

Article

Uncertainty-Aware Lightweight Design of CFRP Battery Enclosure Under Extreme Cold Side-Pole Impact via Bayesian Surrogates

Desheng Zhang ^{1,2}, Jieguo Liao ¹, Longbin Wang ², Zhenxin Sun ^{2,*}  and Han Zhang ²

¹ College of Automobile and Traffic Engineering, Heilongjiang Institute of Technology, Harbin 150050, China; zhangdesheng@hljit.edu.cn (D.Z.); 11819838372@163.com (J.L.)

² College of Mechanical and Electronic Engineering, East University of Heilongjiang, Harbin 150066, China; wanglongbin456@163.com (L.W.); zhanghan57@hljeu.edu.cn (H.Z.)

* Correspondence: samuel@hljeu.edu.cn

Abstract

Mass M (kg) and peak intrusion L (mm) are jointly minimized for a CFRP-enabled battery pack enclosure under the GB 38031-2025 -40° side-pole extrusion condition. A 50-run explicit FE design of experiments is conducted and deterministically partitioned into 37/5/5/3 for initial training, two sequential enrichment batches, and an independent hold-out test. Bayesian additive regression trees are trained as the primary surrogates for M , L , and Stress, and stress acceptability is enforced through a probability-of-feasibility (PoF) gate anchored to a baseline-scaled cap, $\sigma_{\text{lim}} = 1.2 \sigma_{\text{base}} = 410.4$ MPa. NSGA-II performed on the feasible surrogate landscape yields a bimodal feasible non-dominated set. The two branches correspond to two discrete levels of a key thickness variable x_4 : a low-mass regime ($n = 106$) with $M = 100.61$ – 104.81 kg and $L = 5.430$ – 5.516 mm at $x_4 \approx 5.60$ mm, and a stiffer regime ($n = 94$) with $M = 110.69$ – 115.08 kg and $L = 5.362$ – 5.430 mm at $x_4 \approx 8.00$ mm. PoF screening eliminates part of the intermediate region where feasibility confidence is insufficient. Independent FE reruns further indicate that the PoF gate reduces deterministic misclassification near the stress boundary (e.g., one near-threshold candidate exceeds σ_{lim} , whereas others satisfy the cap with margin). Overall, the proposed workflow offers a traceable lightweighting route under extreme-cold uncertainty within a constrained FE budget.

Keywords: battery pack enclosure; CFRP; extreme cold; side-pole extrusion; Bayesian surrogate; BART; Gaussian process regression; probabilistic feasibility; multiobjective optimization



Academic Editor: Pascal Venet

Received: 26 January 2026

Revised: 5 February 2026

Accepted: 12 February 2026

Published: 13 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Operating battery electric vehicles in cold regions imposes stricter requirements on the structural safety and reliability of the battery pack enclosure. Low temperatures can substantially alter the mechanical response and failure mechanisms of polymer-matrix constituents and laminated composites. In particular, reduced matrix ductility and temperature-dependent damage evolution may promote crack initiation and interlaminar/interfacial growth, thereby changing energy-absorption pathways and degrading intrusion resistance under pole-type loading [1,2].

At the regulatory level, tightening safety requirements further amplify the need for early-stage, structure-level verification. China's mandatory national standard GB

38031—2025 has been released with a defined implementation schedule [3], and its official listing specifies the release date and the enforcement date (1 July 2026) [3], reinforcing the expectation that traction-battery systems satisfy stringent safety criteria under prescribed conditions [4].

Among side-pole crash and intrusion scenarios, stress acceptability is enforced via a probabilistic feasibility gate calibrated to a baseline-scaled stress threshold, which governs both the adaptive enrichment and candidate filtering processes. Excessive structural intrusion constitutes the primary precipitating factor for mechanical abuse-induced thermal runaway in traction batteries. Consequently, confining deformation within the available packaging clearance is paramount to preclude module compression and subsequent internal short circuits (ISCs). Representative protocols prescribe a rigid, narrow pole and tightly defined impact kinematics; for example, the Euro NCAP oblique pole protocol specifies a target speed of 32 km/h and a 75° impact angle, while the U.S. compliance procedure for FMVSS 214 rigid-pole testing standardizes key requirements for the vehicle-to-pole configuration. Such narrow-pole loading can trigger rapid local buckling and high curvature deformation in sills, cross-members, and battery-side walls, motivating battery-protection concepts and dedicated energy-absorbing modules tailored to lateral pole loading [4–8].

Lightweighting strategies often consider CFRP-based solutions because of their high specific stiffness and strength; however, cold-region applications face a practical obstacle: low-temperature material data are “available but not self-consistent.” Beyond the commonly cited low-temperature static/impact studies, additional work has quantified temperature effects on tensile behavior and impact damage characteristics and has reported residual strength after impact at low temperatures, collectively indicating that measured trends depend strongly on the specific fiber/resin system, layup, cure route, and test metric [9–11]. Consequently, building a fully consistent, full-temperature coupon database is often cost-prohibitive, whereas assembling parameters across heterogeneous sources can introduce system-mismatch uncertainty that undermines simulation fidelity and design conclusions [12,13]. Given the prohibitive logistics of comprehensively characterizing every layup candidate at -40°C , deterministic optimization is prone to yielding non-robust designs. Consequently, a paradigm shift is essential: rather than presuming the availability of exhaustive empirical data, the design framework must explicitly internalize the predictive uncertainty inherent in sparse or heterogeneous material inputs.

Computational cost forms a second bottleneck. Side pole crash simulations are commonly performed using explicit dynamics, where the stable time increment is constrained by the smallest characteristic element length and the dilatational wave speed; mesh refinement in contact and local buckling regions can therefore reduce the critical time step and inflate runtime. When the design space includes multiple geometric and laminate variables and the assessment must be repeated under multiple temperature scenarios, direct high-fidelity simulation-driven multi-objective exploration becomes prohibitively expensive. Recent work has therefore pursued high-throughput virtual crash datasets and machine-learning surrogates to accelerate crashworthiness screening under side-pole-type loading, while Bayesian optimization and sequential sampling provide a principled route to reduce the number of expensive evaluations under uncertainty [14–17].

Recent work has addressed lateral crash protection for EV battery enclosures, including pole-impact-oriented energy absorbers and enclosure-level crashworthiness concepts under pole-type loading [5]. In parallel, surrogate and machine-learning models have been explored to accelerate crashworthiness assessment when high-fidelity explicit simulations are expensive [14,18–20]. System-level battery-pack models implemented in MATLAB/Simulink have also been reported for electrical-performance studies. These models complement structure-level crashworthiness simulations [21]. Meanwhile, com-

posite mechanics studies have consistently reported that CFRP mechanical response and damage evolution can change markedly with temperature, including low-temperature static behavior and impact-induced damage characteristics.

However, these lines of research are rarely combined into a reusable workflow for cold-region enclosure design under side-pole extrusion. First, enclosure-level studies commonly assume generic or room-temperature conditions, while low-temperature effects are often examined at coupon/laminate level without being propagated to enclosure-level extrusion response and design decisions. Second, low-temperature composite inputs available in the literature are fragmented across material systems, layups, and test definitions; directly assembling parameters from multiple sources can introduce hidden inconsistencies, yet such system-mismatch uncertainty is seldom represented in a traceable manner. Third, explicit pole-impact analyses remain computationally demanding due to stability-limited time increments and local mesh refinement near contact/buckling regions, which makes brute-force multi-objective exploration impractical under realistic budgets. Finally, credibility loops are often limited to internal surrogate validation; independent high-fidelity reruns at selected designs (and scaled verification when feasible) are important to support safety-relevant conclusions [14].

To address these limitations, the present study adopts $-40\text{ }^{\circ}\text{C}$ as the primary scenario and establishes a traceable, uncertainty-aware, and sample-efficient surrogate-assisted workflow for side-pole extrusion design of battery pack enclosures [15].

In summary, this workflow integrates low-temperature composite uncertainty, feasibility-aware optimization, and independent FE reruns to enable reliable lightweight design under a tight simulation budget.

2. Materials and Methods

2.1. FE Model and Side-Pole Extrusion Setup

Lateral extrusion evaluates the crush resistance of the battery pack enclosure. It also reduces the risk of cell/module squeezing during collapse. The simulation follows GB 38031-2025, §8.2.4. After extrusion, the pack/system must show no fire or explosion. The post-test insulation resistance must meet the standard requirements. Figure 1 shows the assembly-level model. Two rigid walls constrain the assembly. A rigid cylindrical pole loads the long-side surface of the enclosure.

A rigid cylindrical pole (150 mm in diameter) moves in the $-Z$ direction under displacement control at 2 mm/s. Loading stops when the reaction force reaches 100 kN. It also stops if deformation reaches 30% of the overall dimension in the loading direction. The earlier condition is used. The standard specifies a 10 min hold at the final displacement. Here, responses are extracted only up to the termination point [6,7].

The enclosure is discretized mainly with shell elements. The nominal element size is 10 mm in the production runs. This size resolves the pole-beam contact region. Numerical stability is checked using commonly energy-balance criteria. Beam-to-case connections use nodal coupling. Skin-core bonding uses TIE constraints. Pole-enclosure interaction uses surface-to-surface contact. A soft-contact formulation is applied ($\text{SOFT} = 1$). Penetration controls are kept consistent.

Key numerical settings are summarized for reproducibility. Shell elements are used. The global $-Z$ axis is the extrusion direction. Shell formulation, through-thickness integration, and hourglass control follow LS-DYNA R11.0 explicit settings. These options are fixed in all cases. The loading rate is 2 mm/s to limit inertial effects. Contact uses a penalty-based surface-to-surface formulation with $\text{SOFT} = 1$. Friction and penetration controls are identical in all runs. Rigid walls suppress rigid-body motion. Mounting constraints follow the baseline installation definition.

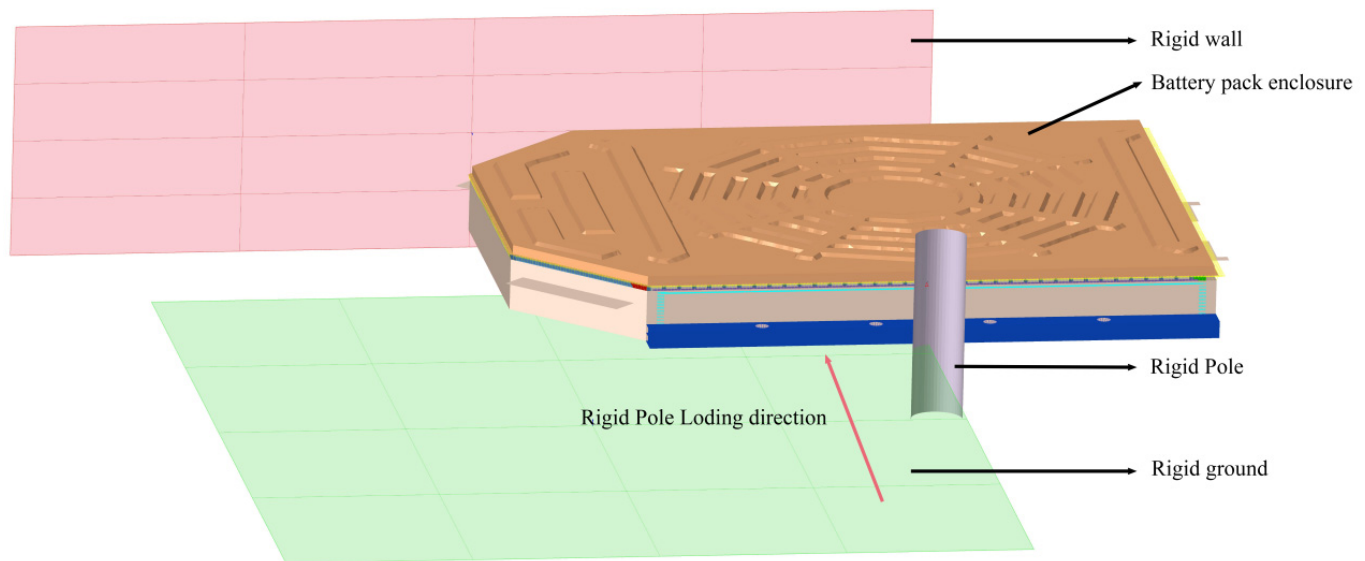


Figure 1. Assembly-level FE model of the battery pack enclosure for the lateral extrusion scenario. Colors are used only to distinguish different components for clarity and do not represent any physical quantity.

Besides the standard pass/fail safety criteria, an application-oriented protection constraint is imposed based on the packaging layout: the maximum intrusion L in the loading direction shall be smaller than the 20 mm packaging clearance between the enclosure and the internal modules. Accordingly, the total enclosure mass M , the maximum intrusion L , and the peak equivalent stress in the enclosure are reported as the key performance metrics for subsequent evaluation and optimization.

2.2. Design Variables, Objectives, and Constraints

2.2.1. Variables and Bounds

Let $\mathbf{x}$ denote the design vector. To keep the extreme-cold campaign computationally affordable, the parameterization is limited to a small set of factors that directly control the main load paths, while secondary details are held constant at baseline settings.

As shown in Figure 2, the design variables are described as follows:

$$\mathbf{x} = [x_1, \dots, x_9]^T \quad (1)$$

The first five variables x_1 – x_5 (mm) are thicknesses assigned to the end bulkhead, cooling plate, load-distribution support plate, lower case, and lateral anti-collision beam, respectively.

The remaining four variables x_6 , x_7 , x_8 , and x_9 describe the CFRP layup using an orientation-grouped weight parameterization for 0° , $+45^\circ$, -45° , and 90° plies. Ply angles are defined with respect to the local e_1 axis in Figure 1. To obtain physically meaningful layup ratios with a unit sum, the weights are normalized prior to laminate construction as follows:

$$r_i = \frac{x_i}{x_6 + x_7 + x_8 + x_9}, \quad i \in \{6, 7, 8, 9\} \quad (2)$$

where $\mathbf{r} (r_6, r_7, r_8, r_9) = (r_0, r_{45}, r_{-45}, r_{90})$ and $\sum r_i = 1$. In this work, x_6 – x_9 serve as unconstrained layup weights in the optimizer, while $r_0, r_{45}, r_{-45}, r_{90}$ are the normalized, unit-sum ratios used to construct the laminate. For implementation, the ratios are mapped to integer ply counts under a prescribed total ply number N . A simple rounding-and-repair step is applied to satisfy basic manufacturability requirements, such as balanced $\pm 45^\circ$ plies.

The DOE design variables and their bounds/levels are summarized in Table 1. Any quantities held fixed across the 50 runs are treated as baseline settings and therefore excluded from the set of design variables.

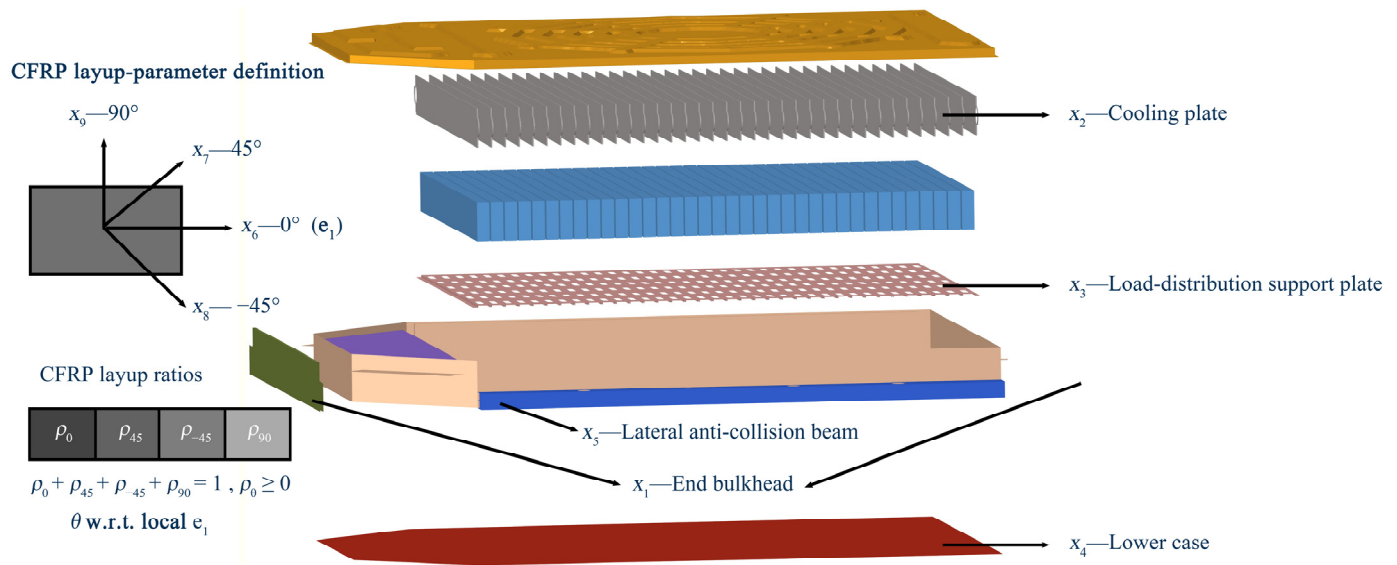


Figure 2. Dominant-factor screening and design-variable definition for the battery pack enclosure under lateral extrusion, including key thickness parameters and the CFRP layup. Colors are used only to distinguish components and have no quantitative physical meaning.

Table 1. Design variables and bounds used for DOE (50-run LHS).

ID	Variable	Type	Bounds/Levels	Unit/Note
x_1	End bulkhead	Continuous	2.0–4.0	mm (thickness)
x_2	Cooling plate	Continuous	1.0–3.0	mm (thickness)
x_3	Load-distribution support plate	Continuous	1.0–3.0	mm (thickness)
x_4	Lower case	Discrete	4.8–8.0	mm (thickness)
x_5	Lateral anti-collision beam	Discrete	6.0–9.0	mm (thickness)
x_6	0°	Discrete	1.0–2.0 (levels: 1.00, 1.25, 1.50, 1.75, 2.00)	layup weight (dimensionless) *
x_7	45°	Discrete	same as x_6	same as x_6
x_8	−45°	Discrete	same as x_6	same as x_6
x_9	90°	Discrete	same as x_6	same as x_6

Note: * Layup weights are dimensionless normalized ratios (e.g., $r_0, r_{45}, r_{-45}, r_{90}$) with $\sum r_k = 1$.

The four layup weights (x_6, x_7, x_8, x_9) are normalized to layup ratios ($r_0, r_{45}, r_{-45}, r_{90}$) by

$$r_i = \frac{x_i}{\sum_{j=6}^9 x_j}, \quad \sum r_i = 1 \quad (3)$$

The layup ratios ($r_0, r_{45}, r_{-45}, r_{90}$) are converted to integer ply counts using a rounding-and-repair procedure [22]. The procedure enforces basic admissibility. It ensures nonnegativity. It enforces $\sum r_i = 1$ It also keeps the $\pm 45^\circ$ plies balanced. This mapping serves as a practical manufacturability screen under the limited FE budget. It does not include process effects. These include cure-induced residual stress and warpage. It also excludes ply-drop and taper rules for thickness transitions. Plant-specific stacking-sequence rules are not enforced either. Examples include symmetry/balance requirements, limits on consecutive plies, and corner drapeability. These constraints can be added later as discrete

rules in the same repair step. They can also be checked during confirmatory FE reruns for shortlisted candidates. The laminates reported here are design-admissible candidates, not manufacturing-release configurations. Within these bounds, a 50-run Latin hypercube sampling (LHS) design was generated and evaluated using the FE model, producing three responses: mass M , peak displacement L , and peak stress ($Stress$). The 50 runs are treated as one DOE dataset and partitioned into 37/5/5/3 for initial training, two sequential enrichment batches, and final independent validation. The complete 50-run dataset will be provided in the Appendix A/Supplementary Materials.

2.2.2. Objective Definition M, L

The cold-driven optimization is posed as a multi-objective design problem under the $T = -40\text{ }^{\circ}\text{C}$ extrusion scenario, described as follows:

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) = [M(\mathbf{x}), L_{-40}(\mathbf{x})] \quad (4)$$

where $\mathbf{x}$ denotes the design vector, $M(x)$ is the total mass of the battery pack enclosure, and $L(x)$ is the maximum deformation/intrusion in the loading direction extracted from the $-40\text{ }^{\circ}\text{C}$ lateral rigid-pole extrusion simulation. In this work, $L(x)$ is defined as the peak displacement component along the prescribed loading axis for the specified enclosure region. The exact definition and extraction procedure of L are provided in Section 2.4.

Energy-related responses are not used in the present dataset; therefore, energy-absorption metrics are not included as optimization objectives. Instead, the stress response is evaluated to avoid designs that achieve low intrusion at the expense of excessive local stresses.

Beyond the variable bounds and laminate admissibility rules, an additional stress-based screen is applied to remove designs that achieve lower intrusion at the cost of excessive local stress concentration in the lateral anti-collision beam. Because the beam and CFRP components are simulated without progressive damage, the reported stress is used as an engineering indicator rather than a direct failure prediction. The screening condition is therefore defined as

$$\text{Stress}(x) \leq \sigma_{\text{lim}} \quad (5)$$

Here, σ_{lim} is an engineering stress cap used for pre-screening. In the surrogate-assisted optimization and candidate selection stage, this requirement is enforced in probabilistic form, $\text{PoFStress}(x) = \Pr(\text{Stress}(x) \leq \sigma_{\text{lim}})$, so that predictive uncertainty is explicitly incorporated (see Section 2.2.3).

Force-related outputs are recorded to verify consistency with the prescribed loading definition and stopping criteria; however, they are not treated as optimization objectives in this work.

2.2.3. Probabilistic Feasibility and Acceptance Criteria

Under $-40\text{ }^{\circ}\text{C}$, the design problem is posed as a bi-objective minimization:

$$\min = \{M(x), L(x)\} \quad (6)$$

where $M(x)$ is the structural mass and $L(x)$ is the peak displacement from the side-pole rigid-pole extrusion response of the enclosure.

Structural acceptability is controlled by the peak-stress response $\text{Stress}(x)$ relative to an engineering cap σ_{lim} . In deterministic terms, this corresponds to the check $\text{Stress}(x) \leq \sigma_{\text{lim}}$.

Because $\text{Stress}(x)$ is predicted by a surrogate and may carry non-negligible uncertainty, feasibility is enforced via a probability-of-feasibility (PoF) gate:

$$\text{PoF}_{\text{Stress}}(x) \geq \eta, \eta = 0.9 \quad (7)$$

where $\text{PoF}_{\text{Stress}}(x)$ is computed from BART posterior predictive samples using the Monte Carlo estimator in Section 2.6.2. A packaging-clearance constraint of $L \leq 20$ mm is rigorously enforced. This threshold corresponds to the physical gap between the enclosure's inner wall and the battery modules; filtering out designs that violate this limit is imperative to preclude mechanical contact that could compromise the electrochemical integrity of the cells.

The engineering stress cap is specified relative to the nominal baseline as

$$\sigma_{\text{lim}} = 1.20\sigma_{\text{base}} \quad (8)$$

where $\sigma_{\text{base}} = 342.0$ MPa is the peak (through-time) stress of the baseline design evaluated under the same -40 °C rigid-pole extrusion setup and termination rule. A 20% margin is adopted as a conservative allowance for screening.

Within NSGA-II, feasibility is prioritized during selection. Individuals with $\text{PoF}_{\text{Stress}}(x) < \eta$ are labeled infeasible and are ranked after feasible solutions in non-dominated sorting; accordingly, the reported Pareto set is the feasible subset under the PoF gate. When predictive uncertainty is small, the PoF gate effectively collapses to the deterministic stress-cap criterion, whereas in uncertain regions it acts conservatively and helps avoid selecting borderline designs that may fail subsequent high-fidelity verification [18–20].

2.3. Response Definitions and Extraction Rules

This section defines the performance metrics used in this study and summarizes how they are obtained from the -40 °C lateral rigid-pole extrusion simulations. All simulations in this study are performed at $T = -40$ °C; accordingly, temperature subscripts are omitted from the response notation for brevity.

The enclosure mass $M(x)$ is evaluated for each design x based on the geometry and material assignment of the battery pack enclosure. The mass reported in this work refers to the enclosure structure only; rigid bodies introduced for boundary representation (rigid walls/ground and the rigid pole) are excluded. In practice, M is taken from the pre-processing mass report mass of the enclosure components.

The deformation metric $L(x)$ characterizes the maximum enclosure intrusion along the prescribed loading direction. Since the rigid pole is driven in the $-Z$ direction under displacement control, $L(x)$ is defined as the maximum nodal displacement component in the $-Z$ direction on the enclosure long-side region subjected to extrusion. The value is extracted over the simulation history up to the termination condition of the simulation. This definition directly supports the packaging-based clearance requirement, where the allowable intrusion is limited to 20 mm.

In plain terms, the reported deformation metric is as follows:

$L(x)$ = maximum displacement component along the $-Z$ loading direction on the extruded long-side region over the simulation history up to the termination condition.

To avoid designs that achieve a small intrusion by inducing excessive local stress, the stress response of the lateral anti-collision beam is monitored. The stress metric is defined as the peak through-time von Mises stress in the lateral anti-collision beam, i.e., the maximum value attained over the simulation history up to the termination condition.

In plain terms:

$Stress(x)$ is the highest von Mises stress reached before the simulation stops. Progressive damage and delamination are not modeled in this campaign. This choice avoids introducing poorly calibrated low-temperature strength and fracture parameters, and it keeps the DOE computationally tractable. Therefore, $Stress(x)$ is used as an engineering screening metric rather than a failure predictor. Under this assumption, intrusion L may be slightly optimistic for near-threshold designs because stiffness degradation is not represented. By contrast, elastic stress hot-spots may be conservative because damage-driven load redistribution is not captured. Section 4.3 discusses the expected changes when damage is enabled.

All response metrics are taken from the simulation history up to the termination point. Specifically, $L(x)$ and $Stress(x)$ are reported as history maxima up to termination, whereas $M(x)$ is obtained from the pre-processing mass report. Unless otherwise specified, M , L , and $Stress$ are reported in kg, mm, and MPa, respectively. The mass M includes only the battery-pack enclosure structure; the rigid pole and rigid fixtures (walls/ground) are excluded, consistent with the DOE dataset and the subsequent verification runs.

2.4. Material Models and Temperature Treatment

All simulations in this study are performed under a single extreme-cold condition, $T = -40\text{ }^{\circ}\text{C}$. Temperature is treated as a discrete scenario rather than a continuous design variable. Its effect is introduced only through the selected material-property set, and no interpolation or extrapolation to other temperatures is carried out. Table 2 lists the $-40\text{ }^{\circ}\text{C}$ material parameters taken from the adopted source, and these values are used throughout this study.

Table 2. CFRP lamina material parameters adopted in this study at $-40\text{ }^{\circ}\text{C}$.

Property	Unit	$-40\text{ }^{\circ}\text{C}$	Source
0° tensile modulus E_1	GPa	158.17	[23]
90° tensile modulus E_2	GPa	9.40	[23]
In-plane shear modulus G_{12}	GPa	12.72	baseline assumption
0° compressive strength X_c	MPa	1488.94	[23]
0° tensile strength X_t	MPa	2521.23	[23]
90° compressive strength Y_c	MPa	279.05	[23]
90° tensile strength Y_t	MPa	72.09	[23]
In-plane shear strength S	MPa	124.97	[24]

Reported low-temperature CFRP properties differ widely due to variations in material system, cure schedule, fiber volume fraction, laminate design, and testing protocols. To prevent untracked inconsistencies, a single-reference approach is adopted: one coupon-level dataset with a clearly specified material system and measured properties at $-40\text{ }^{\circ}\text{C}$ is chosen as the anchor. Additional literature values are only retained after screening for compatibility in both material system and test definitions, and every adopted parameter is recorded with explicit provenance in Table 2. The $-40\text{ }^{\circ}\text{C}$ lamina property set used in the FE DOE is taken from coupon-level measurements reported in [23,24] and is adopted as the anchor dataset. The same reference set is used for all 50 FE runs to maintain internal consistency within the available simulation budget. Table 2 lists the adopted parameters and their sources for traceability.

CFRP components are modeled at the lamina level using an orthotropic constitutive description. In this FE campaign, CFRP is modeled using linear orthotropic elasticity only; progressive damage (e.g., Hashin/Puck-type criteria with stiffness degradation) and delamination are not considered. Accordingly, $Stress(x)$ is treated as an engineering screening metric rather than a direct predictor of material failure (see Section 4.3). This

choice preserves a traceable and affordable explicit-dynamics DOE under a limited FE budget. More importantly, it avoids relying on poorly supported calibration inputs: at $-40\text{ }^{\circ}\text{C}$, intralaminar strengths, fracture energies, and interface/cohesive parameters for the adopted laminate and joints are not sufficiently available. Therefore, the PoF gate is defined using this elastic proxy metric. If damage-enabled metrics (e.g., failure indices, damage variables, or delamination area/energy) are used instead, the same PoF framework remains applicable, but the feasible set and the resulting rankings are expected to change.

In the present study, the lower case and the lateral anti-collision beam are the CFRP components whose properties and layup-related descriptors are explicitly considered as influencing factors in the subsequent analysis/optimization. Other CFRP parts in the enclosure assembly (e.g., the upper cover and additional panels) are also modeled as CFRP, but their layups and thicknesses are kept fixed at the baseline configuration because they are not selected as dominant influencing factors in Figure 2 and are therefore not included in the design vector.

A single $-40\text{ }^{\circ}\text{C}$ orthotropic lamina property set is adopted as an anchor dataset based on coupon measurements from domestic sources [23,24] and is applied consistently across all FE runs to maintain internal consistency under the limited simulation budget. All CFRP inputs are listed in Table 2 for traceability. Cross-source scatter at $-40\text{ }^{\circ}\text{C}$ is represented by bounded ranges that define a plausible envelope around the anchor set and are used only for robustness framing; unless stated otherwise, deterministic FE analyses use the anchor set and the envelope is not embedded in the baseline material card. Probabilistic material perturbation at the FE level was considered, but it would require multiple material realizations per design point to estimate response distributions and would exceed the current FE budget. Uncertainty is therefore handled at the surrogate level via Bayesian predictive distributions and the PoF gate, with confirmatory FE reruns for shortlisted candidates.

Metallic components are modeled with a baseline, temperature-independent material card. This choice is made because the present dataset does not support a full calibration of temperature-dependent metallic plasticity. Available tensile evidence for structural aluminum alloys suggests that decreasing temperature causes only a small increase in elastic modulus (about 5–7%), whereas yield and ultimate strengths typically rise (about 9–13%). Consequently, ignoring explicit temperature dependence in the aluminum alloy card is expected to exert only a secondary effect on stiffness-controlled intrusion predictions and is treated as a deliberate modeling simplification [25].

2.5. DOE, Data Split, and Evaluation Protocol

2.5.1. DOE Dataset and Responses

A 50-run DOE dataset was generated by Latin hypercube sampling (LHS) within the bounds listed in Table 1 and evaluated using the high-fidelity FE model. Each run provides three scalar responses extracted from the FE outputs: mass M , peak displacement L , and peak stress (Stress). The response definitions and post-processing procedures follow Section 2 and are applied consistently to all simulations [22,26].

2.5.2. Data Split and Hold-Out Protocol

The DOE dataset is partitioned into four non-overlapping subsets for the surrogate-assisted sequential workflow. An initial set of 37 runs is used to build the first surrogate models. Two subsequent enrichment rounds contribute 5 runs each, expanding the training set to 42 and 47 runs after Round 1 and Round 2, respectively. The remaining 3 runs are kept as a strict hold-out set and are used only for final independent validation; they are excluded from model fitting, hyperparameter tuning, point selection, and any intermediate analyses.

To ensure reproducibility and to avoid subjective choices in the split, the subset assignment is generated deterministically using a fixed random seed. The resulting indices of the 37/5/5/3 partition are stored and reused throughout the retrospective replay experiments and the cross-model verification reported later. The complete 50-run DOE table (run ID, inputs, and responses) is provided in the Supplementary Material (Table S1). An overview of the subset sizes and their roles is summarized in Table 3.

Table 3. DOE partition for the sequential workflow.

Subset	Size	Purpose	Surrogate Fit/Update	Point Selection
Initial training set	37	Initial training for retrospective replay	Yes	Yes
Enrichment batch (Round 1)	5	Evaluate at $n = 37$; then update to $n = 42$	Yes (after eval)	Yes
Enrichment batch (Round 2)	5	Evaluate at $n = 42$; then update to $n = 47$	Yes (after eval)	No
Hold-out validation set	3	Final validation for model at $n = 47$	No	No

Note: The 3 hold-out runs are reserved for final independent validation and are not used for training, tuning, or point selection.

2.5.3. Retrospective Replay/Prequential Evaluation

A retrospective replay is used to quantify how surrogate accuracy changes as additional FE runs are incorporated. The evaluation is performed strictly under the fixed 37/5/5/3 split. At each stage, the surrogate is trained using only the runs that are currently available. It is then evaluated on the next batch that has not been used for fitting. After the error is recorded, that batch is added to the training set for the next stage. This test-then-update procedure mirrors a sequential campaign while avoiding information leakage across stages.

In Round 0, the surrogate is trained on the initial 37 runs and tested on the first enrichment batch of 5 runs. The training set is then expanded to 42 runs. In Round 1, the surrogate trained on 42 runs is tested on the second enrichment batch of 5 runs and the training set is expanded to 47 runs. Finally, the surrogate trained on 47 runs is evaluated on the hold-out set of 3 runs as an independent check. The hold-out subset remains untouched throughout the replay and is not used for fitting, tuning, or any intermediate decision.

Prediction errors are computed on each test batch using commonly used regression metrics. For a test set of size n_t , let y_i be the true value and, the predicted mean. The root mean squared error is described as follows:

$$RMSE = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} (y_i - \hat{y}_i)^2} \quad (9)$$

To enable comparison across responses with different scales, the normalized RMSE is defined as

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \quad (10)$$

where $y_{\max}$ and $y_{\min}$ are taken from the corresponding test batch. The coefficient of determination is described as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n_t} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_t} (y_i - \bar{y})^2} \quad (11)$$

where $\bar{y}$ is the mean of the true values in the same test batch. These metrics are reported for each response, namely M , L , and $Stress$, at every replay stage.

Figure 3 plots NRMSE against the cumulative training size, which grows from 37 to 42 and then 47 runs. Because each point is tested on a different upcoming batch, the curve reflects a prequential replay pattern rather than a conventional learning curve based on a fixed test set, so a strictly monotonic trend is not expected. In the results, L benefits noticeably from the added training data. The error for M stays low, with small fluctuations that likely arise from varying batch difficulty and the limited number of test samples. By contrast, $Stress$ maintains a higher error level across stages, suggesting that peak stress is harder to model with the same amount of data. This is in line with the highly localized and contact-sensitive character of stress responses.

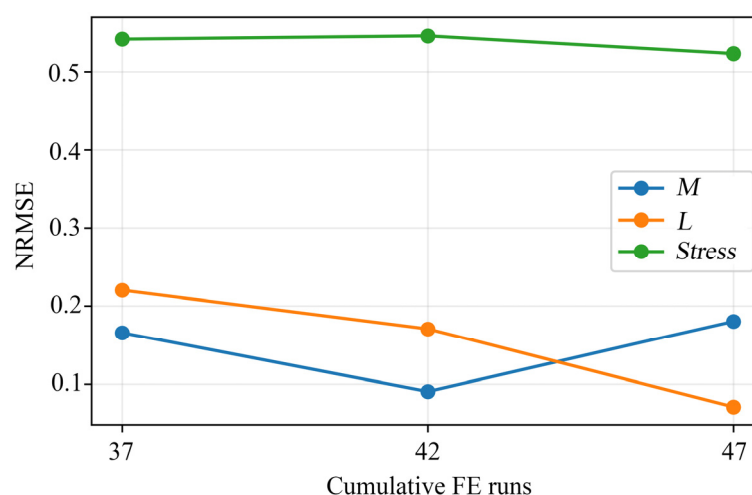


Figure 3. NRMSE across replay stages.

In this work, the sequential workflow is bounded by a fixed FE budget represented by the completed 50-run DOE dataset, and cost is expressed in FE-run units. The model is initialized with 37 runs, and enrichment is capped at 10 additional runs added in two batches of 5. The remaining 3 runs are reserved as a hold-out subset for final validation and are excluded from fitting, tuning, and any intermediate choices. The default stopping rule is to terminate once the enrichment budget is used. Early stopping may also be applied if the prequential test error shows no meaningful change over successive rounds.

2.6. Surrogate Modeling and Feasibility-Guided Enrichment

2.6.1. BART Surrogate (Primary)

Let $\mathbf{x} \in \mathbb{R}^d$ denote the design vector formed by the DOE variables in Table 1 (fixed quantities are not included). Separate surrogates are trained for the three scalar responses M , L , and $Stress$.

BART is used as the primary surrogate because it can capture strong nonlinearity and interaction effects with limited samples, while producing a posterior predictive distribution that supports uncertainty-aware analysis [27].

For each response $y \in \{M, L, Stress\}$, the regression function is represented as follows as a sum of m shallow trees:

$$y(x) = \sum_{j=1}^m g(x; T_j, \Theta_j) + \varepsilon \quad (12)$$

where T_j and Θ_j are the structure and terminal-node parameters of the j -th tree, and ε denotes the residual. Posterior inference follows the Bayesian backfitting procedure, described as follows, and yields posterior predictive draws:

$$\left\{y^{(t)}(x)\right\}_{t=1}^T \quad (13)$$

$$U(x) = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (y^t(x) - \hat{y}(x))^2} \quad (14)$$

The predictive mean is computed from these draws, and the predictive uncertainty is summarized by either the posterior standard deviation or the width of a chosen credible interval. In what follows, the uncertainty measure is denoted by $U(x)$ and is computed in the same manner for all replay rounds.

2.6.2. Probability of Feasibility (PoF) Estimation

Because peak stress is both decision-critical and challenging to predict accurately under a limited explicit-FE budget, feasibility is evaluated from the posterior predictive distribution of Stress rather than from a single mean prediction. For a candidate design x , let $\left\{\text{Stress}^S(x)\right\}_{s=1}^S$ denote the posterior predictive draws produced by the BART surrogate, where S is the number of draws (i.e., pred_draws). Given an engineering stress cap σ_{lim} , the feasibility probability is approximated via Monte Carlo counting:

$$\text{PoF}_{\text{Stress}}(x) = \frac{1}{S} \sum_{s=1}^S \mathbb{I}(\text{Stress}^{(s)}(x) \leq \sigma_{\text{lim}}) \quad (15)$$

No Gaussian or normal approximation is imposed. Designs satisfying $\text{PoF}_{\text{Stress}}(x) \geq \eta$, with $\eta = 0.90$, are treated as feasible for enrichment ranking and for subsequent Pareto exploration. The threshold $\eta = 0.90$ is selected as a pragmatic balance between false rejection and false acceptance under a limited explicit-FE budget. A higher η lowers the chance of accepting borderline, stress-critical designs. However, it can also discard promising lightweight candidates when the stress surrogate remains uncertain. In contrast, a lower η increases exploration and coverage, but it admits more near-boundary designs.

Accordingly, $\eta = 0.90$ is used for enrichment ranking and feasible-Pareto exploration. Final recommendations are not accepted based on PoF alone. They are verified by independent high-fidelity FE reruns. For stress-critical design sign-off, a stricter gate (e.g., $\eta = 0.95$) is recommended.

A packaging-clearance constraint, $L \leq 20$ mm, is applied as an additional post-screening rule. Under the adopted 100 kN termination rule, most DOE and candidate designs satisfy $L(x) \leq 20$ mm with ample margin; therefore, this clearance screen is seldom active in the present dataset. It is retained as an explicit engineering constraint and may become controlling under larger pole travel, alternative termination criteria, or different packaging layouts.

2.6.3. Acquisition and Batch Enrichment

Candidates are ranked using a feasibility-weighted acquisition score, described as follows, where feasibility is quantified by $\text{PoF}_{\text{Stress}}(x)$ and $U(x)$ denotes predictive uncertainty:

$$A(x) = \text{PoF}_{\text{Stress}}(x)U(x) \quad (16)$$

where $U(x)$ measures predictive uncertainty. When multiple responses are considered, $U(x)$ is constructed as a weighted sum of normalized predictive standard deviations, described as follows:

$$U(x) = \sum_{r \in \{M, L, \text{Stress}\}} \omega_r \bar{\sigma}_r(x) \quad (17)$$

With $\bar{\sigma}_r(x)$ denoting the response-wise standard deviation after normalization by its scale and ω_r denoting user-defined weights.

Batch construction follows the ranking while avoiding redundant points. After selecting a candidate, nearby candidates in the normalized design space are temporarily excluded so that the batch maintains diversity. The batch size matches the sequential plan used in the replay, with 5 runs added per round.

The feasibility-guided enrichment procedure is summarized in Figure 4.

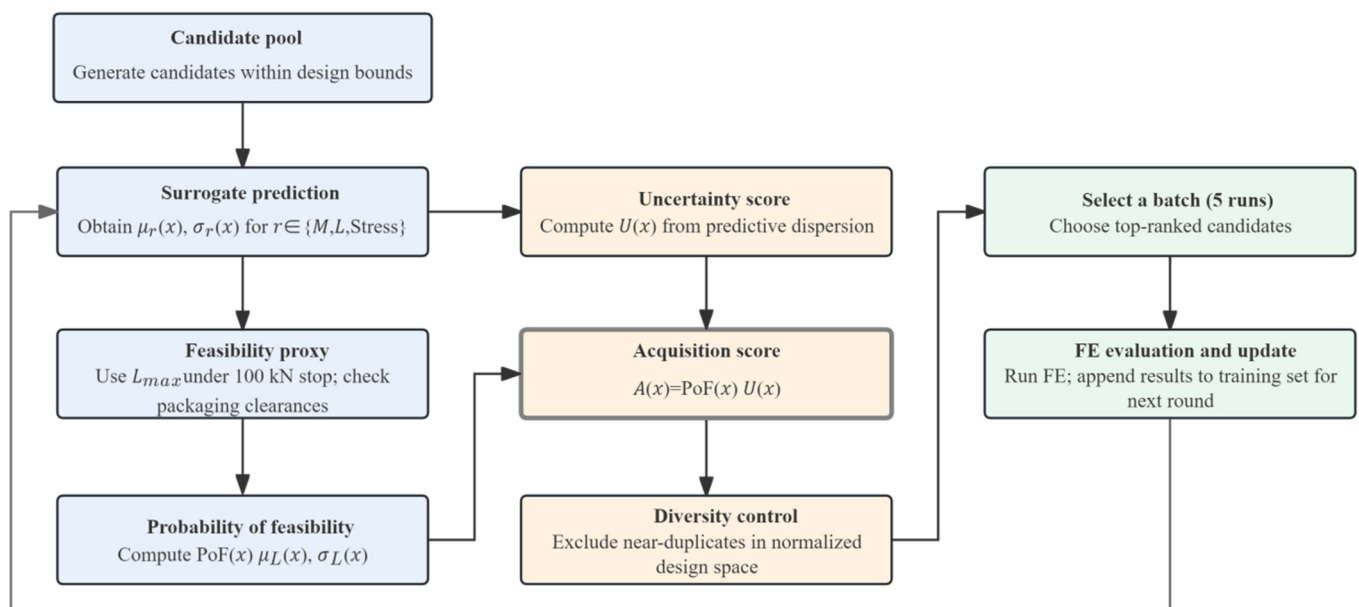


Figure 4. Feasibility-guided enrichment workflow.

Figure 4 illustrates the feasibility-guided enrichment logic adopted in the retrospective replay. Starting from the initial training set, a candidate pool is evaluated using the surrogate models. For each candidate x , BART yields posterior predictive samples for each response, from which the predictive mean and dispersion are estimated. Stress acceptability is measured using the probability of feasibility $\text{PoF}_{\text{Stress}}(x)$, estimated from BART posterior predictive draws as defined in Section 2.6.2. Candidate ranking combines feasibility confidence, such as $\text{PoF}_{\text{Stress}}(x) \geq \eta$ with $\eta = 0.90$, and an uncertainty-driven exploration term.

2.6.4. GPR Reference and Model-Dependence Check

To mitigate model dependence, an additional Gaussian process regression (GPR/Kriging) surrogate is fitted for each response using the same data split and replay protocol. The GPR models act as an independent cross-check to verify that the main conclusions are not unduly driven by the choice of surrogate family [16,17,26,28]. In GPR, the response is modeled as a Gaussian process, described as follows:

$$y(x) \sim \mathcal{GP}(\mu(x), k(x, x')) \quad (18)$$

where $k(\cdot, \cdot)$ is the covariance kernel (with automatic relevance determination when applicable). Hyperparameters are estimated by maximizing the marginal likelihood on the training

subset of the corresponding replay round. The fitted GPR yields closed-form predictions $\mu(x)$ and $\sigma(x)$, which are used as a reference when comparing uncertainty estimates and downstream decisions against those obtained with BART.

Workflow from the 50-run DOE split (37/5/5/3) to surrogate training (BART with GPR check) and downstream use in retrospective replay and feasibility-guided enrichment.

The overall workflow and the roles of the two surrogate models are summarized in Figure 5.

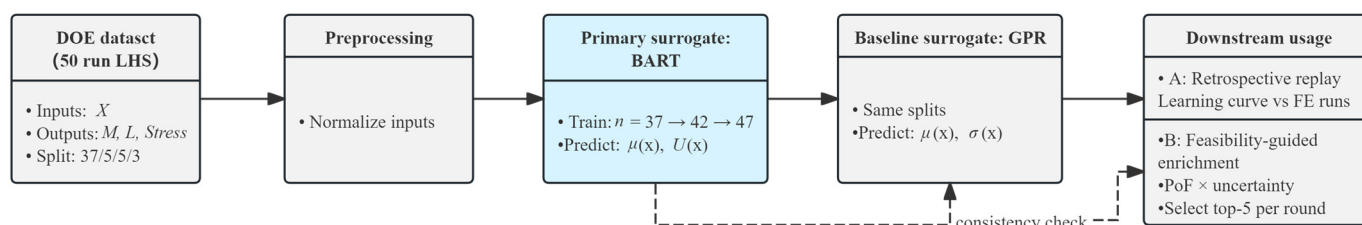


Figure 5. Surrogate-assisted sequential workflow.

To ensure that the conclusions drawn from the sequential study are not tied to a single surrogate type, a dependence check is conducted by repeating the analysis with a Gaussian process regression model. The intention is to use GPR as a transparent reference and to confirm that the observed error trends and the relative difficulty of predicting different responses remain consistent.

For each response, namely M , L , and $Stress$, a separate GPR surrogate is trained using the same input variables as in the BART runs. The 37/5/5/3 partition is kept unchanged. A Matérn kernel with $\nu = 2.5$ is adopted with automatic relevance determination so that different length scales can be learned for different design variables. Hyperparameters are identified by maximizing the log marginal likelihood using only the available training subset at each stage. The fitted model then provides a predictive mean for point prediction and a predictive standard deviation as an uncertainty measure.

The GPR surrogate is evaluated using the identical test-then-update protocol described in Section 2.5.3. In Round 0, training uses the initial 37 runs and testing is performed on the first 5-run batch. In Round 1, the training set is expanded to 42 runs and testing is performed on the second 5-run batch. Finally, the model trained on 47 runs is assessed on the 3-run hold-out subset, which remains excluded from fitting and any intermediate choices throughout.

For each stage and each response, prediction performance is reported using RMSE, NRMSE, and R^2 , matching the metrics used for BART. The side-by-side comparison is summarized in Table 4 and serves as a consistency check for the learning behavior discussed earlier.

The comparison supports the main observation that L is learned more reliably than $Stress$ under the same data regime, while $Stress$ remains challenging due to its contact-sensitive nature.

Figure 6 provides a visual consistency check by plotting the prequential NRMSE trajectories of BART and GPR across the three stages.

Table 4. BART vs. GPR errors (NRMSE, R2).

Response	Stage	n_{train}	Test Subset	n_{test}	BART NRMSE	BART R^2	GPR NRMSE	GPR R^2
<i>M</i>	Round 0 (37 → r1)	37	r1	5	0.165	0.855	0.064	0.978
<i>M</i>	Round 1 (42 → r2)	42	r2	5	0.091	0.924	0.043	0.983
<i>M</i>	Hold-out (47 → holdout)	47	holdout	3	0.180	0.831	0.009	1.000
<i>L</i>	Round 0 (37 → r1)	37	r1	5	0.220	0.597	0.199	0.671
<i>L</i>	Round 1 (42 → r2)	42	r2	5	0.170	0.837	0.058	0.981
<i>L</i>	Hold-out (47 → holdout)	47	holdout	3	0.071	0.972	0.113	0.928
<i>Stress</i>	Round 0 (37 → r1)	37	r1	5	0.541	−0.542	0.738	−1.873
<i>Stress</i>	Round 1 (42 → r2)	42	r2	5	0.545	−1.455	0.320	0.156
<i>Stress</i>	Hold-out (47 → holdout)	47	holdout	3	0.523	−0.572	0.530	−0.616

Note: → denotes the mapping from cumulative sample size to subset label.

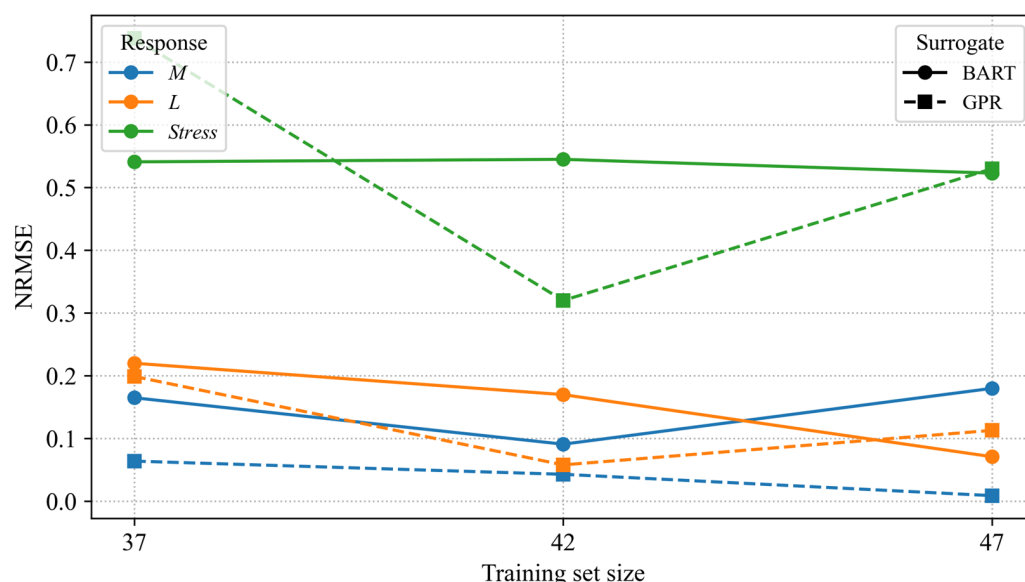
**Figure 6.** NRMSE learning curves of BART and GPR across training set sizes.

Figure 6 indicates that the two surrogate models follow similar learning trends across the replay stages. As more runs are added, the prediction error for *L* generally decreases, whereas *Stress* remains more challenging to capture within the same sampling budget. Small non-monotonic fluctuations are expected, since each stage is tested on a different “next” batch rather than on a single fixed validation set.

Since the enrichment step in Section 2.6.3 relies on a feasibility-weighted uncertainty ranking, a supplementary check can be performed at the ranking level. Using the same candidate pool, acquisition scores produced by BART and by GPR are ranked and compared. Agreement is quantified using Kendall’s τ . This check evaluates robustness of the selection logic without introducing additional FE evaluations.

2.7. Optimization and Decision Pipeline

2.7.1. NSGA-II Configuration and Constraint Handling

Multi-objective optimization is performed exclusively for the $-40\text{ }^{\circ}\text{C}$ side-pole rigid-pole extrusion scenario. Both objective evaluation and feasibility screening are provided by the final surrogate models developed in Section 2.5.3, trained on $n = 47$ high-fidelity samples. Their predictive performance is summarized in Table 4, and the corresponding prequential

behavior is shown in Figure 6; together, these results serve as the error reference for the optimization stage. To emphasize the surrogate-assisted workflow rather than optimizer selection, NSGA-II is used as a standard evolutionary multi-objective baseline to generate a diverse approximation of the Pareto set, in line with typical practice in crashworthiness-driven structural optimization and battery protection design [29–31].

The design vector is $x = [x_1, \dots, x_9]$, where x_1 – x_5 are thickness variables and x_6 – x_9 are discrete layup weights that are normalized into ply ratios. Variable bounds and admissible levels follow the parametric FE model and are listed in Table 1. Two objectives are minimized simultaneously: the enclosure mass $M(x)$ and the peak displacement $L(x)$. Mechanical acceptability is enforced through a stress-based feasibility requirement implemented in probabilistic form.

NSGA-II variation operators (SBX crossover and polynomial mutation) are applied in the continuous decision space. After variation, the discrete thickness variables (x_4 and x_5) are projected to the nearest admissible DOE level. The layup-weight variables (x_6 – x_9) are first normalized into ply ratios and then converted into a manufacturable integer ply schedule using the same rounding-and-repair procedure described in Section 2.2, including enforcement of balanced $\pm 45^\circ$ plies. This projection/repair step is applied to every offspring before surrogate evaluation.

NSGA-II is run with $N_{\text{pop}} = 200$ and $N_{\text{gen}} = 400$, using SBX ($p_c = 0.9$, $\eta_c = 20$) and polynomial mutation ($p_m = 1/9$, $\eta_m = 20$), with a fixed base seed of 2025. All solver settings are kept constant across runs and are summarized in Table 5.

Table 5. Optimization settings.

Item	Setting
Scenario	−40 °C side-pole rigid-pole extrusion only
Design variables	$x = [x_1, \dots, x_9]$ (x_1 – x_5 thickness; x_6 – x_9 layup ratios; see Table 1)
Bounds	See Table 1 (DOE bounds/levels)
Objectives	$\min \{M(x), L(x)\}$
Responses	$M, L, Stress$
Feasibility rule	Feasible if $PoF_{Stress}(x) \geq \eta$
PoF threshold	$\eta = 0.90$
Stress limit	$\sigma_{\text{lim}} = 1.20 \cdot \sigma_{\text{base}}$ ($\sigma_{\text{base}} = 342.0$ MPa from the nominal baseline design; $\sigma_{\text{lim}} = 410.4$ MPa)
Screening criterion (packaging clearance)	Candidates with $L > 20$ mm are removed during post-processing screening
Primary surrogate	BART (final surrogate from Section 2.5.3, trained with $n = 47$)
Cross-check surrogate	GPR (used for surrogate-dependence check)
Accuracy reference	Table 4 (NRMSE, R^2); Figure 6 (prequential NRMSE trajectories)
Optimizer	NSGA-II (baseline optimizer)
Population/generations	$N_{\text{pop}} = 200$; $N_{\text{gen}} = 400$
Operators	SBX: $p_c = 0.9$, $\eta_c = 20$; PM: $p_m = 1/9$, $\eta_m = 20$
Termination	Max generations
Random seed	Base seed = 2025

2.7.2. Screening + TOPSIS

The feasible Pareto set contains many alternatives with nearly indistinguishable trade-offs, so selecting a single “best” design directly is neither transparent nor efficient for verification planning. A two-step procedure was therefore adopted.

First, a compact candidate subset was extracted by interval-based filtering within each Pareto branch. Specifically, solutions in each branch were stratified into equal-mass bins,

and six solutions were retained per branch ($K = 12$ in total). This sampling strategy preserves diversity and branch representativeness while avoiding redundant points clustered in a narrow region of the front.

Second, TOPSIS was applied to the candidate subset using mass M and peak displacement L as cost-type criteria with equal weights ($\omega_M = \omega_L = 0.5$) [32]. This yields a closeness score and a ranked list, from which the highest-ranked candidate was taken as the representative design for engineering interpretation of the feasible trade-off set. Independent validation is then carried out in Section 3 using the hold-out runs that were kept entirely separate from surrogate training.

2.7.3. Candidate Selection for FE Reruns

To close the surrogate-to-FE decision loop, a two-candidate rerun protocol was defined. It included (i) one representative design selected by TOPSIS from the feasible Pareto set and (ii) one near-threshold design with PoF close to the adopted cutoff (~ 0.90) to test robustness at the feasibility boundary. An optional high-margin reference design (PoF ~ 0.95) was also considered. The final candidate IDs and FE rerun results are reported in Section 3.4.

2.8. Implementation Details and Reproducibility

Traceable results are emphasized under a limited simulation budget. All responses are obtained from 50 high-fidelity FE runs, which provide mass M , peak displacement L , and peak stress (*Stress*). These FE outputs are treated as the only ground truth used in the subsequent surrogate and selection steps.

A combined workflow is adopted to link FE evaluation with surrogate-assisted analysis. The FE model generates the response database under the side-pole rigid-pole extrusion condition. Two surrogate families are considered. BART is implemented in Python 3.12 (PyMC 5.27.0 with PyMC-BART 0.11.0) and serves as the primary model, because it offers both nonlinear regression capability and posterior predictive uncertainty required by the selection rule. GPR is implemented in Python and is used as an auxiliary baseline to perform a surrogate-dependence check, ensuring that the main observations are not driven by a single modeling choice. PyMC uses the PyTensor automatic-differentiation backend to support gradient-based samplers (e.g., NUTS) for continuous parameters. In this study, all variables and responses are real-valued. Therefore, only standard real-valued gradients are needed, and no complex-valued differentiation is involved.

The full DOE budget is fixed at 50 FE runs. Within this dataset, only 10 runs are regarded as the sequentially deployable budget and are introduced in two rounds ($5 + 5$). The remaining 3 runs form a hold-out set reserved for independent evaluation and are excluded from training and selection throughout. Accordingly, the data are partitioned into a 37/5/5/3 split and assessed by retrospective replay:

- Stage 1: train on 37 runs and evaluate on the next batch of 5 runs;
- Stage 2: train on 42 runs ($37 + 5$) and evaluate on the next batch of 5 runs;
- Stage 3: train on 47 runs ($37 + 5 + 5$) and evaluate on the hold-out 3 runs.

This setup supports the learning-curve evidence (A) under a fixed budget while maintaining an untouched subset for an independent check.

At each stage, input scaling is performed using statistics computed from the current training subset only, avoiding any leakage from later batches. The replay partition is stored explicitly as a run-ID split file, and fixed random seeds are used for BART sampling and any stochastic components. Stage-specific seeds are generated deterministically from a base seed and the corresponding training size, so that repeated runs follow the same replay trajectory.

In the sequential module, Stress is treated as a feasibility constraint. Candidate ranking follows a feasibility-aware criterion of the form $\text{PoF} \times \text{uncertainty}$ (B), prioritizing points that are likely feasible while still informative for model improvement. For transparency and downstream visualization, the workflow exports stage-wise error summaries (RMSE/NRMSE/ R^2), per-point predictions with uncertainty, learning-curve figures, and PoF/score diagnostic plots used for ranking.

The key settings that determine reproducibility are summarized in Table 6.

Table 6. Key implementation settings for reproducibility.

Item	Setting
DOE budget	50 FE runs
Replay protocol	Fixed 37/5/5/3 split; hold-out (3) never used for training/selection
Sequential budget	10 runs (5 + 5)
Responses	M, L, Stress
Primary surrogate	BART (PyMC + PyMC-BART)
Cross-check surrogate	GPR (Python, scikit-learn) for surrogate-dependence check
Software versions	pymc = 5.27.0, pymc-bart = 0.11.0
seed (base)	2025
BART trees	$m_trees = 50$
MCMC sampling	tune = 400, draws = 400 (per chain)
Chains/cores	Chains/cores 2/2 (training); inference uses posterior predictive evaluation only
Posterior predictive draws	Pred_draws = 400
Performance metric	NRMSE

Overall, this section establishes a traceable surrogate-assisted workflow under a fixed budget of 50 FE runs. Surrogate behavior is assessed via a prequential replay to track accuracy changes as additional data are introduced, while feasibility-guided enrichment targets decision-relevant regions using a PoF-weighted uncertainty criterion with Stress treated as the constraint. Post-training efficiency is quantified through a reproducible benchmark using stored posterior samples. Inference is performed via posterior predictive evaluation only (sans MCMC sampling), separating the one-time overhead (graph reconstruction) from the steady-state latency. Results confirm that surrogate evaluation is orders of magnitude faster than the explicit FE solver. The posterior predictive draw count matches Table 6, and repeated trials are used to summarize median and p95 statistics. The timing results and the FE-to-surrogate acceleration ledger are summarized in Section 3.5.

3. Results

3.1. Hold-Out Validation of Surrogates

The 50-run DOE is organized into four parts: 37 initial runs, two enrichment stages of 5 runs each, and a final hold-out set of 3 runs reserved exclusively for independent validation. The hold-out data are kept completely separate from surrogate fitting and any model-selection decisions, and are used only for independent validation of the final surrogate.

The hold-out results indicate that the peak displacement L is predicted with high accuracy, with relative deviations around one percent across all three cases. By comparison, Stress shows substantially larger errors, reaching roughly twenty percent for one case. Such behavior is consistent with a contact-dominated quantity that depends strongly on localized deformation, stress concentration, and small variations in contact evolution.

These findings support role separation within the workflow. The displacement surrogate is sufficiently accurate to serve as the objective evaluator for mapping the mass-

intrusion compromise, whereas Stress is not suited to be enforced as a deterministic point constraint. Instead, stress acceptability is handled through an uncertainty-aware feasibility criterion.

3.2. Feasible Pareto Set and Trade-Offs

Figure 7 shows the obtained mass–displacement trade-off at $-40\text{ }^{\circ}\text{C}$ under the probabilistic stress constraint, where feasibility is defined by $\text{PoF}_{\text{Stress}}(x) \geq \eta$ ($\eta = 0.90$). As expected, reducing mass generally increases the peak displacement, while tighter displacement control requires additional material. The PoF filter is most influential in the low-mass region, where stress margins are narrow and uncertainty becomes non-negligible; therefore, designs that look attractive under mean predictions may still be rejected when feasibility confidence is enforced. The feasible set in Figure 7 is thus treated as an engineering-relevant candidate pool for subsequent selection and verification, rather than a purely mathematical Pareto boundary.

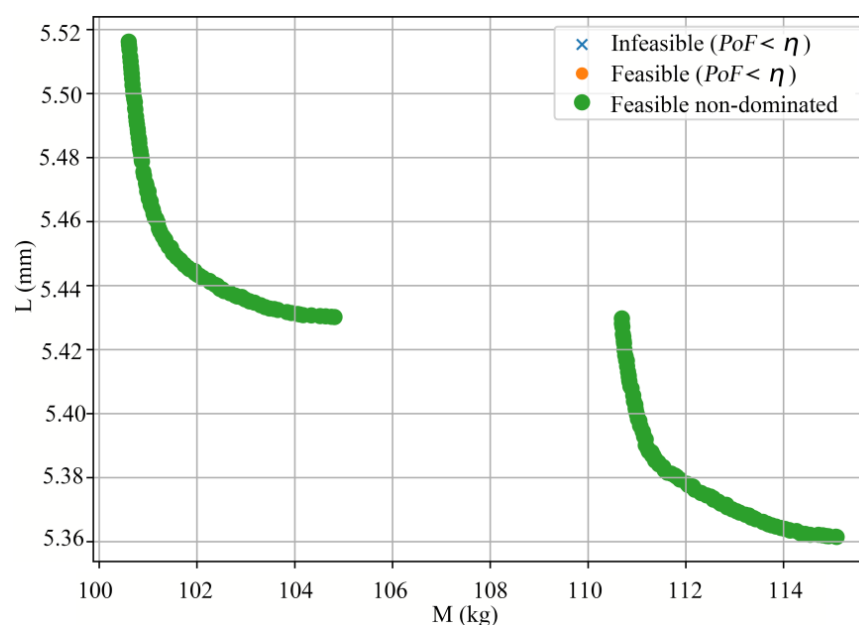


Figure 7. Feasibility-aware Pareto set at $-40\text{ }^{\circ}\text{C}$.

Appendix A, Figure A1 illustrates the distribution of $\text{PoF}_{\text{Stress}}$ across the feasible Pareto pool and marks the adopted screening threshold, $\eta = 0.90$.

A notable feature is that the feasible non-dominated set appears as two separate segments. This split is clarified in Figure 8: the two segments correspond to two distinct levels of the key thickness variable x_4 , indicating two competing design regimes. In the current dataset, x_4 exhibits a bimodal distribution (Figure A2), and the branch statistics are summarized in Table 7. For PoF, Table 7 lists the branch-wise range and the corresponding dispersion (sample standard deviation) over the non-dominated solutions. Practically, increasing x_4 shifts the response to a stiffer deformation pattern, leading to a lower displacement at the expense of higher mass, while the PoF constraint removes part of the intermediate region where feasibility confidence does not meet η . The full branch statistics, including the predicted mean and uncertainty of stress, are provided in Supplementary Material (Table S2).

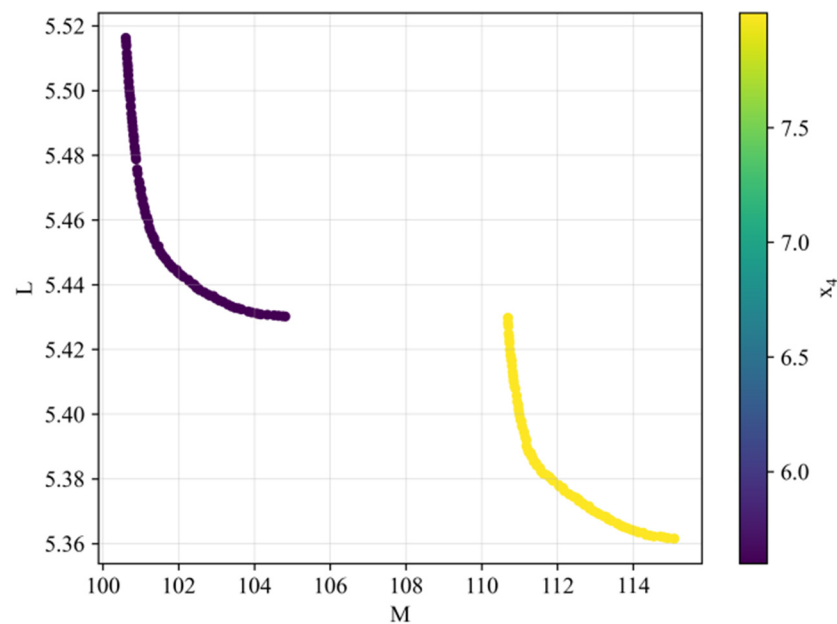


Figure 8. Pareto set colored by x_4 .

Table 7. Statistics of the two Pareto branches.

Branch	n	M (kg)	L (mm)	x_4 (mm)	PoF (Min–Max)	Pof Std
I (low mass)	106	100.61–104.81	5.430–5.516	5.599–5.601	0.900–0.922	0.004685
II (low L)	94	110.69–115.08	5.362–5.430	7.995–8.000	0.900–0.973	0.022131

Note: PoF is summarized by the min–max interval; PoF std is the sample standard deviation across solutions in each branch.

Although both branches follow the same -40 °C rigid-pole extrusion setup and termination rule, they represent two stiffness-controlled deformation regimes governed primarily by the lower-case thickness x_4 . Branch I (lower mass, thinner x_4) is more compliant in out-of-plane bending. It promotes a more global deformation of the enclosure floor–side assembly and spreads the pole-induced load over a broader load path. This regime therefore behaves like a global-compliance response: intrusion reduction is achieved mainly by increasing overall structural stiffness rather than by tuning local contact stiffness.

Branch II (lower L , thicker x_4) increases local bending stiffness and constraint near the pole–contact region. It tends to suppress local instability and section ovalization, and it shifts the response toward a more localized indentation mode. Load transfer is redirected more effectively into adjacent stiff members (e.g., side rails and cross members) along the load path. This yields slightly lower intrusion at the cost of higher mass, and it can also increase sensitivity to contact-driven stress peaks.

Feasibility is therefore governed by a localized peak response, even when the overall displacement pattern is broadly similar across designs (Figure 9). This local–global contrast is consistent with the FE reruns: stress concentrations remain confined to the pole-contact zone of the lateral anti-collision beam (Figure 10).

Feasibility is screened using the probabilistic stress gate defined in Section 2.6.2 ($\eta = 0.90$) together with an intrusion-based packaging-clearance check. This screening step filters out designs that achieve lower intrusion only by inducing excessive stress concentration on the lateral anti-collision beam, thereby yielding a more defensible feasible Pareto set for downstream selection. Notably, the retained solutions become sparse near the extreme low-mass end, implying that stress compliance is the primary limiting factor in that region.

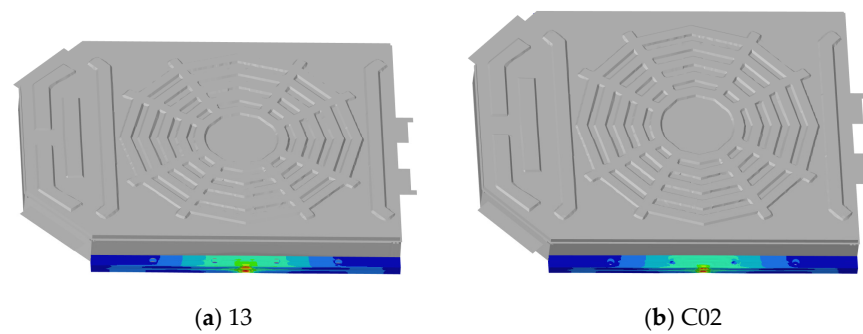


Figure 9. Displacement contours at $-40\text{ }^{\circ}\text{C}$: baseline (Run 13) vs. C02. Note: Contour colors indicate displacement magnitude, where blue and red denote the minimum and maximum values, respectively, and intermediate colors represent intermediate displacement levels.

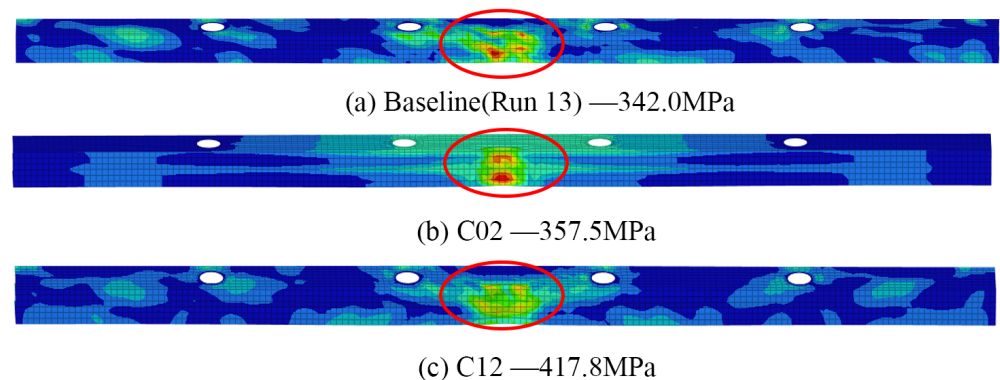


Figure 10. $-40\text{ }^{\circ}\text{C}$ FE stress contours on the lateral anti-collision beam. Colors represent stress magnitude (blue–red: low–high); red circles indicate the pole-contact hotspot for comparison only; a common color scale is used for (a)–(c).

3.3. Candidate Screening and TOPSIS Selection

To reduce sensitivity to a single surrogate formulation, the same candidate subset was re-evaluated using PI-GPR, and the TOPSIS ranking was recomputed. Agreement between the two rankings was quantified by Kendall's τ ($\tau = 0.606$ in this study), indicating moderate ranking consistency. Candidate information and the TOPSIS outcomes before and after the re-evaluation are summarized in Table 8. In this study, C02 from Branch I is adopted as the representative design produced by the decision procedure. The final surrogates are first validated on the reserved hold-out set, followed by additional FE reruns for selected surrogate-recommended candidates (C02 and C12, with C09 included as an optional high-margin case) [33]. The full candidate information is provided as Supplementary Table S3.

Because many feasible Pareto solutions are practically indistinguishable, the set is reduced to a compact candidate pool using interval-based filtering applied separately within each Pareto branch. Six candidates are retained per branch, giving $K = 12$. TOPSIS is then performed on this screened subset, using (M, L) as cost-type criteria with equal weights, $\omega_M = \omega_L = 0.5$.

To reduce sensitivity to a single surrogate family, the same candidates are re-evaluated with PI-GPR and the TOPSIS ranking is recomputed. The agreement between the two rankings, quantified by Kendall's $\tau = 0.606$, indicates moderate stability with limited swapping among the leading candidates. Under this decision chain, design C02 from Branch I is selected as the representative solution for subsequent interpretation.

Table 8. Candidate ranking by TOPSIS with PI-GPR re-evaluation.

ID	Branch	x_4	M	L	PoF _{Stress}	TOPSIS	TOPSIS_rank	PI_rank	rank_diff
C02	I (low mass)	5.599	101.967	5.444	0.903	0.881	1	3	2
C01	I (low mass)	5.599	101.301	5.455	0.903	0.88	2	6	4
C03	I (low mass)	5.599	102.677	5.438	0.901	0.858	3	5	2
C04	I (low mass)	5.6	103.394	5.433	0.902	0.822	4	4	0
C05	I (low mass)	5.6	104.102	5.431	0.903	0.779	5	2	−3
C06	I (low mass)	5.6	104.814	5.43	0.906	0.732	6	1	−5
C07	II (low L)	7.999	111.375	5.385	0.913	0.282	7	7	0
C08	II (low L)	7.999	112.146	5.377	0.92	0.235	8	8	0
C09	II (low L)	7.999	112.845	5.371	0.949	0.196	9	9	0
C10	II (low L)	7.999	113.579	5.366	0.914	0.16	10	11	1
C11	II (low L)	7.999	114.339	5.363	0.9	0.133	11	10	−1
C12	II (low L)	7.999	115.085	5.362	0.9	0.12	12	12	0

3.4. High-Fidelity FE Reruns and Acceptance Outcomes

Table 9 presents the independent validation results for the hold-out set with $n = 3$, where surrogate estimates are benchmarked against the corresponding high-fidelity FE outputs. For each response $y \in \{M, L, \text{Stress}\}$, the absolute and relative deviations are defined as

$$\Delta y = y_{\text{FE}} - y_{\text{sur}} \quad (19)$$

Table 9 indicates that the surrogate reproduces the peak displacement L on the hold-out runs with good accuracy, with relative deviations on the order of 1% for all three cases. For mass M , a small systematic offset is observed on this subset. The two heavier hold-out designs are slightly underestimated by about 4%, whereas the remaining case is overestimated by about 2%, giving an average absolute relative deviation of approximately 3–4%. This discrepancy is acceptable for the present analysis because M is retained mainly to keep the surrogate workflow consistent with the DOE-based learning procedure. In engineering use, mass can be obtained deterministically from the geometry and material density once the design variables are fixed.

Table 9. Surrogate–FE comparison on the hold-out set.

ID	M_{sur} (kg)	M_{FE} (kg)	ΔM (kg)	δM (%)	L_{sur} (mm)	L_{FE} (mm)	ΔL (mm)	δL (%)	$\text{Stress}_{\text{sur}}$ (MPa)	$\text{Stress}_{\text{FE}}$ (MPa)	ΔStress (MPa)	δStress (%)
34	156.136	162.861	−6.725	−4.13	6.246	6.308	−0.062	−0.98	455.434	413.9	41.534	10.03
40	151.193	157.611	−6.418	−4.07	5.513	5.552	−0.039	−0.71	418.91	342.2	76.71	22.42
48	134.531	131.522	3.009	2.29	6.157	6.1	0.057	0.94	408.575	447.9	−39.325	−8.78

Note: Surrogate predictions are generated by the final model trained on 47 non-hold-out runs; the hold-out runs (IDs 34, 40, 48) are not used for training or tuning.

The stress response exhibits larger deviations, which is consistent with its contact-driven nature and its sensitivity to local deformation patterns, stress concentrations, and minor changes in contact evolution. Taken together, the hold-out benchmark provides an unbiased assessment of generalization and supports using the final surrogates as the evaluation engine for feasibility-aware optimization and candidate screening in subsequent analyses.

To complement the tabulated error metrics, Figure A3 presents parity plots comparing the surrogate posterior mean with the FE results for both the hold-out cases and the rerun candidates. The peak displacement predictions closely follow the 1:1 line, whereas the stress response exhibits a visibly broader spread. This contrast motivates the use of a probabilistic feasibility screen in the subsequent selection to mitigate the risk of choosing designs near the stress limit.

Following the protocol in Section 2.7.3, two surrogate-selected candidates were prioritized for independent high-fidelity FE reruns. C02 (Branch I) was chosen as the TOPSIS representative from the feasible Pareto set. C12 (Branch II) was chosen as a near-threshold case ($\text{PoF} \approx 0.90$) to assess the conservativeness of the PoF gate and the risk of borderline misclassification. A higher-margin Branch II design, C09 ($\text{PoF} \approx 0.95$), was retained as an optional within-branch reference.

The reruns strictly follow the FE setup specified earlier, including geometry, boundary conditions, contact formulation, -40°C material properties, and the response extraction procedure; these details are not repeated here. For each candidate, the same termination rule is applied (simulation is terminated when the reaction force reaches 100 kN or the prescribed travel reaches 30%, whichever occurs first). The mass M , peak intrusion L , and peak stress (Stress) are extracted using the same definitions adopted in the DOE simulations.

For surrogate reporting, M_{sur} and L_{sur} correspond to the posterior predictive means from the final BART models, while $\text{Stress}_{\text{sur}}$ is reported as the posterior mean. The feasibility probability $\text{PoF}_{\text{Stress}}(\mathbf{x})$ is computed by Monte Carlo counting over BART posterior predictive draws as defined in Section 2.6.2.

Table 10 summarizes the surrogate–FE comparison for three surrogate-selected candidates, C02, C09, and C12, under the -40°C side-pole rigid-pole extrusion scenario. For each response $y \in \{M, L, \text{Stress}\}$, the relative deviation is computed in the same way as in the hold-out benchmark (Table 9), described as follows:

$$\delta_y(\%) = \frac{y_{\text{sur}} - y_{\text{FE}}}{y_{\text{FE}}} \times 100 \quad (20)$$

These reruns are decision-oriented rather than purely diagnostic. They directly verify the designs recommended by surrogate-assisted optimization and quantify the risk associated with stress prediction, particularly for candidates close to the feasibility threshold.

To ensure comparability, all FE reruns replicate the DOE setup, including geometry definition, boundary conditions, contact parameters, -40°C material properties, and response extraction rules. Mass is reported for the battery-pack enclosure only, excluding the rigid pole and fixtures, consistent with the DOE dataset used for surrogate training and optimization.

Table 10. FE reruns for selected candidates at -40°C .

Candidate	Branch	M_{sur}	M_{FE}	L_{sur}	L_{FE}	$\text{Stress}_{\text{sur}}$	$\text{Stress}_{\text{FE}}$	$\text{PoF}_{\text{Stress}}$	Verdict
C02	I	101.967	108.294	5.444	5.978	355.475	357.5	0.9027	Accept
C09	II	112.845	111.097	5.370	6.134	369.423	389.5	0.9493	Accept
C12	II	115.085	112.776	5.362	6.160	355.567	417.8	0.9001	Reject

Note: All masses M are reported for the battery-pack enclosure only; the rigid pole and test fixtures are excluded to remain consistent with the DOE dataset. Units are M in kg, L in mm, and Stress in MPa. $\text{Stress}_{\text{sur}}$ denotes the posterior mean predicted by the final BART stress surrogate. FE responses are obtained using the same response definitions and the same termination criterion employed in the DOE campaign. Pass/Fail is determined by $\text{Stress}_{\text{FE}} \leq \sigma_{\text{lim}}$ ($\sigma_{\text{lim}} = 410.4 \text{ MPa}$).

Table 10 demonstrates a high degree of concordance between the predicted mass and the FE validation results, with relative errors (δ_M) confined within approximately $\pm 6\%$ across the three candidates. In contrast, the peak displacement (L) is consistently underpredicted (δ_L ranging from -9% to -13%), suggesting that the surrogate model yields slightly optimistic displacement estimates within this local design space. Notwithstanding this systematic bias, the performance trends remain coherent: Branch II candidates are predicted to have a smaller $\$L\$, while FE reruns confirm comparable intrusion levels between the two Branch II points. Regarding stress constraints, candidates C02 and$

C09 remained within admissible limits during high-fidelity validation. Conversely, C12 exceeded the threshold ($\text{Stress}_{\text{FE}} = 417.8 \text{ MPa}$), despite possessing a Probability of Feasibility ($\text{PoF} \approx 0.90$) that satisfied the initial screening criterion. This observation also highlights threshold sensitivity. Using $\eta = 0.90$ can admit near-boundary candidates such as C12 ($\text{PoF} \approx 0.90$) that still carry residual failure risk under highly localized contact stresses. This discrepancy highlights a critical insight: candidates situated near the feasibility boundary entail a residual risk of failure, exacerbated by the highly localized nature of contact stress. The violation in C12 underscores the imperative of a probabilistic framework; a deterministic surrogate, by neglecting predictive variance, would have obscured such borderline risks. Consequently, for stress-critical components in battery safety engineering, adopting a more stringent PoF threshold (e.g., $\eta = 0.95$) is warranted to ensure structural integrity. In contrast, a stricter gate such as $\eta = 0.95$ would favor higher-margin designs (e.g., C09 with $\text{PoF} \approx 0.95$). This reduces false acceptance, but it may also exclude borderline lightweight candidates.

Figure 9 juxtaposes the -40°C displacement fields of the baseline (Run 13) and C02 at the 100 kN cut-off, to confirm that the surrogate-selected design maintains the same deformation mechanism while mitigating intrusion.

Relative to the baseline, C02 achieves a smaller peak intrusion while exhibiting a comparable global deformation pattern. This visual evidence agrees with the reduced peak displacement L reported in Table 10.

As illustrated in Figure 10, stress concentrations are confined to the pole-contact zone of the lateral anti-collision beam, suggesting that feasibility is dictated by a local peak response rather than the overall deformation pattern. In the FE reruns, C02 attains a maximum stress of 357.5 MPa, which satisfies the limit $\sigma_{\text{lim}} = 410.4 \text{ MPa}$, whereas C12 rises to 417.8 MPa and exceeds the cap. These results motivate the probabilistic feasibility screening adopted in the surrogate-assisted optimization, helping avoid recommendations that lie too close to the stress boundary.

In addition to demonstrating numerical consistency, the FE reruns in Table 10 provide direct confirmation of the surrogate-selected candidates using the same response definitions adopted throughout this work. Final engineering sign-off is based on a transparent pass–fail scheme that couples a baseline-referenced stress cap with a packaging-clearance requirement. Stress compliance is judged against σ_{lim} as defined in Section 2.6.2; under this check, C02 and C09 satisfy the limit, whereas C12 exceeds it in the rerun.

The clearance criterion is also met for all three candidates, since the predicted intrusions remain at the millimeter level and are well within the available packaging gap.

Taken together, these checks create a consistent bridge from PoF-based screening to the deterministic acceptance decision from explicit FE, thereby supporting C02 as the final representative design while keeping C09 as a feasible Branch II reference.

3.5. Sample Economy and Acceleration Ledger

High-fidelity FE runs for the -40°C side-pole case are time-consuming. With the adopted termination criterion, one FE evaluation takes about $t_{\text{FE}} \approx 23 \text{ min}$ on the workstation used in this study. After training, surrogate evaluation is sub-second: the median wall-clock time is 0.394 s per design (p95 0.595 s), and 0.160 s per design (p95 0.231 s) when averaged over batch scoring, including, PoFStress. Compared with $t_{\text{FE}} \approx 23 \text{ min}$, this implies an evaluation-level acceleration of roughly 3.5×10^3 (single-point median) to 8.6×10^3 (batch-amortized median), with conservative p95-based bounds of 2.3×10^3 to 6.0×10^3 .

During model setup, short pilot runs were performed to verify the termination response. Each pilot case was halted when the reaction force reached 100 kN, which in most

instances occurred within the first six recorded output steps. These pilot runs took about 23 min to complete and are excluded from the acceleration ledger, as they were not treated as production-grade high-fidelity evaluations.

To clarify the stopping rule, Figure 11 presents the reaction force–intrusion histories of the baseline (Run 13) and the surrogate-selected rerun candidates (C02 and C12). The horizontal line denotes the 100 kN termination threshold, and the marker indicates the first output step at which the recorded force exceeds this limit.

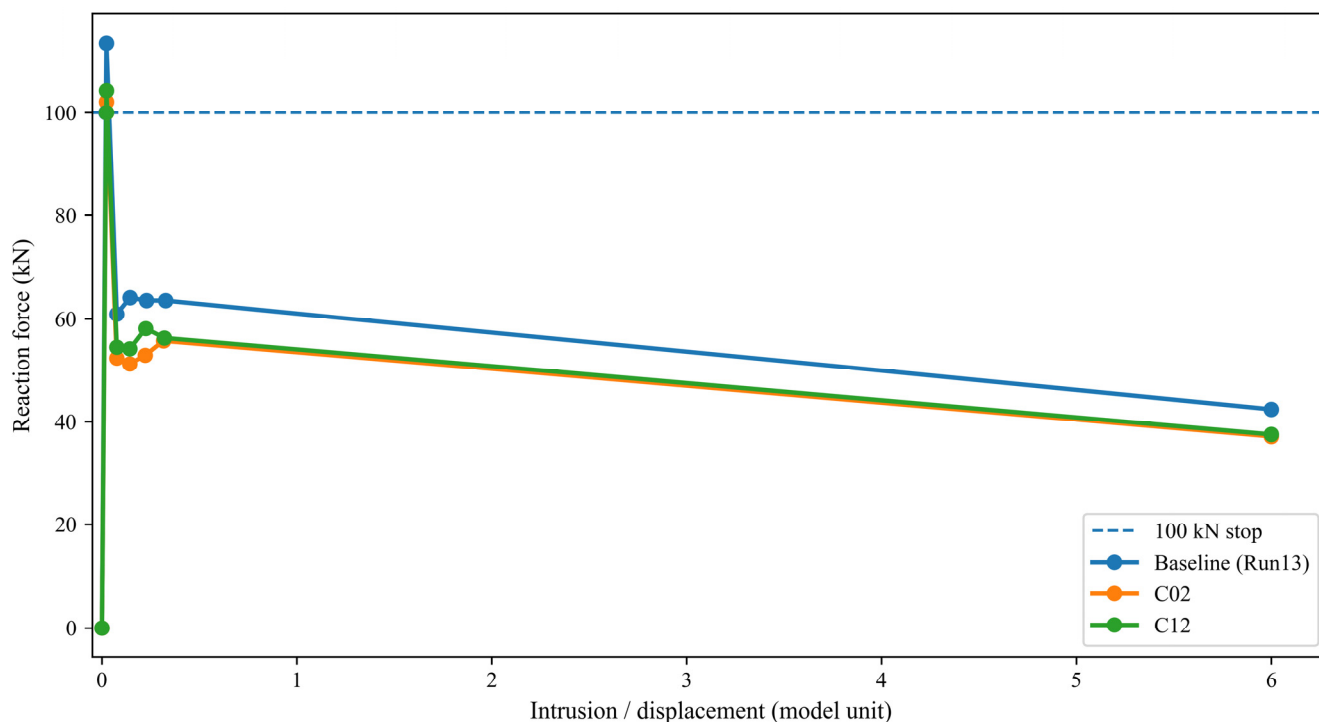


Figure 11. Force–displacement curves with a 100 kN cutoff.

4. Discussion

4.1. Implications for Feasibility Screening and Final Acceptance

Table 9 provides two observations that directly affect the surrogate-assisted decision process. First, the peak displacement L is reproduced consistently, indicating that the displacement surrogate is adequate for objective evaluation and for tracing the mass–intrusion compromise that forms the Pareto set in subsequent analyses. The second is that the stress response exhibits visibly larger deviations. This is not unexpected, as peak stress in a contact-dominated event is highly sensitive to local deformation modes, stress concentrations, and small differences in contact evolution. For this reason, a deterministic pass–fail rule based on a single stress point estimate would tend to admit borderline candidates that may exceed the limit once re-evaluated by explicit FE, and a probabilistic screen is preferred.

During optimization, feasibility is therefore imposed using the gate $\text{PoF}_{\text{Stress}}(x) \geq \eta$ with $\eta = 0.90$, so that the uncertainty of the stress surrogate is explicitly reflected in candidate screening. Final engineering acceptance, however, is determined deterministically from explicit FE outputs using the same response definitions and termination rule adopted in the DOE model. The stress requirement is checked against the baseline-referenced cap in Equation (8) (with the screening inequality in Equation (5), which gives $\sigma_{\text{lim}} = 410.4$ MPa (≈ 410 MPa). A packaging-clearance check is also applied by rejecting designs with excessive intrusion ($L > 20$ mm). Together, these checks yield a traceable pass–fail rule that is

consistent with the adopted constraint-handling strategy and can be used directly when reporting the outcomes of subsequent FE reruns.

The runtime results distinguish one-time overhead (posterior loading and graph reconstruction) from the steady-state per-design inference cost. Since FE analyses in this study are terminated at the 100 kN criterion, the reference cost t_{FE} reflects an early-terminated evaluation (≈ 23 min) rather than a full-length crash simulation. Although absolute timings vary with hardware and the number of posterior predictive draws, the settings used here keep inference in the sub-second range and support uncertainty-aware screening in routine use.

4.2. Applicability and Reproducibility Considerations

This section summarizes the outcomes of the surrogate-assisted workflow for the battery pack enclosure under the extreme-cold design condition $T = -40^\circ\text{C}$ in the side-pole rigid-pole extrusion scenario.

All results reported here correspond to the discrete extreme-cold condition $T = -40^\circ\text{C}$ under the adopted side-pole configuration, and no interpolation to other temperatures is assumed. Stress feasibility is enforced through $\text{PoF}_{\text{Stress}}$ to reflect surrogate uncertainty and to limit the admission of borderline candidates. Reproducibility is supported by fixed data partitions with split indices 37/5/5/3, an explicit replay protocol, and a deterministic screening-to-TOPSIS procedure once the random seed is specified. The full DOE dataset, screened candidate lists, and ranking outputs should be provided as Supplementary Materials to enable auditability and reuse.

4.3. Limitations and Robustness

This study relies on high-fidelity FE simulations parameterized via literature-derived material properties, in the absence of component-level physical crash validation for the specific enclosure designs. While the proposed Bayesian framework effectively quantifies the uncertainty propagated from the surrogate model, it cannot entirely eliminate the epistemic uncertainty inherent in the baseline material inputs. Future work will address this by integrating experimental coupon data directly into the Bayesian prior to bridge the discrepancy between literature values and in situ material behavior. Furthermore, the violation observed in the C12 candidate underscores that while the PoF gate enhances reliability compared to deterministic approaches, ‘borderline feasible’ designs necessitate rigorous confirmatory verification. Future work will incorporate material variability directly in FE by using coupon-informed parameter sampling and a small number of material-perturbed replicate runs to calibrate response-level variability.

CFRP components are modeled using linear orthotropic elasticity, without progressive damage or delamination. This choice keeps the explicit-dynamics DOE traceable and computationally feasible under a limited FE budget. It also avoids additional model-form uncertainty from damage parameters that are difficult to calibrate at -40°C , including intralaminar strengths, fracture energies, and cohesive interface properties for the adopted laminate and joints. Accordingly, the stress-based feasibility criterion used here should be interpreted as an elastic screening proxy, not as a definitive failure criterion.

If progressive damage were enabled (e.g., Hashin/Puck-type intralaminar criteria with stiffness degradation, together with cohesive-zone or continuum-damage formulations for delamination), three effects are expected. First, matrix-dominated damage may initiate earlier at low temperature. Feasibility defined by stress margin could then become more conservative when reformulated in terms of damage indices. Second, as damage accumulates, local stress hot-spots may be reduced by load redistribution, while stiffness degradation increases global compliance. As a result, intrusion L may increase for designs

near the constraint boundary, meaning the elastic model can be slightly optimistic for borderline cases. Third, feasibility would be more appropriately expressed using damage-relevant metrics (e.g., peak failure index, damage variables, or delamination area/energy) rather than a single von Mises stress scalar. The same uncertainty-aware PoF framework can be applied to these metrics, but PoF values and ranking near the threshold are expected to change. This motivates confirmatory reruns for borderline designs, as illustrated by C12.

Despite this limitation, the observed two-regime Pareto structure governed by x_4 is primarily driven by global load-path stiffness and contact kinematics. It is therefore expected to remain qualitatively stable. In contrast, stress-margin ranking and near-threshold feasibility may shift once damage is enabled, which we view as the logical next step toward design sign-off.

In addition, strain-based metrics and beam energy absorption provide practical alternatives to stress for characterizing crash response, and they can be screened using the same PoF framework once the corresponding response definitions are established.

5. Conclusions

This study develops a traceable surrogate-assisted decision workflow for the $-40\text{ }^{\circ}\text{C}$ side-pole design scenario of a battery pack enclosure under a constrained high-fidelity FE budget. The main conclusions are summarized as follows.

(1) Surrogate usefulness is response-dependent.

Evidence from the strict hold-out evaluation supports using peak intrusion L as a dependable objective evaluator for Pareto exploration. In contrast, peak stress is more uncertainty-prone, which is consistent with its contact-driven localization and sensitivity to stress concentration.

(2) Stress feasibility is best enforced probabilistically.

When the stress surrogate carries non-negligible uncertainty, a probability-of-feasibility criterion, for example $\text{PoF}_{\text{Stress}}(x) \geq 0.90$, is more defensible than a hard cutoff applied to point predictions. This strategy naturally separates exploratory screening from final pass–fail decisions based on explicit FE confirmation. Using $\eta = 0.95$ provides a more conservative gate for stress-critical sign-off. It reduces false acceptance of near-boundary designs such as C12, but it may also reject some borderline lightweight candidates.

(3) The feasible Pareto set exposes interpretable regimes.

The feasible nondominated solutions exhibit distinct trade-off regimes that can be linked to a dominant thickness variable acting as a regime selector, such as the lower-case thickness parameter. This mechanism explains why the feasible front appears segmented rather than forming a single continuous curve.

(4) Transferability is workflow-level, not surrogate-level.

The workflow can be transferred to other temperatures, but the trained surrogate cannot. Extending the framework requires (i) a temperature-specific anchor property set (and uncertainty envelope) derived from coupon data; (ii) a new or augmented FE DOE at the target temperature to retrain the surrogate; and (iii) re-evaluation of feasibility thresholds under the same termination rule. A multi-temperature extension would additionally require DOE coverage across temperatures, with temperature included as an input variable.

(5) The compute-to-coverage gain is substantial.

Production-grade explicit FE typically costs $t_{\text{FE}} \approx 30$ min per run (and can be longer). Surrogate inference is sub-second. This yields an evaluation-level speedup of about $10^4 \sim 10^5$.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/batteries12020061/s1>, Table S1: 50-run DOE/FE dataset; Table

S2: Full statistics of the two Pareto branches; Table S3: Full list of feasible Pareto candidates; Code: Python 3.12 scripts for data processing and figure/table generation.

Author Contributions: All authors contributed to the study's conception and design. Conceptualization, D.Z., J.L. and Z.S.; methodology, Z.S. and D.Z.; software, Z.S., H.Z. and J.L.; validation, L.W., Z.S. and D.Z.; formal analysis, Z.S.; investigation, L.W. and Z.S.; resources, D.Z.; data curation, Z.S. and L.W.; writing—original draft preparation, Z.S.; writing—review and editing, D.Z., H.Z., J.L. and L.W.; visualization, Z.S.; supervision, D.Z.; project administration, D.Z.; funding acquisition, D.Z. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was Supported by Heilongjiang Provincial Natural Science Foundation of China (Grant No. LH2022E100) and the Self-financed Project of Harbin Science and Technology Program (Grant No. 2023ZCZJCG036).

Data Availability Statement: The code supporting the findings of this study is available in the Supplementary Materials (Code S1).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

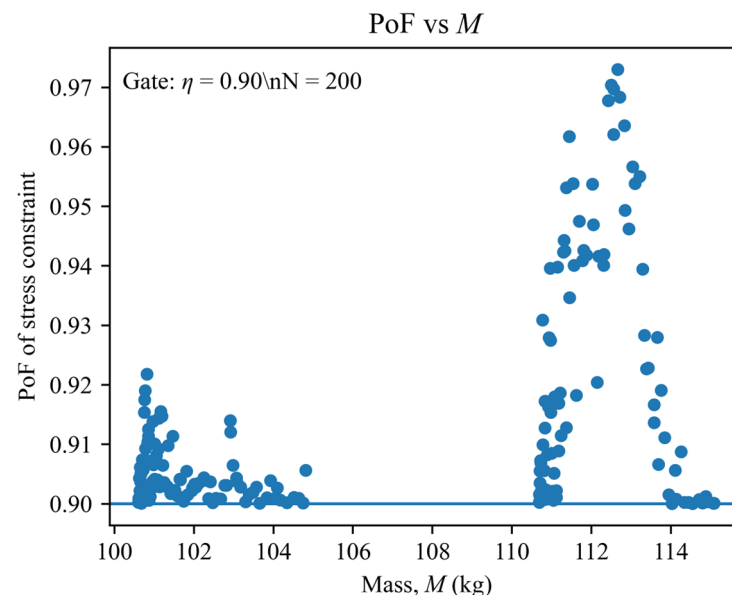


Figure A1. Visualization of the PoF-based feasibility gate ($\eta = 0.90$). Dots denote candidate designs, and the horizontal line indicates the PoF feasibility threshold ($\eta = 0.90$).

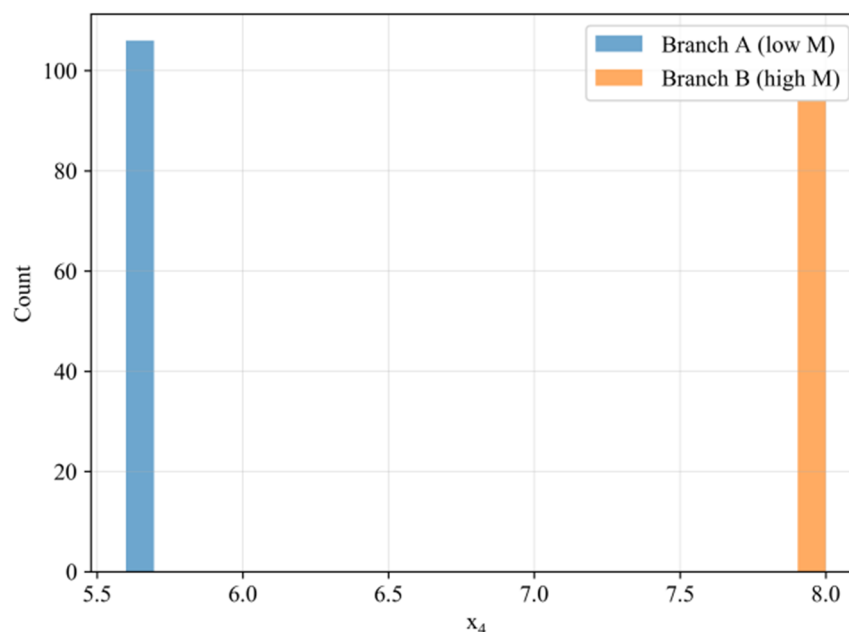


Figure A2. Bimodal distribution of x_4 supporting the branch separation.

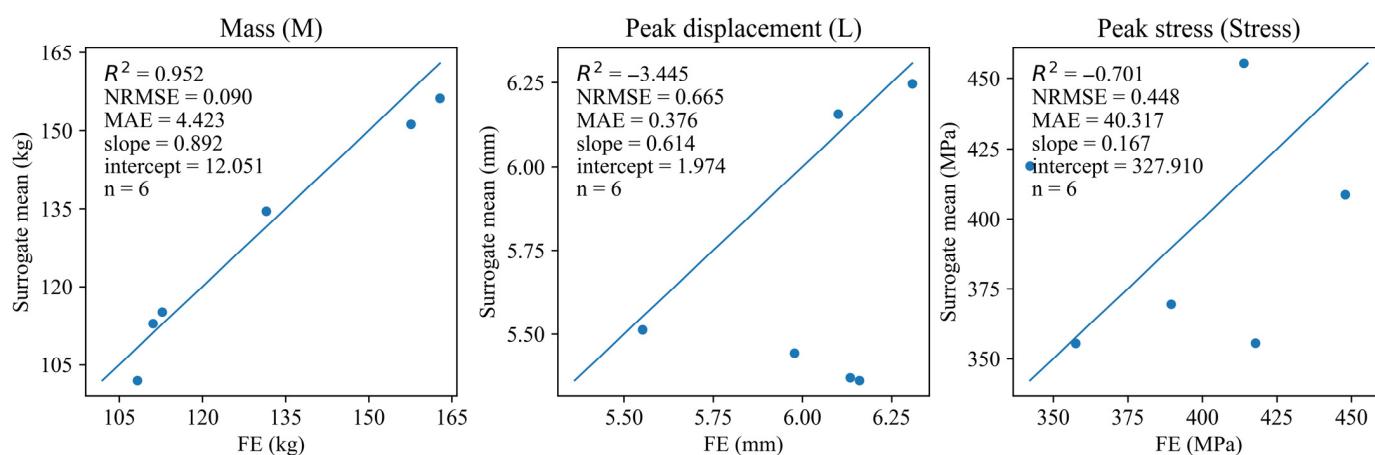


Figure A3. Parity plots (surrogate mean vs. FE) for M , L , and $Stress$. Dots denote validation samples, and the solid line represents the ideal parity line ($y = x$).

References

1. Sánchez-Sáez, S.; Gómez-del Río, T.; Barbero, E.; Zaera, R.; Navarro, C. Static behavior of CFRPs at low temperatures. *Compos. Part B Eng.* **2002**, *33*, 383–390. [\[CrossRef\]](#)
2. Gómez-del Río, T.; Zaera, R.; Barbero, E.; Navarro, C. Damage in CFRPs due to low velocity impact at low temperature. *Compos. Part B Eng.* **2005**, *36*, 41–50. [\[CrossRef\]](#)
3. GB 38031-2025; Electric Vehicles Traction Battery Safety Requirements. Standardization Administration of China: Beijing, China, 2025.
4. Kawsar, I.; Li, H.; Liu, B.; Zhang, Y.; Pan, Y. Enhancing mechanical reliability and safety performance of a battery pack system for electric vehicles: A review. *ETransportation* **2025**, *26*, 100469. [\[CrossRef\]](#)
5. Mortazavi Moghaddam, A.; Kheradpisheh, A.; Asgari, M. An integrated energy absorbing module for battery protection of electric vehicle under lateral pole impact. *Int. J. Crashworthiness* **2023**, *28*, 321–333. [\[CrossRef\]](#)
6. Chen, J.; Bian, K.; Mao, H. *Analyzing EV Battery Package Responses During Side Pole Impacts with Multiple Speeds and Locations*; SAE Technical Paper: Warrendale, PA, USA, 2025.
7. Muresanu, A.D.; Dudescu, M.C.; Tica, D. Study on the Crashworthiness of a Battery Frame Design for an Electric Vehicle Using FEM. *World Electr. Veh. J.* **2024**, *15*, 534. [\[CrossRef\]](#)
8. Zhang, D.; Sun, Z.; Zhang, H.; Liao, J. Manufacturability-Constrained Multi-Objective Optimization of an EV Battery Pack Enclosure for Side-Pole Impact. *World Electr. Veh. J.* **2025**, *16*, 632. [\[CrossRef\]](#)

9. Sun, G.; Zuo, W.; Chen, D.; Luo, Q.; Pang, T.; Li, Q. On the effects of temperature on tensile behavior of carbon fiber reinforced epoxy laminates. *Thin-Walled Struct.* **2021**, *164*, 107769. [\[CrossRef\]](#)
10. Im, K.-H.; Cha, C.-S.; Kim, S.-K.; Yang, I.-Y. Effects of temperature on impact damages in CFRP composite laminates. *Compos. Part B Eng.* **2001**, *32*, 669–682. [\[CrossRef\]](#)
11. Sánchez-Sáez, S.; Barbero, E.; Navarro, C. Compressive residual strength at low temperatures of composite laminates subjected to low-velocity impacts. *Compos. Struct.* **2008**, *85*, 226–232. [\[CrossRef\]](#)
12. Kulkarni, S.S.; Hale, F.; Taufique, M.; Soulami, A.; Devanathan, R. Investigation of crashworthiness of carbon fiber-based electric vehicle battery enclosure using finite element analysis. *Appl. Compos. Mater.* **2023**, *30*, 1689–1715. [\[CrossRef\]](#)
13. Zhang, X.; Lin, Q.; Xiao, Y.; Jia, L.; Yang, T.; Wang, L.; Ma, Q.; Wang, B. Top-Down Design Approach of Lightweight Composite Battery Pack Enclosure for Electric Vehicles Based on Numerical Modeling and Topology Optimization. *Polymers* **2025**, *17*, 2897. [\[CrossRef\]](#)
14. Shaikh, S.A.; Taufique, M.F.N.; Balusu, K.; Kulkarni, S.S.; Hale, F.; Oleson, J.; Devanathan, R.; Soulami, A. Finite element analysis and machine learning guided design of carbon fiber organosheet-based battery enclosures for crashworthiness. *Appl. Compos. Mater.* **2024**, *31*, 1475–1493. [\[CrossRef\]](#)
15. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2015**, *104*, 148–175. [\[CrossRef\]](#)
16. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [\[CrossRef\]](#)
17. Forrester, A.; Sobester, A.; Keane, A. *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
18. Gelbart, M.A.; Snoek, J.; Adams, R.P. Bayesian optimization with unknown constraints. *arXiv* **2014**, arXiv:1403.5607. [\[CrossRef\]](#)
19. Hernández-Lobato, J.M.; Gelbart, M.; Hoffman, M.; Adams, R.; Ghahramani, Z. Predictive entropy search for Bayesian optimization with unknown constraints. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 1699–1707.
20. Picheny, V. A stepwise uncertainty reduction approach to constrained global optimization. In Proceedings of the Artificial Intelligence and Statistics, Reykjavik, Iceland, 22–25 April 2014; pp. 787–795.
21. Theodore, A.M.; Şahin, M.E. Modeling and simulation of a series and parallel battery pack model in MATLAB/Simulink. *Turk. J. Electr. Power Energy Syst.* **2024**, *4*, 2–12. [\[CrossRef\]](#)
22. McKay, M.D.; Beckman, R.J.; Conover, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **2000**, *42*, 55–61. [\[CrossRef\]](#)
23. Sun, S.S.; Fang, Y.W.; Wen, Z.T. The Influence of High and Low Temperature on the Shear Properties of Carbon Fiber Reinforced Plastics. *Fiber Glass* **2023**, *52*, 16–19. [\[CrossRef\]](#)
24. Sun, S.S.; Wen, Z.T.; Feng, G.M.; Li, J.; Fang, Y.W.; Hao, Z.T.; Chen, J.M. Effects of High and Low Temperature on Mechanical Properties of Carbon Fiber Reinforced Plastics. *Fiber Glass* **2023**, *52*, 7–13. [\[CrossRef\]](#)
25. Xi, R.; Xie, J.; Yan, J.-B. Evaluations of low-temperature mechanical properties and full-range constitutive models of AA 5083-H112/6061-T6. *Constr. Build. Mater.* **2024**, *411*, 134520. [\[CrossRef\]](#)
26. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and analysis of computer experiments. *Stat. Sci.* **1989**, *4*, 409–423. [\[CrossRef\]](#)
27. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298. [\[CrossRef\]](#)
28. An, J.; Lee, H.; Kim, C.-W. Weight Minimization of Type 2 Composite Pressure Vessel for Fuel Cell Electric Vehicles Considering Mechanical Safety with Kriging Metamodel. *Machines* **2024**, *12*, 132. [\[CrossRef\]](#)
29. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [\[CrossRef\]](#)
30. Dong, S.; Jing, T.; Zhang, J. A few-shot learning-based crashworthiness analysis and optimization for multi-cell structure of high-speed train. *Machines* **2022**, *10*, 696. [\[CrossRef\]](#)
31. Guo, W.; Xu, P.; Yi, Z.; Xing, J.; Zhao, H.; Yang, C. Variable stiffness design and multiobjective crashworthiness optimization for collision post of subway cab cars. *Machines* **2021**, *9*, 246. [\[CrossRef\]](#)
32. Behzadian, M.; Otaghsara, S.K.; Yazdani, M.; Ignatius, J. A state-of-the-art survey of TOPSIS applications. *Expert Syst. Appl.* **2012**, *39*, 13051–13069. [\[CrossRef\]](#)
33. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.