

Transfer learning in Bayesian optimization for aircraft design

A. Tfaily, Y. Diouane, N. Bartoli, M. Kokkolaras

G-2026-15

March 2026

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

Citation suggérée : A. Tfaily, Y. Diouane, N. Bartoli, M. Kokkolaras (Mars 2026). Transfer learning in Bayesian optimization for aircraft design, Rapport technique, Les Cahiers du GERAD G-2026-15, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2026-15>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

Suggested citation: A. Tfaily, Y. Diouane, N. Bartoli, M. Kokkolaras (March 2026). Transfer learning in Bayesian optimization for aircraft design, Technical report, Les Cahiers du GERAD G-2026-15, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (<https://www.gerad.ca/en/papers/G-2026-15>) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2026
– Bibliothèque et Archives Canada, 2026

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2026
– Library and Archives Canada, 2026

Transfer learning in Bayesian optimization for aircraft design

Ali Tfaily ^{a, b, c}

Nathalie Bartoli ^e

Youssef Diouane ^{c, d}

Michael Kokkolaras ^{c, f}

^a *Department of Mechanical Engineering, McGill University, Montréal (Qc), Canada, H3A 0C3*

^b *Advanced Product Development, Bombardier, Dorval (Qc), Canada, H4S 1Y9*

^c *GERAD, Montréal (Qc), Canada, H3T 1J4*

^d *Department of Mathematical and Industrial Engineering, Polytechnique Montréal, Montréal (Qc), Canada, H3T 1J4*

^e *ONERA/DTIS, Université de Toulouse, 31055 Toulouse, France*

^f *School of Mechanical and Automotive Engineering, Clemson University, Greenville SC 29607, USA*

ali.tfaily@mail.mcgill.ca
youssef.diouane@polymtl.ca
nathalie.bartoli@onera.fr
mkokkol@clemson.edu

March 2026
Les Cahiers du GERAD
G–2026–15

Copyright © 2026 Tfaily, Diouane, Bartoli, Kokkolaras

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs. Les auteurs conservent leur droit d'auteur et leurs droits moraux sur leurs publications et les utilisateurs s'engagent à reconnaître et respecter les exigences légales associées à ces droits. Ainsi, les utilisateurs:

- Peuvent télécharger et imprimer une copie de toute publication du portail public aux fins d'étude ou de recherche privée;
- Ne peuvent pas distribuer le matériel ou l'utiliser pour une activité à but lucratif ou pour un gain commercial;
- Peuvent distribuer gratuitement l'URL identifiant la publication.

Si vous pensez que ce document enfreint le droit d'auteur, contactez-nous en fournissant des détails. Nous supprimerons immédiatement l'accès au travail et enquêterons sur votre demande.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*. Copyright and moral rights for the publications are retained by the authors and the users must commit themselves to recognize and abide the legal requirements associated with these rights. Thus, users:

- May download and print one copy of any publication from the public portal for the purpose of private study or research;
- May not further distribute the material or use it for any profit-making activity or commercial gain;
- May freely distribute the URL identifying the publication.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Abstract : The use of transfer learning within Bayesian optimization addresses the disadvantages of the so-called *cold start* problem by using source data to aid in the optimization of a target problem. We present a method that leverages an ensemble of surrogate models using transfer learning and integrates it in a constrained Bayesian optimization framework. We identify challenges particular to aircraft design optimization related to heterogeneous design variables and constraints. We propose the use of a partial-least-squares dimension reduction algorithm to address design space heterogeneity, and a *meta* data surrogate selection method to address constraint heterogeneity. Numerical benchmark problems and an aircraft conceptual design optimization problem are used to demonstrate the proposed methods. Results show significant improvement in convergence in early optimization iterations compared to standard Bayesian optimization, with improved prediction accuracy for both objective and constraint surrogate models.

Acknowledgements: The authors would like to thank Jasveer Singh from Bombardier for his insights and support in setting up the industrial aircraft conceptual design problems.

1 Introduction

In the context of aircraft design, Bayesian optimization is a promising and growing field as presented in [4, 20, 24, 36, 39, 41]. However, Bayesian optimization suffers from slow local convergence in early iterations of some applications [2, 36]. One of the reasons for this slow convergence can be attributed to the so-called *cold start* problem where each new Bayesian optimization is initiated without assuming any prior knowledge of the blackbox functions of the objective and constraints.

Transfer learning can be beneficial for addressing the cold start issue in Bayesian optimization that contributes to slow convergence [2, 33, 48, 55]. Transfer learning in Bayesian optimization has been investigated and applied in the field of machine learning as surveyed by Bai *et al.* [2] where the authors break down transfer learning in Bayesian optimization into four different components: (1) surrogate model design, (2) acquisition function definition, (3) initial design of experiments (DOE) strategy, and (4) design space selection. In this work, we focus on using transfer learning in the surrogate model design component of Bayesian optimization to address inaccurate surrogate model predictions in early iterations that slow down convergence considerably.

Most Bayesian optimization applications use Gaussian processes (or variations thereof) as the surrogate model for the objective and constraint functions. In machine learning applications, Gaussian processes with a large number of design variables face computational and accuracy challenges. To that end, other surrogate modeling methods have been proposed, e.g., neural networks [44], deep Gaussian processes [50], neural processes [47], and prior fitted networks [32]. Such methods are not ideal for the types of problems considered here, where the number of design variables is less than 100 and the amount of data is relatively low. We will therefore focus on the use of Gaussian process surrogate models based on the methods presented in [45].

The use of transfer learning on surrogate models of Bayesian optimization can be based on a single surrogate model that uses all the source and target data at once such as the approach proposed by [28, 53], or it can be based on a combination of surrogate models separated by source and target tasks as in [14, 52]. Yogotama and Mann proposed a sequential model-based optimization that transfers information by constructing a common response surface for all source and target datasets [53]. Feurer *et al.* developed an ensemble model for Bayesian optimization that can incorporate the results of past optimization runs using so-called ranking-weighted Gaussian process ensembles (RGPEs) [13]. They first train the Gaussian process posterior for every source task, and use the posterior mean function to compute the number of so-called misranked pairs between one source problem and the target problem. Wistuba *et al.* proposed a two-stage transfer surrogate model (TST-R) for hyperparameter optimization of machine learning models [51]. The first stage of the surrogate model approximates the hyperparameter response functions of a new data set and each data set from the source data individually with Gaussian processes. The second stage combines the first-stage models by taking into account the similarity between the new target data and the source data from previous experiments. They construct a ranking of hyperparameter configurations as well as a prediction about the uncertainty of this ranking. They use the models from source problems to obtain the mean of a target problem. Another approach is the use of *meta* data that describe properties of the data used for the source and target problems [8, 38]. Chen and Liu applied a weighted ensemble of Gaussian processes within Bayesian optimization (similar to the one in [14]) on a one-dimensional turbomachinery problem [10].

Other research uses transfer learning to define a prior distribution of a Gaussian process using source data and then apply the prior to a target problem as proposed in [46] and [18]. More recently, Mahboubi *et al.* proposed a point-by-point transfer learning method that integrates mixtures of Gaussian experts with Bayesian optimization to address the cold start problem, demonstrating improvements on both synthetic and real-world datasets [29]. Chen *et al.* provided a comprehensive review of ensemble of surrogates methods in blackbox engineering optimization, highlighting their effectiveness in improving surrogate robustness but noting that most ensemble methods focus on combining different surrogate types rather than transferring knowledge across problems [9]. In the area of

constrained Bayesian optimization without transfer learning, Yu proposed a constraint-handling framework leveraging Prior-data Fitted Networks, a foundation transformer model that evaluates objectives and constraints simultaneously through in-context learning, demonstrating significant speedups over Gaussian process-based methods [54].

Notably, existing transfer learning methods for Bayesian optimization focus exclusively on unconstrained problems or transfer only the objective function surrogate. None of the aforementioned approaches address the transfer of constraint surrogate models, which is critical for engineering design optimization where constraints often outnumber objectives and significantly impact convergence. Furthermore, these methods assume homogeneous design spaces between source and target problems, limiting their applicability to real-world scenarios where design variables and constraints may differ across optimization tasks. This work addresses these gaps by extending transfer learning to constrained Bayesian optimization with explicit handling of heterogeneous design spaces and constraints.

This paper is organized as follows. In [Section 2](#), we identify a set of motivating scenarios faced by aircraft designers when conducting aircraft design Bayesian optimization, and we present an aircraft conceptual design problem. We provide background on Bayesian optimization and transfer learning of surrogates in [Section 3](#). [Section 4](#) presents the method we propose for using transfer learning in Bayesian optimization, including the ensemble of surrogates method and novel approaches for handling heterogeneity. Numerical and aircraft design problem results are presented in [Section 5](#), where we compare the optimization results of the proposed transfer learning algorithm with existing Bayesian optimization methods. Conclusions are drawn and perspectives are offered in [Section 6](#).

2 Aircraft design and transfer learning

2.1 Motivation and context

Aircraft design Bayesian optimizations are initiated with a cold start, where designers trade initial design of experiments (DOE) size (and amount of corresponding evaluations) with optimization convergence. A designer formulates an optimization problem by defining the objective function, constraint functions, and design variables. The typical Bayesian optimization (see [Algorithm 1](#) in [Section 3](#)) conducts a DOE, builds an initial surrogate model, and initiates the optimization loop. In scenarios where the optimization problem shares the same design variables, constraints, and objective function with an existing DOE, a designer may use that existing DOE to build the initial surrogate models and skip the cold start DOE step; this is the so-called *hot start* Bayesian optimization [5]. However, in most practical applications, designers would not repeat the same optimization problem with the same design space, output and their ranges. Therefore, such a hot start approach is typically not possible. For aircraft conceptual design optimization, we identify the following scenarios where a source optimization or DOE and target optimization differ such that a conventional hot start cannot be used:

- A change in aircraft design requirements such as aircraft range, cruise speed, balanced field length, approach speed, etc. A source optimization or DOE based on a set of design requirements can be different from the target optimization due to differing market requirements of the target optimization, management decisions based on the source optimization or DOE results, or new certification regulations that were not applicable to the source optimization or DOE.
- A change in constraints driven by configuration or technological selections such as landing gear or flight control constraints. A source optimization may have been based on a specific landing gear configuration (e.g., trailing arm type), a specific flight control system configuration (e.g., a slatted wing), or a specific flight control system actuation technology (e.g., linear electrohydraulic servo-actuator); whereas a target optimization may have differing configurations or technologies (a cantilever landing gear type, an unslatted wing configuration, and a rotary electro-mechanical actuator respectively). Each configuration or technology may necessitate its own set

of constraints which leads to a potential differences between the constraints and their limits between the source and target optimizations.

- A change in fixed inputs, e.g., a fixed weight or cabin geometry. A target optimization may require a specific avionics function that is performed by a technology associated with a fixed input weight to the optimization, and this technology (and its associated weights) were not used in the source optimization setup. Similarly, a target optimization may change the cabin geometry relative to the source optimization which leads to several implications on an optimization's output (e.g., fuselage weight or tail moment arm). These different fixed inputs lead to different output values which prevents a designer from using a source optimization data in the target optimization.
- A change of fidelity of models within the multi-disciplinary analysis (MDA). A source optimization may have been performed using a low fidelity aerodynamics model. Whereas a target optimization may use a newly calibrated low-fidelity aerodynamics model or a high-fidelity model. This change of model within the MDA leads to different results even for the same set of inputs between the source and target optimizations.
- A change in simulation codes in the MDA between the source and target optimization. Updates to simulation codes due to bug fixes or changes in capabilities can lead to differing output for the same input.

The motivation of this work is to enable transferring knowledge between source and target optimizations in the presence of changes in requirements, constraints, inputs, or models listed above. The aim is to reduce convergence time of new optimizations by creating transferable surrogate models of disciplines within an optimization. Existing methods for transfer learning in Bayesian optimization do not modify surrogate models prior to transferring knowledge from the source to the target problem. In addition, previous work on the weighting of surrogate ensembles does not separate the design space prior to defining these surrogate ensembles, as explained in [45]. Existing methods for transfer learning using ensembles of Gaussian processes as surrogates do not address the need for accurate constraint models in Bayesian optimization where the value of predictions is just as important as the shape of the constraint functions. Finally, existing methods using ensembles of Gaussian processes as surrogates in Bayesian optimization do not handle heterogeneous design variables or heterogeneous constraints, i.e., when there is a mismatch between the design variables or the constraints output of the source and target problems.

The objective of this work is to develop algorithms to enable faster Bayesian optimization by applying transfer learning techniques while addressing the shortcomings of existing work presented above.

2.2 An aircraft conceptual MDO problem

We present an aircraft conceptual design optimization problem inspired by the examples identified previously. The problem of interest is a single target optimization problem. We consider available existing data based on five source DOEs (which are the results of previous optimizations). The definition of target and source problems is presented in Table 1 showing top views of selected aircraft from the DOE of each problem in Figure 1.

Table 1: Setup of source and target problems in the aircraft conceptual MDO transfer learning problem.

| Problem | Description |
|---------------|--|
| Target (T) | BRAC |
| Source 1 (S1) | BRAC with a modified engine weight (+500 lb) and a relaxed BFL constraint (+10%) |
| Source 2 (S2) | BRAC with a longer fuselage (+100 inch) and longer range (+20%) |
| Source 3 (S3) | BRAC with higher cruise speed requirement (+ Mach 0.05) |
| Source 4 (S4) | BRAC with a different engine architecture and rear spar configuration |
| Source 5 (S5) | BRAC with slats |

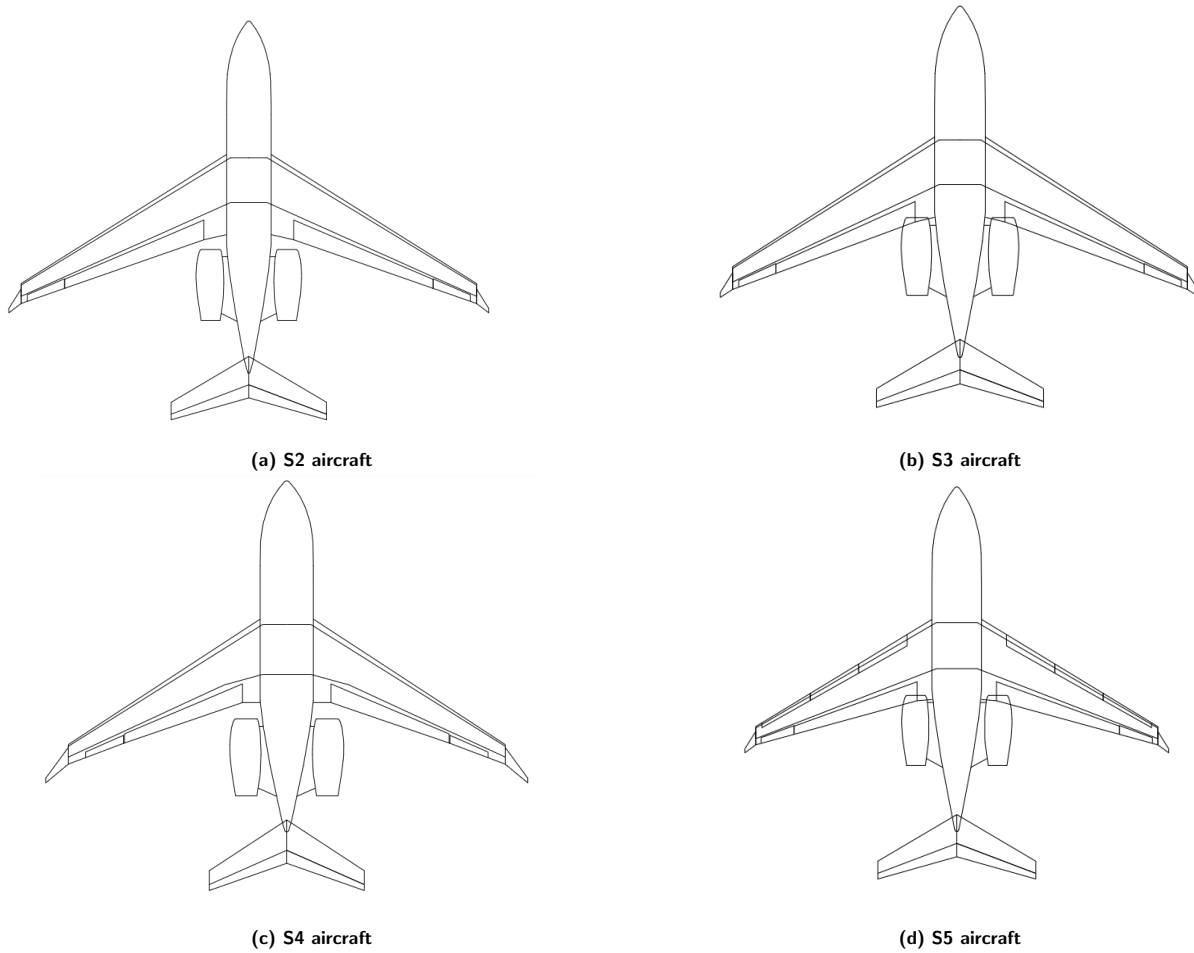


Figure 1: Scaled top views of selected aircraft from the source problem's DOE highlighting the differences in wing and fuselage geometries.

The target and source problems share a homogeneous objective function: minimize the maximum takeoff weight (MTOW) of the Bombardier Research Aircraft (BRAC) [36, 37] using heterogeneous requirements, constraints, and design variables. The purpose is to leverage existing source data to improve the performance of the target optimization.

$$\begin{cases} \text{minimize} & \text{MTOW}(x) \\ & x \in \Omega \\ \text{subject to} & c_i(x) \leq 0, \quad i = 1, \dots, m, \end{cases}$$

where $x \in \Omega \subseteq \mathbb{R}^{16}$ is the vector of bounded design variables listed in Table 2 and $c_i(x)$ $i = 1, \dots, 13$ are the inequality constraints described in Table 3. We use Bombardier's conceptual multidisciplinary optimization (MDO) framework [35] considering tube-and-wing aircraft configurations with a T-tail design and turbofan engines. The MDO framework includes an aircraft multidisciplinary analysis (MDA) environment which is comprised of sizing and simulation models of all major disciplines in aircraft conceptual design. The extended design structure matrix (XDSM) [25] of the MDA environment is presented in Figure 2.

Table 2: Design variables and their applicability (noted as a checkmark ✓) to the target (T) and each source (S1 to S5).

| Input | Description | T | S1 | S2 | S3 | S4 | S5 |
|-------------------------|--------------------------------------|---|----|----------------|----------------|----|----------------|
| x_1 | Rubber engine scaling factor | ✓ | ✓ | ✓ ¹ | ✓ ¹ | ✓ | ✓ |
| x_2 | Rubber engine bypass ratio | | | | | ✓ | |
| x_3 | Rubber engine overall pressure ratio | | | | | ✓ | |
| x_4 | Wing aspect ratio | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_5 | Wing area | ✓ | ✓ | ✓ ¹ | ✓ ¹ | ✓ | ✓ ¹ |
| x_6 | Wing trailing edge sweep | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_7, x_8 | Wing rear spar chord-wise location | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_9 | Wing sweep | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_{10} | Wing taper ratio | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_{11}, \dots, x_{14} | Wing thickness-to-chord ratios | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| x_{15} | Rear spar kink span-wise location | | | | | ✓ | |
| x_{16} | Slat chord ratio | | | | | | ✓ |

¹ using different bounds for the same design variable

Table 3: A list of the aircraft conceptual design problem constraints.

| Constraint | Description | T | S1 | S2 | S3 | S4 | S5 |
|-------------|--|-----|------|------|-----|-----|----|
| $c_1(x)$ | Balanced field length (BFL) | = | +10% | = | = | = | = |
| $c_2(x)$ | Initial cruise altitude (ICA) | = | = | = | = | = | = |
| $c_3(x)$ | Aircraft reference speed (V_{ref}) | = | = | = | = | = | = |
| $c_4(x)$ | Excess fuel weight | = | = | = | = | = | = |
| $c_5(x)$ | Inboard Wing flight controls actuation height clearance | * | * | * | * | * | * |
| $c_6(x)$ | Outboard Wing flight controls actuation height clearance | * | * | * | * | * | * |
| $c_7(x)$ | Inboard Wing flight controls actuation chord clearance | * | * | * | * | * | * |
| $c_8(x)$ | OutboardWing flight controls actuation chord clearance | * | * | * | * | * | * |
| $c_9(x)$ | Wing chord clearance for landing gear integration | * | * | * | * | * | * |
| $c_{10}(x)$ | Wing tip chord | = | = | = | = | = | = |
| $c_{11}(x)$ | Aircraft climb performance | = | = | = | = | = | = |
| $c_{12}(x)$ | Aircraft mission range | = | = | +20% | = | = | = |
| $c_{13}(x)$ | Slat actuator height | N/A | N/A | N/A | N/A | N/A | * |

* Constraints defined using internal simulation models within the MDA

= Equivalent constraint bounds for target and source problems

% Heterogeneous constraint thresholds that vary by a fixed percentage relative to the target problem

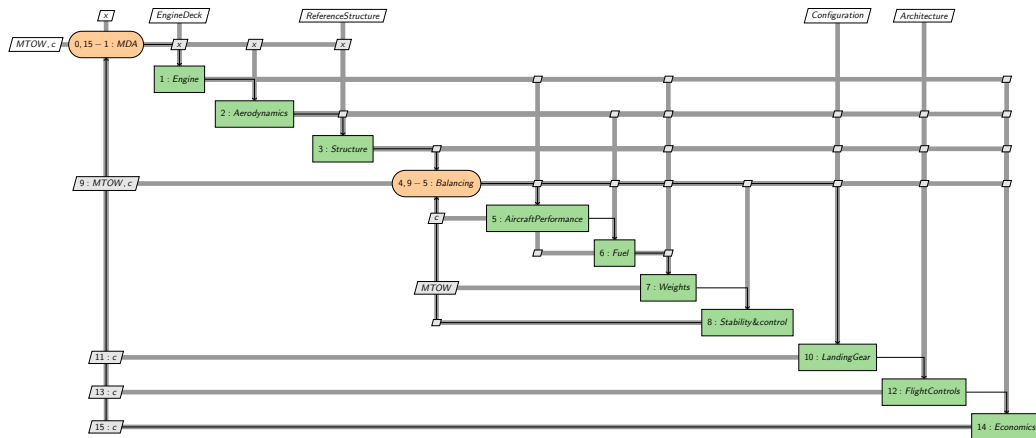


Figure 2: XDSM representation of the aircraft conceptual design MDA framework.

Note that the reference engine weight and fuselage length are fixed inputs to the MDA; therefore, they do not appear in the list of design variables in Table 2. In addition, the cruise speed requirement is set in the MDA and the aircraft performance simulation model analyzes its impact on the mission, i.e., cruise speed is not a constraint.

3 Bayesian optimization and transfer learning

Bayesian optimization is thought of having its origins in Efficient Global Optimization (EGO) [21]. It is a surrogate-assisted optimization approach that has gained significant attention in the last two decades. It targets problems with expensive or noisy blackbox functions, promising reduced computational cost and efficient convergence using a probabilistic approach. Consider the unconstrained optimization problem

$$\min_{x \in \Omega} y(x), \quad (1)$$

where $y(x)$ is a noisy blackbox function and x is its input vector within a design space $\Omega \subseteq \mathbb{R}^D$. In Bayesian optimization, a surrogate $\hat{y}(x)$ is created based on the input and output of the function $y(x)$ and is used in a sequential optimization framework. There are three components in Bayesian optimization: (1) an initial DOE, (2) a surrogate model, and (3) an acquisition function that is used to select the points to be evaluated using the blackbox function $y(x)$.

3.1 Gaussian process surrogate models

In Bayesian optimization, the most popular surrogate models are Gaussian processes [49], also known as Kriging [23]. Let us assume a DOE with N samples evaluated using a blackbox function $y(x)$ and denote its input vectors of dimensions d as $X_N = \{x_n\}_{n=1}^N$ and its output vector as $Y_N = \{y_n\}_{n=1}^N$ with $x_n \in \mathbb{R}^D$ and $y_n \in \mathbb{R}$, $\forall n \in N$. A Gaussian process is a stochastic process that approximates $y(x)$ at any point x by a Gaussian random variable $\hat{Y}(x)$ which is defined by its mean and covariance functions, $\hat{y}(x)$ and $k(\cdot, \cdot)$ respectively. The mean function is defined as

$$\hat{y}(x) = \hat{\beta} + r^T(x)K^{-1}(Y_N - \mathbf{1}\hat{\beta}) \quad (2)$$

and the variance is defined as

$$\hat{s}^2(x) = \hat{\sigma}^2(x)(1 - r^T(x)K^{-1}r(x)), \quad (3)$$

where $\hat{\beta}$ is a mean bias, $r(x) = \{k(x, x_1), \dots, k(x, x_N)\}$, K is an $N \times N$ covariance matrix of the combination of components of the DOE $K_{ij} = k(x_i, x_j)$, $\mathbf{1}$ is an $N \times 1$ column vector of 1's, and $\hat{\sigma}$ is the variance of the covariance function. In this work, we use the squared exponential covariance function:

$$k(x_i, x_j) = \hat{\sigma}^2 \prod_{d=1}^D \exp\left(-\frac{(x_i^d - x_j^d)^2}{2l_d}\right), \quad (4)$$

where x_i and $x_j \in \mathbb{R}^D$, the length scale vector $l = \{l_1, \dots, l_D\} \in \mathbb{R}$, the variance $\hat{\sigma}$, and the mean bias $\hat{\beta}$ are obtained by the training data of the DOE. The most common method to obtain these hyperparameters is the maximum likelihood estimation (log-ML) [15]:

$$\log\text{-ML} = -\frac{1}{2} \left(N \ln(2\pi\hat{\sigma}^2) + \ln(\det K) + \frac{(Y_N - \mathbf{1}\hat{\beta})^T K^{-1}(Y_N - \mathbf{1}\hat{\beta})}{\hat{\sigma}^2} \right). \quad (5)$$

3.2 Acquisition functions

The acquisition function is a critical part of the Bayesian optimization framework. Many acquisition functions have been proposed and investigated in literature. Expected Improvement (EI) [21] is one of the most widely used:

$$\text{EI}(x) = \int_{-\infty}^{\infty} I(x)\phi(z) dz, \quad (6)$$

where $I(x)$ is the improvement at point x relative to the the minimum value of the objective function observed so far $y_{\min} - \hat{y}(x)$, y_{\min} is the minimum value of the objective function observed so far, $\hat{y}(x)$ and $\hat{s}(x)$ are mean and standard deviation of the Gaussian process, and ϕ is the probability distribution function, and $z = (y_{\min} - \hat{y}(x))/\hat{s}(x)$.

3.3 Constraint handling

We consider an inequality constrained optimization formulation. We use the Bayesian optimization extension to handle constrained blackbox optimization problems as proposed by [43] where they propose creating surrogate models of the inequality constraints $c_1(x), \dots, c_m(x)$ similar to the objective function and apply these constraints on the acquisition function maximization problem. We define the constrained optimization problem as

$$\begin{aligned} \min_{x \in \Omega} \quad & y(x) \\ \text{subject to} \quad & c_i(x) \leq 0, \quad i = 1, \dots, m. \end{aligned} \quad (7)$$

The BO problem is then defined as

$$\begin{aligned} \max_{x \in \Omega} \quad & \alpha(x) \\ \text{subject to} \quad & \hat{c}_i(x) \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (8)$$

where $\alpha(x)$ is the selected acquisition function, and $\hat{c}_1(x), \dots, \hat{c}_m(x)$ are surrogate models of the inequality constraints. The overall algorithm for a constrained Bayesian optimization framework is defined in Algorithm 1.

Algorithm 1: Constrained Bayesian optimization framework.

Input: Initial DOE
Output: Best feasible point from the DOE

- 1 **for** $i = 1, \dots, \text{max_iter}$ **do**
- 2 Build Gaussian process surrogate models of the objective and constraint functions.;
- 3 Maximize an acquisition function to find x^{i+1} .;
- 4 Evaluate objective and constraint functions at x^{i+1} .;
- 5 Update the DOE.;
- 6 **end**

3.4 Ensemble of surrogates using transfer learning

We use an ensemble of Gaussian process surrogate models in lieu of the conventional Gaussian processes used in Bayesian optimization based on the work presented in [45]. The method starts by transferring the posterior of source models to target models as follows. Then, surrogate models are weighted based on multiple criteria according to the intended function.

3.4.1 Weighted ensemble of surrogates

A source model $(\hat{y}_S^n, \hat{s}_S^n)$ is adjusted by considering both a scaling factor and a bias, as presented in [45]

$$\hat{y}_{S'}^n(x) = \alpha_n \hat{y}_S^n(x) + \beta_n, \quad (9)$$

where the scaling term α_n and the bias term β_n are constant parameters that can be fitted to the target data DOE_T to create a modified set of source models using a prespecified error metric ε (e.g., mean squared error). The corresponding adjusted variance is then calculated as follows

$$\hat{s}_{S'}^n(x) = \alpha_n \hat{s}_S^n(x). \quad (10)$$

All source models go through this adjustment step as presented in Algorithm 2. The method then combines the modified source surrogate models into an ensemble using a weighted combination approach. The ensemble prediction \hat{y} at a point x is given by:

$$\hat{y}(x) = \sum_{j=1}^N P(\hat{y}_j = \hat{y}^*) \hat{y}_j(x), \quad (11)$$

Algorithm 2: Source surrogate model with posterior information [45, Algorithm 1].

Input: N Source datasets $\text{DOE}_S^{1,\dots,N}$ and target dataset $\text{DOE}_T : (X_T, Y_T)$.
Output: Modified source GP models $(\hat{y}_{S'}^{1,\dots,N}, \hat{s}_{S'}^{1,\dots,N})$.

- 1 **for** $n = 1, \dots, N$ **do**
- 2 Build source GP model $(\hat{y}_S^n, \hat{s}_S^n)$ based on DOE_S^n using Equations (2) to (5);
- 3 Use the posterior of the \hat{y}_S^n to calculate $\hat{y}_{S'}^n(X_T)$;
- 4 Calculate $(\alpha_n, \beta_n) = \underset{\alpha, \beta \in \mathbb{R}}{\text{argmin}} \varepsilon(\alpha, \beta)$ where ε is an error metric between $\hat{y}_{S'}(X_T) = \alpha \hat{y}_S^n(X_T) + \beta$ and Y_T
 for a given $\alpha, \beta \in \mathbb{R}$;
- 5 Compute the hyperparameters σ_0^n and l^n ;
- 6 Set the modified source GP model: $\hat{y}_{S'}^n = \alpha_n \hat{y}_S^n + \beta_n$ and variance $\hat{s}_{S'}^n = \alpha_n \hat{s}_S^n$;
- 7 **end**

where N is the number of source models, \hat{y}^* represents the true underlying function, and $P(\hat{y}_j = \hat{y}^*)$ is the probability that source model j is the correct model. The computation of these probabilities is described in Section 3.4.2. This formulation was written for a mixture of experts approach with multiple clusters in [45] where the design space is divided into K clusters and model probabilities are computed per cluster; however, in this work we use a single cluster ($K = 1$) to minimize computational costs.

To obtain the standard deviation \hat{s} of the ensemble of surrogates \hat{y} , we present two methods:

1. Target surrogate model variance \hat{s}_t calculated using Equation (3):

$$\hat{s}^2(x) = \hat{s}_t^2(x). \quad (12)$$

2. Weighted ensemble of surrogates variances:

$$\hat{s}^2(x) = \sum_{j=1}^N (P(\hat{y}_j = \hat{y}^*))^2 \hat{s}_j^2(x), \quad (13)$$

where $\hat{s}_j(x)$ is the standard deviation of the j th source model.

3.4.2 Multi-criteria weighted ensembles

To compute the model probabilities $P(\hat{y}_j = \hat{y}^*)$ in Equation (11), a multi-criteria ranking approach was used. The selectable criteria are based on the shape, accuracy, and variance of the surrogates in an ensemble. A summation of weighted multi-criteria $C(\hat{y}_j = \hat{y}^*)$ to calculate a score for a model \hat{y}_j that being the correct model is presented as:

$$C(\hat{y}_j = \hat{y}^*) = \sum_{l=1}^{N_C} w_l c_l(\hat{y}_j = \hat{y}^*), \quad (14)$$

where $w_l \in [0, 1]$ for $l = 1, \dots, N_C$ are the weights assigned to the selected criteria such that $\sum_{l=1}^{N_C} w_l = 1$, c_l is the measure of the selected criterion, and N_C is the number of measured criteria.

For the shape criterion, the discordant pairs method is used from [14, 45]. The Epanechnikov quadratic kernel [11] is used as follows:

$$\kappa_\rho(\chi_S, \text{DOE}_T) = \delta \left(\frac{\tau(\chi_S, \text{DOE}_T)}{\rho} \right) \quad (15)$$

with

$$\delta(t) = \begin{cases} \frac{3}{4}(1-t^2) & \text{if } t \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

and

$$\tau(\chi_S, \text{DOE}_T) = \frac{\sum_{m=1}^n \sum_{k=2}^n \mathbf{1}((y_m^s < y_k^s) \oplus (y_m^t < y_k^t))}{n}, \quad (17)$$

where $\rho > 0$ is a predefined bandwidth, DOE_T is the input-output data of the target problem T , and χ_S are the data predictions Y_S on the inputs X_T of DOE_T using a source model S with n points.

For the accuracy criterion, a measure of the absolute relative error metric is used:

$$\tau_a(\chi_S, \text{DOE}_T) = \frac{1}{n} \sum_{m=1}^n \mathbf{1} \left(\frac{|y_m^s - y_m^t|}{|y_m^t|} > \epsilon_{\max} \right), \quad (18)$$

where ϵ_{\max} is a predefined maximum relative error. Similarly, for the variance criterion:

$$\tau_v(\chi_S, \text{DOE}_T) = \frac{1}{n} \sum_{m=1}^n \mathbf{1} \left(\frac{\sigma_m^s}{y_{\max}^t} > \sigma_{\max} \right), \quad (19)$$

where σ_{\max} is a predefined maximum relative variance, y_{\max}^t is the maximum observed absolute output value in DOE_T , and σ_m^s is the variance of point m measured using source model S .

The probability of a source model j being the correct model is then estimated as:

$$P(\hat{y}_j = \hat{y}^*) \approx \frac{C(\hat{y}_j = \hat{y}^*)}{\sum_{n=1}^N C(\hat{y}_n = \hat{y}^*)} \quad (20)$$

where N is the number of source models.

The choice of criteria weights depends on the role of the surrogate model in the optimization. For the objective function surrogate, shape similarity is prioritized because the acquisition function relies on the relative ranking of predictions to identify promising regions; accurate absolute values are less critical since the optimizer seeks improvement directions rather than exact function values. For constraint surrogates, accuracy becomes equally important because constraint satisfaction depends on the actual predicted values relative to constraint bounds—a surrogate that correctly ranks points but systematically over- or under-predicts may misclassify feasibility. The variance criterion is included in both cases to favor source models with lower uncertainty, improving confidence in the ensemble predictions.

3.4.3 Ensemble of surrogates algorithm

The ensemble of surrogates using the transfer learning algorithm can be summarized in Algorithm 3. This algorithm is a simplified instance of [45, Algorithm 3] where only one cluster is regarded.

Algorithm 3: Ensemble of surrogate models using transfer learning.

Input: Source datasets $\text{DOE}_S^{1, \dots, N}$, target dataset DOE_T , selected criteria N_C , and criteria weights w_l and bandwidths ρ_l for $l \in 1, \dots, N_C$

Output: Ensemble of surrogates using transfer learning

- 1 Build target surrogate model \hat{y}_t using DOE_T ;
 - 2 **for** $n = 1, \dots, N$ **do**
 - 3 Build n th source surrogate model \hat{y}_{s_n} using DOE_S^n ;
 - 4 Use the posterior of \hat{y}_{s_n} and DOE_T to create a modified source surrogate model $\hat{y}_{s'_n}$ using Algorithm 2;
 - 5 Calculate the criteria scores $C(\hat{y}_{s'_n} = \hat{y}^*)$ using Equation (14);
 - 6 Calculate the probability $P(\hat{y}_{s'_n} = \hat{y}^*)$ using Equation (20);
 - 7 **end**
 - 8 Return the ensemble of surrogates model (\hat{y}, \hat{s}) using Equations (11) and (12) or (13);
-

4 Methodology

In this section, we describe the algorithms and methods we propose for heterogeneous transfer learning in Bayesian optimization. Section 4.1 proposes an approach to handle heterogeneous design spaces using partial-least-squares dimension reduction. Section 4.2 presents a method to handle heterogeneous constraints based on meta data, which are features that describe the design variables and constraints, e.g., variable names. In Section 3.4, we present the ensemble of surrogates using transfer learning approach, including the transfer of prior method, weighted ensemble formulation, and multi-criteria weighting scheme. Section 4.3 presents the overall algorithm to perform transfer learning in constrained Bayesian optimization.

4.1 Heterogeneous design space

Heterogeneity in transfer learning is encountered when the design space or the outputs of the source model(s) do not match the design space or outputs of the target model [3, 22, 48, 55]. In real-world scenarios, the use of existing data typically comes with a variation of the design variables used between source and target problems. For example, in aircraft design settings source data may contain an engine by-pass ratio as a design variable whereas a target problem may have a fixed engine size and architecture. In this case, the target problem still has the engine by-pass ratio as a fixed parameter that can be used to solve the heterogeneity issue. However, in certain cases, source data may include design variables that do not even exist as fixed parameters in the target problem. For example, source data may be obtained from an aircraft that has slats with a design variable for the ratio of slats chord to the wing chord. Assuming that the target problem is an unslatted aircraft, this design variable does not exist and therefore cannot be used as a fixed parameter. In [3], Bao *et al.* surveyed methods that handle heterogeneity in transfer learning problems and categorize the types of algorithms as shown in Figure 3. They distinguished between data-based and model-based methods. Data-based methods involve the transfer of either the original data or their transformed design variables to a target design space, allowing the target models to be trained with this augmented data, thereby enriching the available data within the target design space.

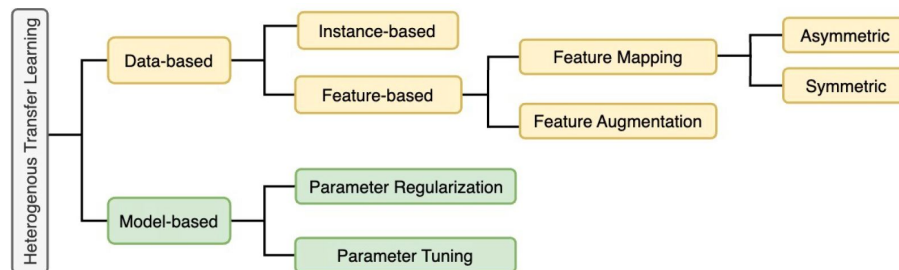


Figure 3: Heterogeneous transfer learning categories (adapted from [3]).

Conversely, model-based methods center around constructing models within the source design space then applying target data to adapt model results to the target design space. In literature, heterogeneous transfer learning in Bayesian optimization is presented in [31] and [12]. In [31], Min *et al.* proposed a model-based approach that builds a neural network to estimate design variables mapping from a source problem to a target problem using a neural network. In [12], Fan, Han, and Wang proposed a neural network model that estimates a prior of the target model based on the source models and design variables. The neural network is trained using source design variables and posteriors of source surrogate models. In this work, the use of data-based methods is more appropriate for ensembles of surrogates since we are interested in maintaining each surrogate model as a separate entity.

Assuming a set of design variables of a source problem X_S and a set of design variables of a target problem X_T , Bao *et al.* categorized the differences between the types of data-based transfer

learning methods, i.e., instance or feature based, in Figure 4 [3]. Instance-based methods are best suited for situations where X_S and X_T are weakly related. An intermediate set of design variables is defined to build both the source and target surrogate models. The concept of feature-based methods is to increase the similarity between the source and target design variables. A common measure of similarity is the so-called Maximum Mean Discrepancy (MMD) that assesses the distances between the means of distributions between the source and target models [17]. Feature mapping methods entail modifying the set of design variables of the source and target problems into a common design space, so-called symmetric, or projected design space on the source or target problems, so-called asymmetric. Feature augmentation methods add dimensions d to the dimensions of the source and target problems, d_S and d_T respectively [26]. In this work, we propose the use of a partial-least-squared dimension

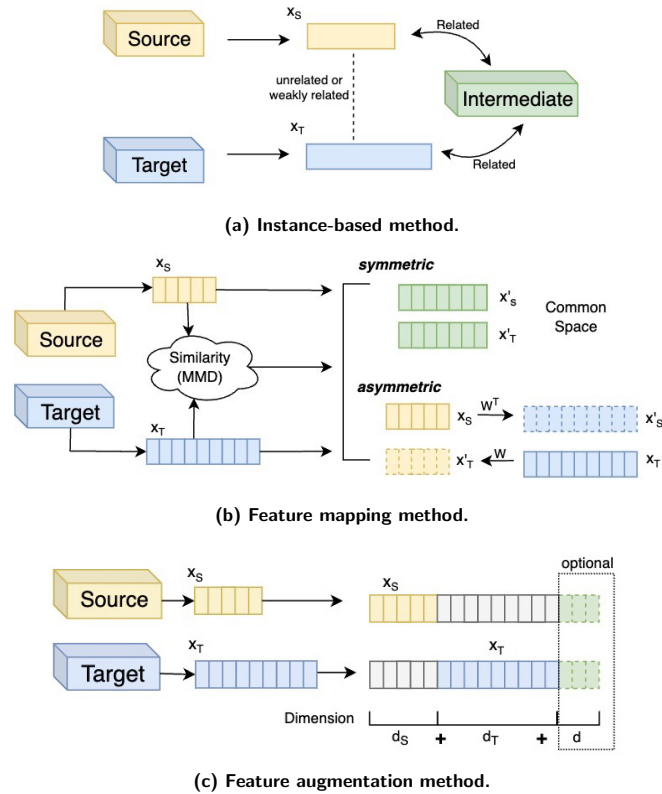


Figure 4: Heterogeneous data-based transfer learning methods (adapted from [3]).

reduction algorithm to address heterogeneous design spaces between source and target problems similar to feature mapping methods in [3] since the source and target problems will share a number of design variables in our aircraft optimization settings as we describe in Table 1. In addition, we are interested in reducing the computational cost of the surrogate models in Bayesian optimization. Bouhleb *et al.* proposed a method for dimension reduction of Gaussian processes using partial least squares [7]. They developed a kernel that reduces the number of parameters of a Gaussian process. The key assumption is that some input variables have a less significant impact on the output of a Gaussian process than others. Therefore, they proposed to project the design space parameters onto smaller dimensions while monitoring the correlation between input variables and the output. The partial least squares method builds a relationship between these input variables and the projected variables using linear combinations of the input variables. An input vector X of d dimensions is used to create t principal components with h dimensions to predict the output of a function y . These principal components represent the new coordinate system and are defined for $l = 1, \dots, h$ as

$$t^l = Xw_*^{(l)}, \quad (21)$$

where $w_*^{(l)}$ for $l = 1, \dots, h$ represent the weights of the input vector X on principal component t . The vector $w_*^{(l)}$ is obtained by maximizing the covariance of $X^{(l-1)t} y^{(l-1)} y^{(l-1)t} X^{(l-1)}$. The number h of principal components is chosen by minimizing the leave-one-out cross-validation error. Consequently, a point x is transformed into principal component t as follows

$$x \mapsto [w_{*1}^{(l)} x_1, \dots, w_{*d}^{(l)} x_d]^t. \tag{22}$$

Finally, the authors obtain the following Gaussian exponential kernel between points x and x' :

$$k(x, x') = \sigma^2 \prod_{l=1}^h \prod_{i=1}^d [-\theta_l (w_{*i}^{(l)} x_i - w_{*i}^{(l)} x'_i)^2], \tag{23}$$

where σ^2 and $\theta_l \in [0, +\infty[$ are the covariance and hyper-parameters of the Gaussian, respectively. Bouhlel *et al.* applied their method on numerical problems obtaining significant computational gain while maintaining prediction accuracy compared to that of conventional Gaussian processes.

In the aircraft design optimization problems considered here, source problems share a number of design variables with the target problem. Input and output variables are also defined in advance. Therefore, we can assign meta data that define each input variable as presented in Figure 5, where a transfer learning problem with two source input vectors and a target input vector is considered. Knowing these meta data, we propose to modify the KPLS algorithm in [7] by adding or removing variables to the modified input vectors of source problems to match the input variables of the target problem. Consequently, when performing the optimization to obtain the weights vector of the principal components $w_*^{(l)}$ for $l = 1, \dots, h$, we assign a NULL value for $w_*^{(l)}$ of each added or removed input variable of the source problems. The illustration in Figure 5 shows that input vectors X_{S1}, X_{S2}, X_T are assigned defining meta data (a,b,c,...). Therefore, when creating the KPLS components, variables with meta data g and h, $w_g^{(l)}$ and $w_h^{(l)}$ are assigned NULL values when creating the modified input vector X'_{S1} of source problem 1 for all $l = 1, \dots, h$. Similarly, $w_e^{(l)}$ and $w_f^{(l)}$ are assigned NULL values when creating the modified input vector X'_{S2} of source problem 2. The resulting principal components t_{S1}, t_{S2}, t_T will depend on the relationships between each modified input vector and the corresponding output. We note that for this approach, the target problem does not need to apply KPLS dimension reduction for predictions as opposed to the source problems.

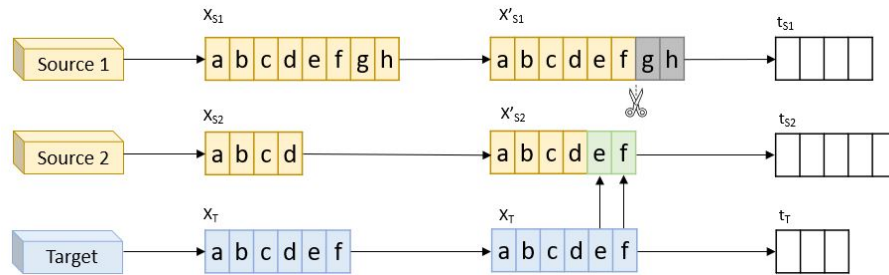


Figure 5: Heterogeneous design space in transfer learning using meta data.

Another approach to solve the heterogeneous design space problem in transfer learning of Bayesian optimization is the use of Hierarchical Gaussian Processes [42]. The idea is to combine all source and target data into a single Gaussian process and to add a set of so-called dimensional variables that affect the dimensions of a problem and decide if other so-called decreed variables, i.e. conditionally active or inactive variables, are acting or non-acting [1]. We determine that the use of hierarchical Gaussian processes is not ideal for solving the heterogeneity problem of transfer learning in Bayesian optimization since the need to add all source and target data points into a single Gaussian process significantly increases computational training needs at every iteration compared to the ensemble of surrogates approach where multiple source Gaussian processes are prebuilt and only the target Gaussian process is reconstructed at every iteration.

4.2 Heterogeneous constraints

In [Section 2](#), we considered scenarios where constraints of source problems are not equivalent to target problem constraints. In this section, we present an approach to handle heterogeneous optimization constraints for transfer learning. Let's first consider a target problem with constraints vector C_T and two source problems with constraints vectors C_{S1} and C_{S2} as shown in [Figure 6](#). One could use the ensemble of surrogates approach defined for the objective function using all the constraints data from both of the source problems to define each of the target constraint surrogates. However, for scenarios where there is a large number of constraints and a large number of source problems, such an approach may not be practical due to the additional computational cost needed to create ensembles for each target problem constraint. Therefore, we propose the use of meta data to select the source constraints surrogate models to be used for the ensemble of surrogates for a target constraint.

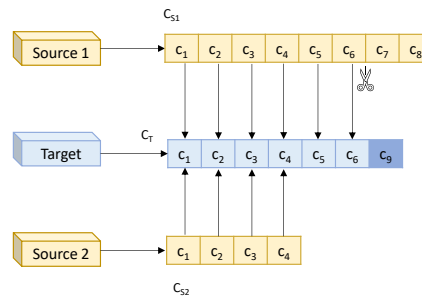


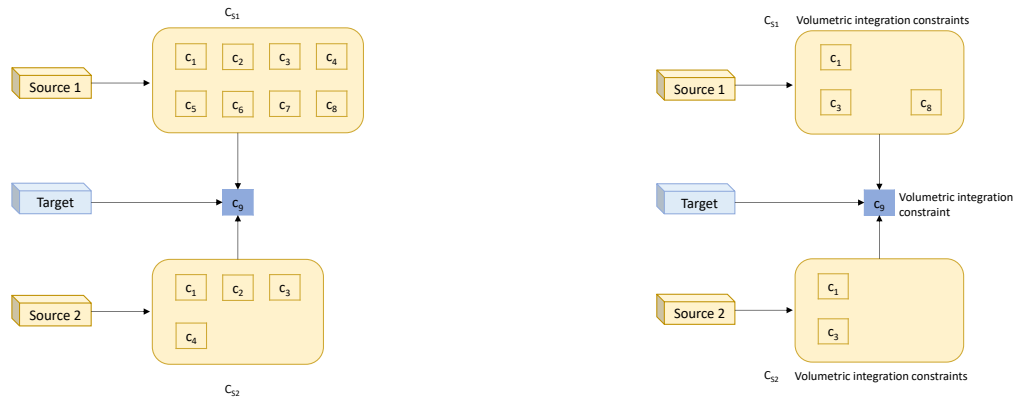
Figure 6: Heterogeneous constraints transfer of surrogates approach showing the matching of source and target constraints using meta data.

Assuming there exist meta data to describe each constraint, e.g., parameter name, we simply propose to match the same parameters from the source problems when creating the target problem constraint ensemble of surrogates. Therefore, in the schematic shown in [Figure 6](#), target constraints $c_1 - c_4$ are built using source constraint models from both C_{S1} and C_{S2} whereas constraints c_5 and c_6 are built solely based on source constraint models from C_{S1} .

We also present a case where the target constraint does not exist in any of the source constraint data in [Figure 6](#), i.e., c_9 does not have a matching parameter in C_{S1} nor C_{S2} . In such a case, we can use all the source problems constraints for the ensemble of surrogates of c_9 as presented in [Figure 7\(a\)](#) which may be computationally costly as explained above. Alternatively, we propose creating categories for constraints to reduce the number of models generated for the ensemble, as presented in [Figure 7\(b\)](#). For aircraft conceptual design problems, we categorize optimization constraints by discipline, as per the examples presented in [Table 4](#). Let's assume that c_9 is a volumetric integration constraint in aircraft conceptual design, therefore, only parameters under the same category can be used to generate the ensemble of surrogates for c_9 as we present in [Figure 7\(b\)](#).

4.3 Transfer learning for constrained Bayesian optimization

The overall algorithm for a constrained Bayesian optimization framework is defined in [Algorithm 4](#). Variations of this algorithm can be created to cater for the problems at hand. For example, building the probabilities for the ensemble of surrogates can be created prior to the Bayesian optimization loop to reduce the computational cost of the algorithm (or fixing these probabilities after a set number of iterations). In this work, we aim to maximize the use of target data, so we elect to re-calculate the model probabilities in the ensemble of surrogates after every new evaluation point of the target functions (objective and constraints).



(a) Use of all source constraint models for the ensemble of surrogates.

(b) Use of subset of source constraint models for the ensemble of surrogates.

Figure 7: Target problem constraints without matching meta data in source problems.

Table 4: Aircraft MDO constraints.

| Constraints type | Constraints description |
|------------------------|--|
| Performance | Minimum range at different speeds and payloads |
| | Maximum takeoff (or balanced) field length |
| | Maximum landing field length |
| | Minimum initial cruise altitude |
| | Maximum approach (or reference) speed |
| Volumetric integration | Maximum climb duration |
| | Minimum flight control actuation space envelope |
| | Minimum landing gear integration space envelope |
| | Minimum empty fuselage volume (outside of payload volume) |
| | Minimum fuselage underfloor height for avionics integration |
| | Maximum wing span |
| Operational | Minimum excess fuel volume after design mission |
| | Maximum external noise in different flight phases (flyover, approach, takeoff) |
| Environmental | Maximum operating cost |
| | Ability of engines to provide power for anti-icing |
| | Emissions , noise, and life cycle analysis |

Algorithm 4: Constrained Bayesian optimization with transfer learning.**Input:** Source datasets $DOE_S^{1,\dots,N}$ and initial target dataset DOE_T **Output:** Best feasible point from the target DOE

- 1 Build source Gaussian processes for the objective and constraint functions $\hat{y}_s^{1,\dots,N}$ using $DOE_S^{1,\dots,N}$
- 2 **for** $i = 1, \dots, max_iter$ **do**
- 3 Build a transfer learning ensemble of Gaussian processes using [Algorithm 3](#) for the objective function and each constraint;
- 4 Maximize an acquisition function to find x^{i+1} . ;
- 5 Evaluate objective and constraint functions of the target model at x^{i+1} .;
- 6 Update DOE_T .;
- 7 **end**

5 Numerical results

We present analytical and aircraft conceptual design optimization examples to demonstrate the proposed methods.

5.1 Implementation details

The code is developed in Python 3 building on the following tool boxes: Scikit-learn v1.5.2 [34], SMT 2.0 [40], SMAC3 [27], RGPE [14], TST-R [51], and the Bayesian optimization library in [16]. All results are obtained using an Intel® Xeon® CPU E5-1650 v3 @ 3.50 GHz core with 32 GB of memory.

In the analytical and aircraft optimization examples in the following sections, we use the following notation:

- a) Optimization algorithms:
 - VBO: a vanilla Bayesian optimization per [Algorithm 1](#).
 - TLBO: Bayesian optimization with transfer learning per [Algorithm 4](#).
- b) Surrogate models within TLBO:
 - ETL: ensemble of surrogates using transfer learning per [Algorithm 3](#).
- c) Variance estimation for ensemble methods:
 - TV: Target variance using from [Equation \(12\)](#).
 - AV: Weighted variance based on ensemble of surrogates without clustering using [Equation \(13\)](#).

Combinations of the above methods are concatenated when presented. For example, TLBO optimizer with ETL and AV is presented as TLBO-ETL-AV.

We use a single cluster assumption for creating the ensembles of surrogates in ETL to minimize its impact on the computational costs of optimizations. We set fixed weights for the ensemble criteria from [Section 3.4.2](#). For the objective function, we are interested in the shape of the source functions and how they compare to the target function. Therefore, we set the criteria weights to be based on shape and variance. Whereas for the constraint functions, accuracy is just as important as the shape of the function as it determines feasibility in the acquisition function maximization loop as described in [Section 3.3](#). Therefore, we assign equal weights to each of the shape, accuracy, and variance criteria.

5.2 Analytical examples

We first illustrate an unconstrained optimization problem based on the Bohachevsky functions [6, 19]. We use three homogeneous source functions, f_{S1} , f_{S2} , f_{S3} , and one target function, f_T , as follows. For a given $x_1, x_2 \in [-5, 5]$, we consider

$$f_{S1}(x_1, x_2) = x_1^2 + 2x_2^2 - 0.3 \cos(\pi x_1) - 0.4 \cos(2\pi x_2) + 0.7 \quad (24)$$

$$f_{S2}(x_1, x_2) = x_1^2 + 2x_2^2 - 0.3 \cos(\pi x_1) \cos(2\pi x_2) \quad (25)$$

$$f_{S3}(x_1, x_2) = 2x_1^2 + 4x_2^2 - 0.3 \cos(3\pi x_1 + 4\pi x_2) - 0.5 \quad (26)$$

$$f_T(x_1, x_2) = 0.5x_1^2 + x_2^2 - 0.3 \cos(3\pi x_1 + 4\pi x_2) + 0.4. \quad (27)$$

We conduct a DOE of size 50 using Latin hypercube sampling (LHS) for each of the source problems to create the source surrogate models. Then, we initiate the Bayesian optimization algorithm using an initial DOE of the target function of 2 random evaluations. We set the number of iterations to 20 and perform optimizations using both VBO and TLBO using ETL. For TLBO we also test two possible methods for variance calculation: TV and AV. We repeat the process for 20 runs to obtain the convergence results presented in [Figure 8](#). We note that TLBO using ETL achieves faster convergence than VBO as shown in [Figure 8\(a\)](#). We note that both TV and AV yield similar convergence results, as seen in the box plots in [Figure 8\(b\)](#), which can be explained by the fact that all source surrogate models are built based on LHS sampling across the same design space as the target problem. For the remainder of this work, we use TV to maintain acquisition function exploration based on the target model.

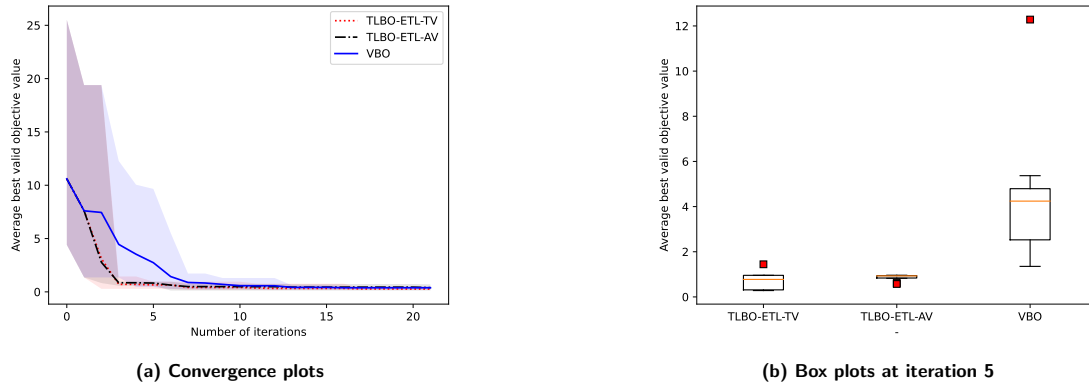


Figure 8: Results of the analytical problem showing the improved convergence using TLBO over the standard VBO over 20 optimization runs.

We also tested our methodology using a benchmark problem typically used for multi-fidelity optimization [30]. The problem is based on variations of the Rosenbrock function $h_{\mathbb{T}}$ by creating a medium fidelity level $h_{\mathbb{S}1}$ and a low fidelity level $h_{\mathbb{S}2}$. In this transfer learning scenario, we consider the low and medium fidelity functions as source problems and the higher fidelity function as the target problem.

$$h_{\mathbb{S}1}(x) = \sum_{i=1}^{D-1} 50(x_{i+1} - x_i^2)^2 + (-2 - x_i)^2 - \sum_{i=1}^D 0.5x_i \quad (28)$$

$$h_{\mathbb{S}2}(x) = \frac{f_1(x) - 4 - \sum_{i=1}^D 0.5x_i}{10 + \sum_{i=1}^D 0.25x_i} \quad (29)$$

$$h_{\mathbb{T}}(x) = \sum_{i=1}^{D-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2. \quad (30)$$

For this problem, we select $D = 5$ and $-2 \leq x_i \leq 2$ for $i = 1, \dots, D$ as the so-called MF2.2 benchmark problem from [30]. We create a DOE of size 100 using LHS using each of $h_{\mathbb{S}1}$ and $h_{\mathbb{S}2}$. We set the size of the initial DOE for the target problem as the number of design variables d . We conduct 20 optimization runs of 100 iterations each and compare TLBO and VBO. We present the results in Figure 9 showing the improved convergence when using TLBO over VBO. In Figure 9(b), we assume that the computational cost of each high fidelity function evaluation is 10 seconds. It is noted that TLBO produces lower average minima throughout the optimization even with the added computational costs as seen in Figure 9(c).

5.3 Aircraft conceptual design optimization

We use an aircraft conceptual design problem to demonstrate our Bayesian optimization transfer learning approach. We present a scenario based on the problem defined in Section 2.2. We select a low number of evaluations for the initial DOE of the target optimization (less than a quarter of the dimension of the input vector x), and 150 evaluations for the DOEs of the source problems. Note that DOE size does not have to be equal across all source problems.

We first perform 10 sets of the initial DOEs consisting of 3 samples each using the target BRAC optimization problem. We use these DOEs to conduct 10 optimization runs for each of TLBO and VBO. We present the results of the optimizations in Figure 10 demonstrating the superior performance of TLBO over VBO. Convergence plots show the average of the minimum normalized objective values at every iteration (shown as a line per method) and the variance at every iteration (shown as pastel colors of the corresponding line). We note the improved convergence behaviour of TLBO in the early

iterations of the optimization even when considering the additional computation cost as can be seen in Figure 10(b). This improvement is also demonstrated in Figure 10(c) where we show result boxplots of the two tested methods at the 15th optimization iteration.

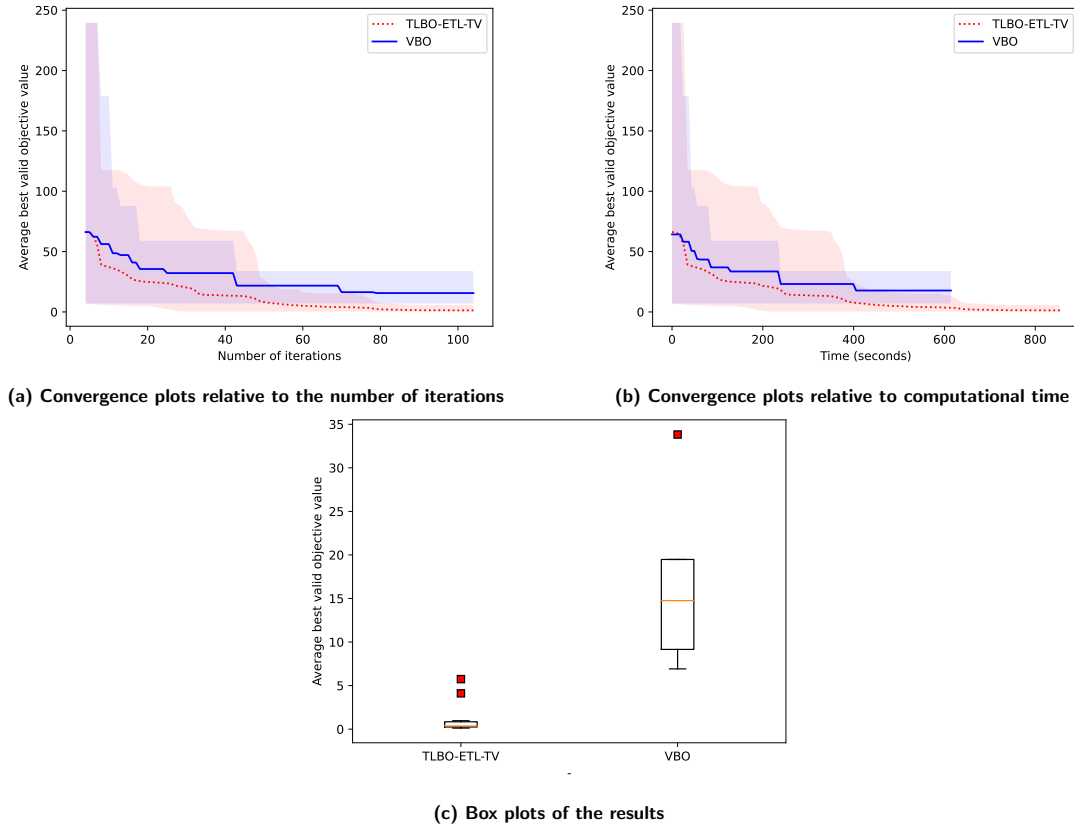


Figure 9: Results of the 5-dimensional Rosenbrock multi-fidelity problem showing the improved convergence of TLBO over VBO over 20 optimizations runs.

We compare the prediction accuracy of the surrogate models in the Bayesian optimization frameworks at the 15th iteration using 100 test points in Figure 11, showing the improved accuracy of TLBO across all the surrogate models for constraints and the objective function, which explains the improved convergence results early in the optimization per Figure 10(a).

5.4 Discussion

While the proposed TLBO method demonstrates improved convergence in the presented examples, practitioners should be aware of scenarios where the method may underperform or fail to provide benefits over standard Bayesian optimization.

Highly dissimilar source and target problems. When source data originates from fundamentally different optimization landscapes—for example, when the source problem is unimodal while the target is highly multimodal—the transferred surrogate models may mislead the optimization search. Although the multi-criteria weighting scheme assigns lower probabilities to poorly matching sources, sufficiently dissimilar problems may still introduce bias in the ensemble predictions, particularly in early iterations when limited target data is available for weight calibration.

Conflicting source models. When multiple source models disagree significantly and suggest optima in different regions of the design space, the weighted ensemble produces an averaged prediction that may not accurately represent any actual optimum. This averaging effect can result in a surrogate landscape

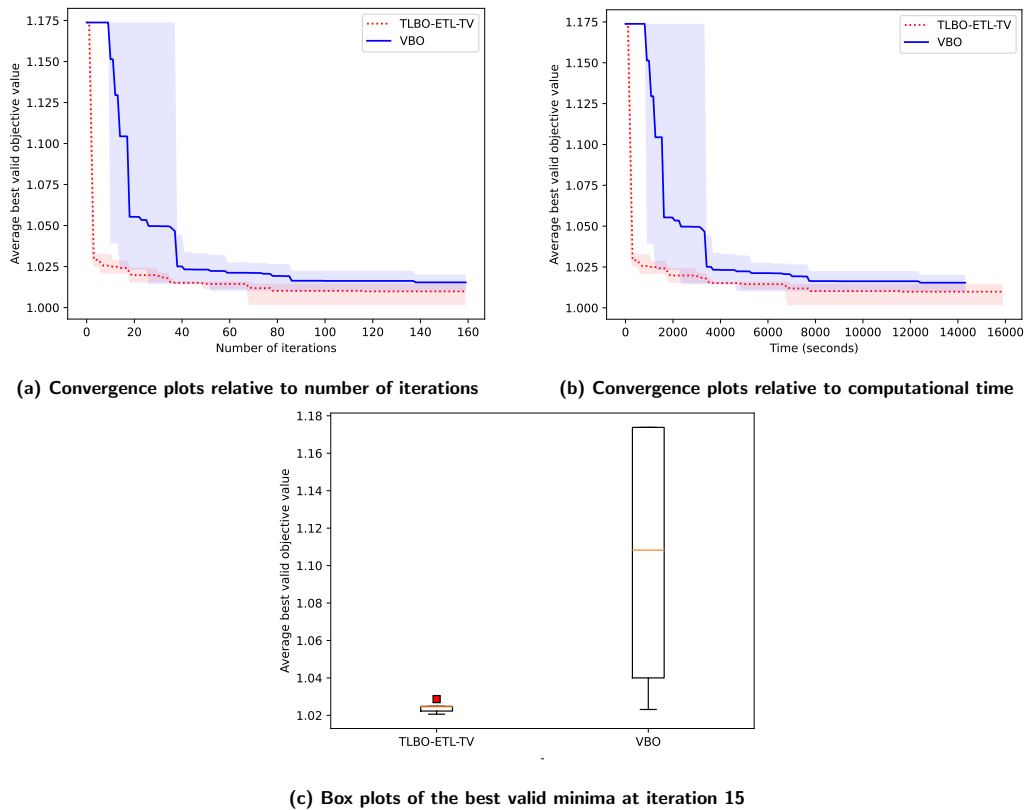


Figure 10: Results of 10 runs of the BRAC optimization problem showing the improved convergence and the variance of TLBO over VBO.

that is smoother than the true target function, potentially causing the optimizer to miss local optima or converge to suboptimal regions.

Insufficient target data for weight calibration. The multi-criteria weighting relies on the target DOE to assess source model quality. With very few initial target samples (e.g., fewer than the number of design variables), the weight estimation may be unreliable, leading to inappropriate source model rankings. In such cases, the alternating TLBO/VBO procedure described in Section 5.4 can help mitigate this risk.

Constraint boundary shifts. When constraints are active in different regions between source and target problems—for instance, if a structural constraint becomes critical at different wing loading conditions—transferring the constraint surrogate may incorrectly predict feasibility boundaries. This is particularly problematic for inequality constraints near their bounds, where small prediction errors can lead to misclassification of feasible and infeasible regions.

Computational overhead. The ensemble of surrogates approach adds computational cost due to source model construction, weight calculation, and ensemble predictions at every iteration. For simple low-dimensional problems, fast blackbox evaluations, or when source data quality is poor, this overhead may not be justified by convergence improvements. Users should assess whether the expected reduction in blackbox evaluations outweighs the added surrogate modeling costs.

Addressing Negative Transfer Negative transfer occurs when source data degrades target optimization performance, typically when source and target problems are insufficiently related. Our methodology mitigates negative transfer through several mechanisms. First, the multi-criteria weighting

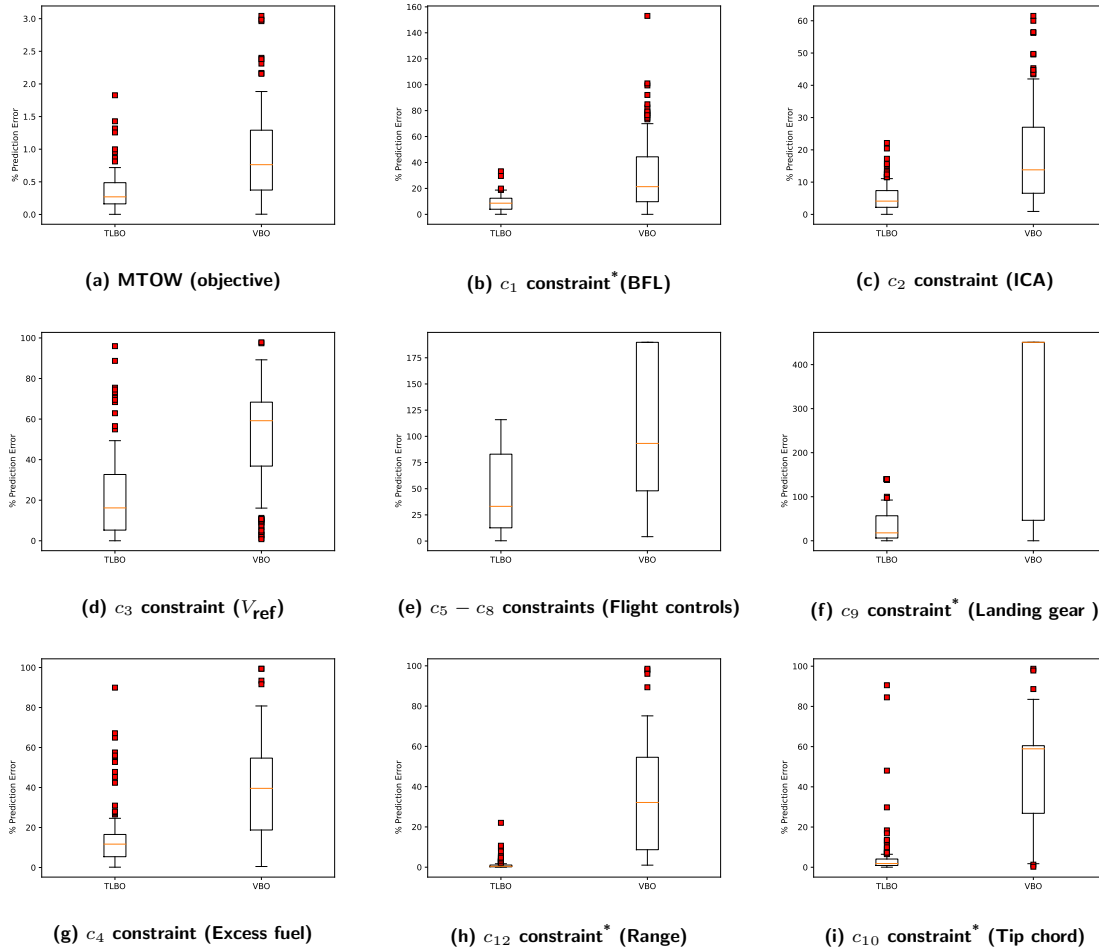


Figure 11: Box plots comparing the prediction errors relative to the objective function and the constraints of TLBO and VBO at iteration 15 of the optimization. * denotes active inequality constraints (less than 2% margin) at the minimum observed valid objective function using 100 test evaluations.

scheme (Section 3.4.2) automatically assigns lower probabilities to source models that poorly match the target data, effectively down-weighting irrelevant sources. Second, the scaling and bias adjustment in the transfer of prior (Equation (9)) allows source models to adapt to the target function’s range, reducing the impact of value mismatches. Third, for cases where negative transfer may still occur, an alternating procedure can be employed: the optimization alternates between TLBO and the standard Bayesian optimization (VBO) at specified intervals, allowing the algorithm to rely on target-only surrogates when the transfer learning ensemble under performs. This hybrid approach ensures robustness by periodically resetting to vanilla Bayesian optimization, thereby bounding the potential negative impact of misleading source information while still benefiting from positive transfer when source data is informative.

6 Conclusions

In this work, we advanced the state-of-the-art for transfer learning in Bayesian optimization by presenting a novel method to address constraints and heterogeneity. Our method is based on an ensemble of surrogates using transfer learning algorithm which we embedded in a Bayesian optimization framework. We propose two approaches to handle heterogeneity between source and target optimization problems: the first is based on meta data adjusted dimension reduction aimed to solve design space heterogeneity,

and the second approach addresses heterogeneity of constraints by selecting which constraints are to be used as source constraints in a target optimization problem. The method adds computational costs to the standard Bayesian optimization algorithm due to the additional source model selection step at every iteration. We validated the proposed methodology using an aircraft conceptual design optimization scenario and compared the results with standard Bayesian optimization. The results show significant improvement in convergence early in the optimization even with the increased computational cost due to the introduction of the ensemble of surrogates.

Future work includes extending the weighted ensemble to a mixture of experts approach with multiple clusters, where each cluster can have different source model weights based on local performance as presented for surrogate modeling in [45]. This would allow the ensemble to adapt spatially across the design space, potentially improving performance in regions where different source models are more accurate.

Additional future research directions include: extending the application to more aircraft design optimization test cases with varying fidelity of simulation models to understand the impact of source data fidelity on optimization results; investigating adaptive schemes that automatically select between TLBO and standard Bayesian optimization based on detected negative transfer; and developing methods to quantify source-target similarity *a priori* to guide practitioners in selecting appropriate source data for transfer.

References

- [1] Charles Audet, Edward Hallé-Hannan, and Sébastien Le Digabel. A general mathematical framework for constrained mixed-variable blackbox optimization problems with meta and categorical variables. In *Operations Research Forum*, volume 4, page 12. Springer, 2023.
- [2] Tianyi Bai, Yang Li, Yu Shen, Xinyi Zhang, Wentao Zhang, and Bin Cui. Transfer learning for Bayesian optimization: a survey. *arXiv:2302.05927*, 2023.
- [3] Runxue Bao, Yiming Sun, Yuhe Gao, Jindong Wang, Qiang Yang, Haifeng Chen, Zhi-Hong Mao, Xing Xie, and Ye Ye. A survey on heterogeneous transfer learning. *arXiv:2310.08459*, 2023.
- [4] Nathalie Bartoli, Mohamed-Amine Bouhleb, Igor Kurek, Rémi Lafage, Thierry Lefebvre, Joseph Morlier, Rémy Priem, Vivien Stilz, and Rommel Regis. Improvement of efficient global optimization with application to aircraft wing design. In *17th AIAA/ISSMO Multidisciplinary analysis and optimization conference*, pages AIAA 2016-4001, 2016.
- [5] Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Romain Olivanti, Rémy Priem, Nicolas Bons, Joaquim R.R.A. Martins, and Joseph Morlier. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerospace Science and technology*, 90:85-102, 2019.
- [6] Ihor O. Bohachevsky, Mark E. Johnson, and Myron L. Stein. Generalized simulated annealing for function optimization. *Technometrics*, 28(3):209-217, 1986.
- [7] Mohamed Amine Bouhleb, Nathalie Bartoli, Abdelkader Otsmane, and Joseph Morlier. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53:935-952, 2016.
- [8] Pavel Brazdil, João Gama, and Bob Henery. Characterizing the applicability of classification algorithms using meta-level learning. In *Machine Learning: ECML-94: European Conference on Machine Learning Catania, Italy, April 6-8, 1994 Proceedings 7*, pages 83-102. Springer, 1994.
- [9] Hao Chen, Zhilang Zhang, Weikun Li, Qimeng Liu, Kai Sun, Dixia Fan, and Weicheng Cui. Ensemble of surrogates in black-box-type engineering optimization: Recent advances and applications. *Expert Systems with Applications*, page 123427, 2024.
- [10] Junfeng Chen and Changxing Liu. Efficient one-dimensional turbomachinery design method based on transfer learning and Bayesian optimization. *SN Applied Sciences*, 4(10):255, 2022.
- [11] Vassiliy A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153-158, 1969.
- [12] Zhou Fan, Xinran Han, and Zi Wang. Transfer learning for Bayesian optimization on heterogeneous search spaces. *Transactions on Machine Learning Research*, 1616, 2024.

- [13] Matthias Feurer, Benjamin Letham, and Eytan Bakshy. Scalable meta-learning for Bayesian optimization using ranking-weighted gaussian process ensembles. In *AutoML Workshop at ICML*, volume 7, 2018.
- [14] Matthias Feurer, Benjamin Letham, Frank Hutter, and Eytan Bakshy. Practical transfer learning for Bayesian optimization. *arXiv:1802.02219*, 2018.
- [15] Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, New Jersey, NY, 2008.
- [16] Jacob R. Gardner, Matt J. Kusner, Zhixiang Eddie Xu, Kilian Q. Weinberger, and John P Cunningham. Bayesian optimization with inequality constraints. In *ICML*, volume 2014, pages 937–945, 2014.
- [17] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. *Advances in Neural Information Processing Systems*, 19, 2006.
- [18] Sigrid Passano Hellan, Christopher G. Lucas, and Nigel H. Goddard. Data-driven Prior Learning for Bayesian Optimisation. *arXiv:2311.14653*, 2023.
- [19] Momin Jamil and Xin-She Yang. A literature survey of benchmark functions for global optimization problems. *CoRR*, abs/1308.4008, 2013.
- [20] Timothy M.S. Jim, Ghifari A. Faza, Pramudita S. Palar, and Koji Shimoyama. Bayesian optimization of a low-boom supersonic wing planform. *AIAA journal*, 59:4514–4529, 2021.
- [21] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.
- [22] Siraj Khan, Pengshuai Yin, Yuxin Guo, Muhammad Asim, and Ahmed A Abd El-Latif. Heterogeneous transfer learning: recent developments, applications, and challenges. *Multimedia Tools and Applications*, pages 1–37, 2024.
- [23] Daniel G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.
- [24] Rémi Lam, Matthias Poloczek, Peter Frazier, and Karen E Willcox. Advances in Bayesian optimization with applications in aerospace engineering. In *2018 AIAA Non-Deterministic Approaches Conference*, pages AIAA 2018–1656, 2018.
- [25] Andrew B. Lambe and Joaquim R.R.A. Martins. Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes. *Structural and Multidisciplinary Optimization*, 46:273–284, 2012.
- [26] Wen Li, Lixin Duan, Dong Xu, and Ivor W. Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1134–1148, 2013.
- [27] Marius Lindauer, Katharina Eggenesperger, Matthias Feurer, André Biedenkapp, Difan Deng, Carolin Benjamins, Tim Ruhkopf, René Sass, and Frank Hutter. SMAC3: A versatile Bayesian optimization package for hyperparameter optimization. *Journal of Machine Learning Research*, 23(54):1–9, 2022.
- [28] Marius Lindauer and Frank Hutter. Warmstarting of model-based algorithm configuration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [29] Negareh Mahboubi, Junyao Xie, and Biao Huang. Point-by-point transfer learning for bayesian optimization: An accelerated search strategy. *Computers & Chemical Engineering*, 194:108952, 2025.
- [30] Laura Mainini, Andrea Serani, Markus P. Rumpfkeil, Edmondo Minisci, Domenico Quagliarella, Hüseyin Pehlivan, Sihmehmet Yildiz, Simone Ficini, Riccardo Pellegrini, Francesco Di Fiore, et al. Analytical benchmark problems for multifidelity optimization methods. *arXiv:2204.07867*, 2022.
- [31] Alan Tan Wei Min, Abhishek Gupta, and Yew-Soon Ong. Generalizing transfer Bayesian optimization to source-target heterogeneity. *IEEE Transactions on Automation Science and Engineering*, 18(4):1754–1765, 2020.
- [32] Samuel Müller, Matthias Feurer, Noah Hollmann, and Frank Hutter. PFNs4BO: In-context learning for Bayesian optimization. In *International Conference on Machine Learning*, pages 25444–25470. PMLR, 2023.
- [33] Shuteng Niu, Yongxin Liu, Jian Wang, and Houbing Song. A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, 1(2):151–166, 2020.
- [34] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, and et al. *Scikit-learn: Machine learning in Python*. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [35] Pat Piperni, Antoine DeBlois, and Ryan Henderson. Development of a multilevel multidisciplinary-optimization capability for an industrial environment. *AIAA journal*, 51:2335–2352, 2013.

- [36] Remy Priem, Hugo Gagnon, Ian Chittick, Stephane Dufresne, Youssef Diouane, and Nathalie Bartoli. An efficient application of Bayesian optimization to an industrial MDO framework for aircraft design. *AIAA Aviation 2020 Forum*, pages AIAA 2020-3152, 2020.
- [37] Thomas A. Reist, David Koo, David W. Zingg, Pascal Bochud, Patrice Castonguay, and David Leblond. Cross-validation of high-fidelity aerodynamic shape optimization methodologies for aircraft wing-body optimization. *AIAA Journal*, 2019.
- [38] Adriano Rivolli, Luís P.F. Garcia, Carlos Soares, Joaquin Vanschoren, and André CPLF de Carvalho. Characterizing classification datasets: a study of meta-features for meta-learning. *arXiv:1808.10406*, 2018.
- [39] Satadru Roy, Kenneth Moore, John T. Hwang, Justin S. Gray, William A. Crossley, and Joaquim R.R.A. Martins. A mixed integer efficient global optimization algorithm for the simultaneous aircraft allocation-mission-design problem. In *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pages AIAA 2017-1305, 2017.
- [40] P. Saves, R. Lafage, N. Bartoli, Y. Diouane, and et al. SMT 2.0: A Surrogate Modeling Toolbox with a focus on Hierarchical and Mixed Variables Gaussian Processes. *Advances in Engineering Software*, 188:103571, 2024.
- [41] Paul Saves, Nathalie Bartoli, Youssef Diouane, Thierry Lefebvre, Joseph Morlier, Christophe David, Eric Nguyen Van, and Sébastien Defoort. Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design. In *AIAA SciTech 2022 Forum*, pages AIAA 2022-0082, 2022.
- [42] Paul Saves, Youssef Diouane, Nathalie Bartoli, Thierry Lefebvre, and Joseph Morlier. A general square exponential kernel to handle mixed-categorical variables for Gaussian process. In *AIAA AVIATION 2022 Forum*, pages AIAA 2022-3870, 2022.
- [43] Matthias Schonlau. Computer experiments and global optimization. PhD thesis, University of Waterloo, 1997.
- [44] Jost Tobias Springenberg, Aaron Klein, Stefan Falkner, and Frank Hutter. Bayesian optimization with robust Bayesian neural networks. *Advances in Neural Information Processing Systems*, 29, 2016.
- [45] Ali Tfaily, Nathalie Bartoli, Youssef Diouane, and Michael Kokkolaras. Transfer learning in surrogate modeling with emphasis on aircraft design. *Structural and Multidisciplinary Optimization*, 69(2):41, 2026.
- [46] Jan N Van Rijn and Frank Hutter. Hyperparameter importance across datasets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2367-2376, 2018.
- [47] Ying Wei, Peilin Zhao, and Junzhou Huang. Meta-learning hyperparameter performance prediction with neural processes. In *International Conference on Machine Learning*, pages 11058-11067. PMLR, 2021.
- [48] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3:1-40, 2016.
- [49] Christopher Williams and Carl Rasmussen. *Gaussian processes for machine learning*. MIT press, Cambridge, MA, 2006.
- [50] Martin Wistuba and Josif Grabocka. Few-shot Bayesian optimization with deep kernel surrogates. *arXiv:2101.07667*, 2021.
- [51] Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Two-stage transfer surrogate model for automatic hyperparameter optimization. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16*, pages 199-214. Springer, 2016.
- [52] Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. Scalable gaussian process-based transfer surrogates for hyperparameter optimization. *Machine Learning*, 107(1):43-78, 2018.
- [53] Dani Yogatama and Gideon Mann. Efficient transfer learning method for automatic hyperparameter tuning. In *Artificial intelligence and statistics*, pages 1077-1085. PMLR, 2014.
- [54] Rosen Ting-Ying Yu. *Constrained and high-dimensional Bayesian optimization with transformers*. Master of science in computational science and engineering, Massachusetts Institute of Technology, 2025.
- [55] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43-76, 2020.