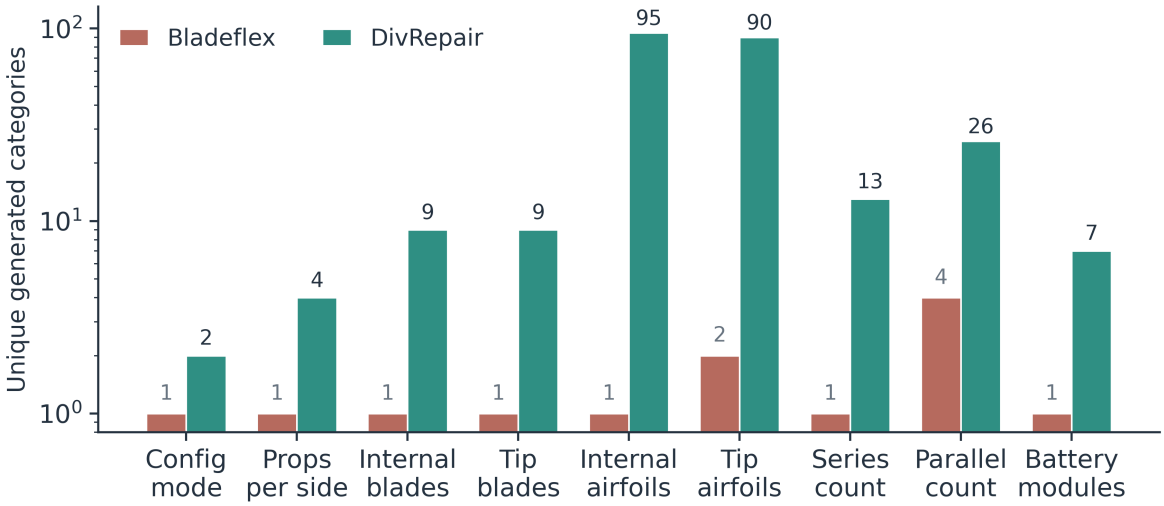


Figure 3 – Categorical coverage.

Table 5 – Physics-model accuracy.

Target	Validation R^2	Held-out test R^2
$\overline{L/D}$	0.9816	0.9771
$\overline{\alpha}$	0.9891	0.9884
ΔT_b	0.9734	0.9679
ΔSOC	0.9883	0.9802
$SPL_{\max}$	0.9367	0.9174
$\overline{SPL}$	0.9345	0.9196

how the performance-diversity trade-off should be interpreted when generated-sample metrics are obtained from an embedded physical-response model.

4.1 Ablation and sensitivity analysis

The ablation study is organized at two levels. The first level compares physical-response model structures, because this model is the differentiable evaluator embedded in the generator loss. The second level compares generator modules, with emphasis on the quality-oriented physics loss, categorical marginal regularization, entropy regularization, moment matching, and diversity restoration. These ablations provide more diagnostic information than a single final-score comparison because the main observed failure mode is not only infeasibility, but also compression of the mixed-variable design distribution.

Table 6 – Physics-model architecture ablation.

Architecture	Principal structural change	Val. min R^2	Test min R^2	Interpretation
Target-wise heads	Independent target heads with no explicit acoustic-delta coupling	0.9199	0.9108	Acoustic targets limit the accuracy floor.
Cascade single	Average acoustic response plus non-negative maximum-noise increment	0.9308	0.9159	Acoustic coupling improves validation stability; seed variance remains visible.
Cascade ensemble	Three cascade-delta members averaged at inference	0.9345	0.9174	Highest validation stability and test floor among the tested structures.

Table 6 reports the weakest target on the validation and held-out test splits for models trained on the same response targets and grouped split. The comparison indicates that the final physics model is not chosen merely by increasing network capacity. The cascade-delta structure embeds a physically meaningful decomposition of acoustic responses: the maximum sound level is represented as a baseline acoustic response plus a non-negative peak increment. This improves the stability of the acoustic heads compared with fully independent target-wise heads. Ensembling further reduces

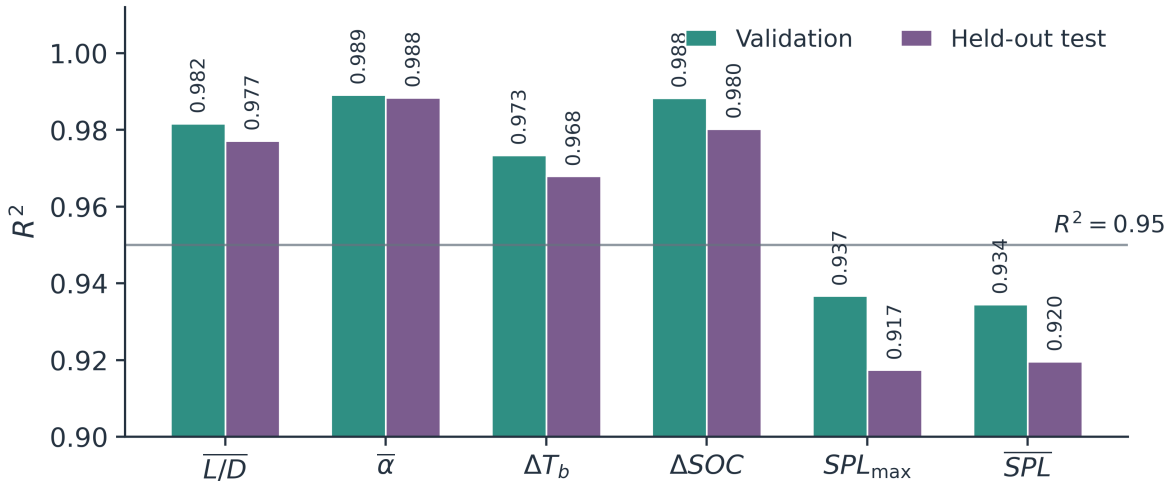


Figure 4 – Physics-model R^2 values.

seed-level variance, which is relevant because the generator receives gradients from the physics model during training.

The generator-module ablation is summarized in Table 7. The first two rows isolate the effect of replacing a feasibility-only physical objective with a quality-oriented physical objective under the same 3000-sample audit protocol. The last two rows isolate the diversity-restoration modules under the final 10000-sample audit protocol. The rows should be interpreted pairwise rather than as a single monotonic ranking.

Table 7 – Generator-module ablation.

Variant	Active or ablated module setting	Audit	Screened	Mean constraint	Main effect
Feasibility-only	Uses response-limit penalties without the utility term $\mathscr{J}$ in Equation (5)	3000	2295	8.72×10^{-4}	Learns response bounds; usable rate and mean L/D are lower.
Quality-oriented	Adds the utility term $\mathscr{J}$ for higher lift-to-drag ratio, lower temperature rise, lower noise, and suitable final SOC	3000	2995	5.24×10^{-4}	Usable rate reaches 99.8%; mean L/D increases by 0.93 and noise decreases by about 1.9 dBA.
Bladeflex	Strong global physical objective; weak $\mathscr{L}_{div}$; no categorical marginal loss	10000	9757	7.59×10^{-3}	Concentrated predicted metrics with major categorical collapse and 6/10 geometric realizations.
DivRepair	Adds $\mathscr{L}_{cat}$, stronger moment/diversity terms, and reduced single-point physical pressure	10000	9956	1.97×10^{-3}	Broad categorical support, 74% lower mean constraint loss, and 10/10 geometric realizations.

The first generator pair illustrates why the physical loss benefits from more than feasibility constraints. A feasibility-only loss can make samples satisfy lower and upper bounds, but once a sample lies inside the feasible region it provides little gradient for ranking better aircraft. Adding the quality-oriented objective turns the physics module into a continuous preference model over the feasible region. In the corresponding audit, the usable count increases from 2295/3000 to 2995/3000, the mean constraint loss decreases, mean L/D increases by 0.9274, and the mean maximum and average acoustic levels decrease by approximately 1.9 dBA.

The second generator pair illustrates why a pure performance objective is not sufficient for generative design. The performance-concentrated model indicates that large physical pressure can concentrate the generator around a favorable predicted basin, but it also collapses major discrete variables such as propulsor count, blade count, airfoil family, battery-module count, and series count. The diversity-restored model reintroduces categorical marginal matching, entropy regularization over observed category support, stronger moment matching, and explicit diversity pressure. These terms change the generator behavior from implicit point optimization toward distribution modeling over feasible air-

craft families.

Table 8 provides the detailed 10000-sample comparison of the last two rows in Table 7. Bladeflex denotes the performance-concentrated generator, whereas DivRepair denotes the selected diversity-restored generator used for the main results. The performance-concentrated variant obtains lower predicted acoustic values and a tight final-SOC distribution, but it collapses to a narrow design family. The diversity-restored variant recovers broad categorical and continuous coverage while preserving a high physics-model-usable rate and reducing mean constraint loss. The observed cost is a shift from single-point high performance to a broader feasible distribution with lower mean L/D and higher mean acoustic levels.

Table 8 – Generator comparison.

Metric	Bladeflex	DivRepair
Physics-model-usable samples	9757	9956
Mean constraint loss	7.590×10^{-3}	1.974×10^{-3}
Mean $\widehat{L/D}$	18.9506	17.5608
Std. $\widehat{L/D}$	0.0524	0.6792
Mean SOC_f	0.4915	0.5293
Mean $\widehat{SPL}_{\max}$ (dBA)	59.9494	69.6622
Mean $\widehat{SPL}$ (dBA)	57.1461	65.6523
Range std. (nmi)	5.3572	18.2853

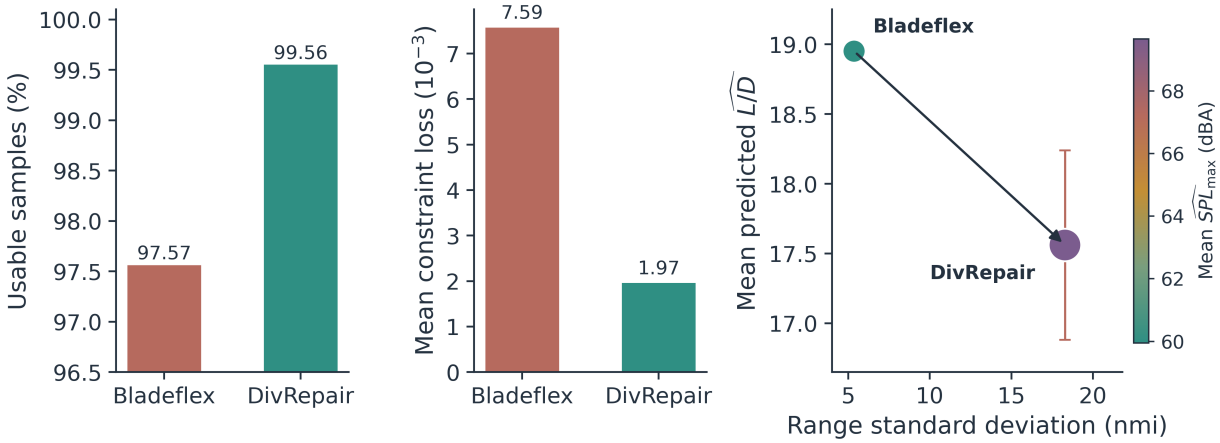


Figure 5 – Performance-diversity trade-off.

In Figure 6, constraint quality is normalized by the lower mean constraint loss of the diversity-restored generator, and category coverage is normalized by the recovered categorical support. The ablation indicates that feasibility and diversity are not equivalent objectives. Increasing the physical-performance weight alone improves concentration around favorable predicted metrics, but it also drives the generator toward a small subset of propeller, battery, and airfoil categories. The diversity-restored objective increases the usable-sample count by 199 samples over 10000 generated designs, reduces mean constraint loss by approximately 74%, and restores blade-count support across the declared 2–10 range. Its lower mean predicted L/D should be interpreted as the cost of recovering a broader design distribution, rather than as a simple degradation of the method.

4.2 Interpretation and limitations

The performance-diversity trade-off is a key consideration in generative aircraft design. A deterministic optimizer can legitimately return one optimum, whereas a generative model is expected to return a useful distribution of alternatives. In this study, the physics objective improves predicted performance only when paired with diversity controls that preserve blade count, airfoil family, battery topology, and propulsor-count support. The categorical marginal and entropy terms serve a methodological role

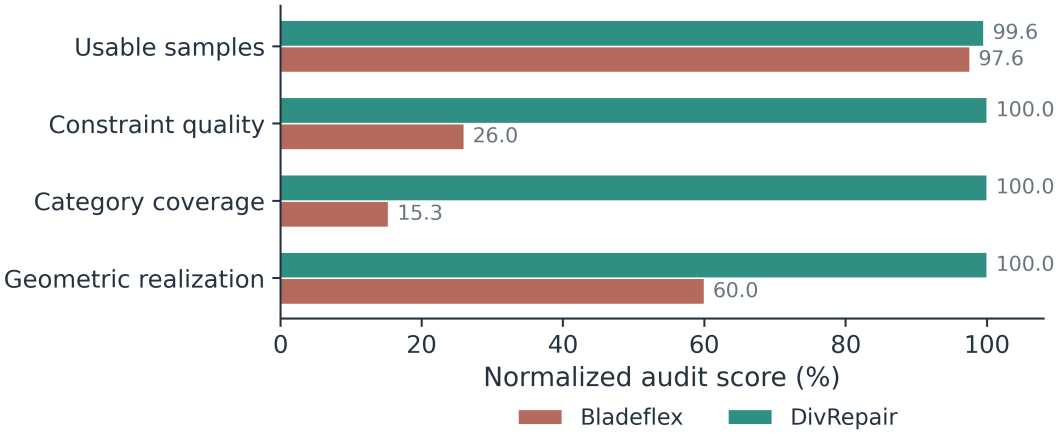


Figure 6 – Objective-level ablation summary.

beyond visual diversity: they reduce the tendency of the learned distribution to become an implicit point optimizer.

Geometric realization success is necessary but not sufficient for aerodynamic or acoustic validity. The geometry files document materialization of generated designs; independent multidisciplinary evaluation is still required for final performance labels. The protocol separates three levels of evidence: differentiable model audit, geometric realization, and independent response evaluation. This separation keeps the interpretation conservative: a generated design is not considered verified merely because it is feasible under the neural physics model or because it can be materialized geometrically.

5. Conclusion

This study leads to three main conclusions. First, a mixed-variable design representation can express DEP aircraft candidates across aircraft geometry, propulsion layout, battery topology, and mission variables while remaining compatible with differentiable generative training. The selected physical-response ensemble provides useful guidance for this task, with higher held-out accuracy for aerodynamic, trim, battery-temperature, and SOC targets and lower but still informative accuracy for the acoustic targets.

Second, physical feasibility and design diversity are not interchangeable objectives. Large physical-performance pressure can produce favorable predicted metrics, but it may also compress the mixed-variable design distribution into a narrow family. Restoring categorical marginal matching, entropy, moment matching, and explicit diversity pressure increases the usable count to 9956/10000, reduces the mean constraint loss to 1.97×10^{-3}, and recovers broad coverage of propulsor count, blade count, airfoil family, battery topology, and mission variables.

Third, the representative family analysis indicates that the generated aircraft are not small perturbations of one baseline. The generated set contains compact low-blade layouts, large tip-propeller layouts, dense distributed-propulsion layouts, high-blade variants, long-range candidates, low-noise candidates, high-altitude mission variants, and an alternate configuration mode. The associated trade-offs are consistent with expected physical couplings: longer range and higher predicted L/D tend to increase acoustic burden, whereas low-noise candidates sacrifice part of the aerodynamic optimum while preserving feasible energy reserve.

The present evidence supports the use of physics-constrained generative modeling as an early-stage DEP concept-generation tool. Final design decisions should still rely on independent multidisciplinary verification before quantitative performance claims are made. Future work should extend independent response evaluation in generated design families and develop family-aware objectives that improve performance within each family rather than driving all generated samples toward a single predicted optimum.

Funding

This work was supported by the National Key Research and Development Program of China (No. 2023YFB3002800), the National Natural Science Foundation of China (No. 52502451), the Natural Science Basic Research Plan in the Shaanxi Province of China (No. 2025CY-YBXM-083), and the Fundamental Research Funds for the Central Universities (No. G2024KY05104).

Contact Author Email Address

All correspondence regarding this paper should be addressed to: y.ma@nwpu.edu.cn.

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] Nicholas K Borer, Michael D Patterson, Jeffrey K Viken, Mark D Moore, JoeBen Bevirt, Alex M Stoll, and Andrew R Gibson. Design and performance of the nasa sceptor distributed electric propulsion flight demonstrator. In *16th AIAA aviation technology, integration, and operations conference*, page 3920, 2016.

[2] Michael D Patterson, Joseph M Derlaga, and Nicholas K Borer. High-lift propeller system configuration selection for nasa's sceptor distributed electric propulsion flight demonstrator. In *16th AIAA aviation technology, integration, and operations conference*, page 3922, 2016.

[3] Benjamin J Brelje and Joaquim RRA Martins. Electric, hybrid, and turboelectric fixed-wing aircraft: A review of concepts, models, and design approaches. *Progress in Aerospace Sciences*, 104:1–19, 2019.

[4] Smruti Sahoo, Xin Zhao, and Konstantinos Kyprianidis. A review of concepts, benefits, and challenges for future electrical propulsion-based aircraft. *Aerospace*, 7(4):44, 2020.

[5] Andreas W Schäfer, Steven RH Barrett, Khan Doyme, Lynnette M Dray, Albert R Gnadt, Rod Self, Aidan OSullivan, Athanasios P Synodinos, and Antonio J Torija. Technological, economic and environmental prospects of all-electric aircraft. *Nature Energy*, 4(2):160–166, 2019.

[6] Peter F Pelz, Philipp Leise, and Marvin Meck. Sustainable aircraft designa review on optimization methods for electric propulsion with derived optimal number of propulsors. *Progress in Aerospace Sciences*, 123:100714, 2021.

[7] Pedro D Bravo-Mosquera, Fernando M Catalano, and David W Zingg. Unconventional aircraft for civil aviation: A review of concepts and design methodologies. *Progress in Aerospace Sciences*, 131:100813, 2022.

[8] Thierry Lefebvre, Nathalie Bartoli, Sylvain Dubreuil, Marco Panzeri, Riccardo Lombardi, Pierluigi Della Vecchia, Luca Stingo, Fabrizio Nicolosi, Agostino De Marco, Pier Davide Ciampa, et al. Enhancing optimization capabilities using the agile collaborative mdo framework with application to wing and nacelle design. *Progress in Aerospace Sciences*, 119:100649, 2020.

[9] Jichao Li, Xiaosong Du, and Joaquim RRA Martins. Machine learning in aerodynamic shape optimization. *Progress in Aerospace Sciences*, 134:100849, 2022.

[10] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[11] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. Pmlr, 2017.

[12] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[13] Hairun Xie, Jing Wang, and Miao Zhang. Parametric generative schemes with geometric constraints for encoding and synthesizing airfoils. *Engineering Applications of Artificial Intelligence*, 128:107505, 2024.

[14] WANG Jing, LI Runze, HE Cheng, CHEN Haixin, ZHAI Chen, et al. An inverse design method for supercritical airfoil based on conditional generative models. *Chinese Journal of Aeronautics*, 35(3):62–74, 2022.

[15] Shi-Yi Jin, Shu-Sheng Chen, Shi-Qi Che, Jin-Ping Li, Jia-Hao Lin, and Zheng-Hong Gao. Airfoil aerodynamic/stealth design based on conditional generative adversarial networks. *Physics of Fluids*, 36(7), 2024.

[16] Kazunari Wada, Katsuyuki Suzuki, and Kazuo Yonekura. Physics-guided training of gan to improve accuracy in airfoil design synthesis. *Computer Methods in Applied Mechanics and Engineering*, 421:116746, 2024.

[17] Dule Shu, James Cunningham, Gary Stump, Simon W Miller, Michael A Yukish, Timothy W Simpson, and Conrad S Tucker. 3d design using generative adversarial networks and physics-based validation. *Journal of Mechanical Design*, 142(7):071701, 2020.

[18] Amin Heyrani Nobari, Wei Chen, and Faez Ahmed. Range-constrained generative adversarial network: Design synthesis under constraints using conditional generative adversarial networks. *Journal of Mechanical Design*, 144(2):021708, 2022.

[19] Sam Sisk and Xiaosong Du. Physics-constrained generative artificial intelligence for rapid takeoff trajectory design. In *AIAA AVIATION FORUM AND ASCEND 2025*, page 3802, 2025.

[20] Brenda M Kulfan. Universal parametric geometry representation method. *Journal of aircraft*, 45(1):142–158, 2008.

[21] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

[22] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.

[23] Matthew A Clarke, Matteo Guidotti, Sai Shekar, Aidan Molloy, Cade Boggan, and Yair N Guerrero. Rcaide: A multidisciplinary analysis toolbox for aircraft design and flight simulation. *Aerospace Science and Technology*, page 112328, 2026.