



**HAL**  
open science

## Multi-objective bayesian optimization with mixed-categorical design variables for expensive-to-evaluate aeronautical applications

Nathalie Bartoli, Thierry Lefebvre, Rémi Lafage, Paul Saves, Youssef Diouane, Joseph Morlier, Jasper Bussemaker, Giuseppa Donelli, Joao Marcos Gomes de Mello, Massimo Mandorino, et al.

### ► To cite this version:

Nathalie Bartoli, Thierry Lefebvre, Rémi Lafage, Paul Saves, Youssef Diouane, et al.. Multi-objective bayesian optimization with mixed-categorical design variables for expensive-to-evaluate aeronautical applications. AEROBEST 2023, ECCOMAS, Jul 2023, Lisbonne, Portugal. pp.436. hal-04170287

**HAL Id: hal-04170287**

**<https://hal.science/hal-04170287>**

Submitted on 25 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# MULTI-OBJECTIVE BAYESIAN OPTIMIZATION WITH MIXED-CATEGORICAL DESIGN VARIABLES FOR EXPENSIVE-TO-EVALUATE AERONAUTICAL APPLICATIONS

N. Bartoli<sup>1\*</sup>, T. Lefebvre<sup>1</sup>, R. Lafage<sup>1</sup>, P. Saves<sup>1,2</sup>, Y. Diouane<sup>3</sup>, J. Morlier<sup>2</sup>,  
J. H. Bussemaker<sup>4</sup>, G. Donelli<sup>4</sup>, J. M. Gomes de Mello<sup>5</sup>, M. Mandorino<sup>6</sup>, P.  
Della Vecchia<sup>6</sup>

1: ONERA/DTIS,  
Université de Toulouse,  
31400 Toulouse, France  
{nathalie.bartoli,thierry.lefebvre,remi.lafage,paul.saves}@onera.fr

2: ICA, ISAE-SUPAERO, INSA, CNRS, MINES ALBI, UPS  
Université de Toulouse,  
31400 Toulouse, France  
joseph.morlier@isae-supero.fr

3: Polytechnique Montréal  
Montreal, QC, Canada  
youssef.diouane@polymtl.ca

4: German Aerospace Center (DLR)  
Institute of System Architectures in Aeronautics  
Hein-Saß-Weg 22, 21129 Hamburg, Germany  
{jasper.bussemaker,giuseppa.donelli}@dlr.de

5: EMBRAER S.A.  
Av. Brigadeiro Faria Lima, 2170, São José dos Campos, Brazil  
joao.mello@embraer.com.br

6: UNINA  
University di Napoli Federico II  
Naples, 80125, Italy  
{massimo.mandorino,pierluigi.dellavecchia}@unina.it

**Abstract.** *This work aims at developing new methodologies to optimize computational costly complex systems (e.g., aeronautical engineering systems). The proposed surrogate-based method (often called Bayesian optimization) uses adaptive sampling to promote a trade-off between exploration and exploitation. Our in-house implementation, called SEGOMOE, handles a high number of design variables (continuous, discrete or categorical) and nonlinearities by combining mixtures of experts for the objective and/or the constraints. Additionally, the method handles multi-objective optimization settings, as it allows the construction of accurate Pareto fronts with a minimal number of function evaluations. Different infill criteria have been implemented to handle multiple objectives with or without constraints. The effectiveness of the proposed method was tested on practical aeronautical applications within the context of the European Project AGILE 4.0 and*

*demonstrated favorable results. A first example concerns a retrofitting problem where a comparison between two optimizers have been made. A second example introduces hierarchical variables to deal with architecture system in order to design an aircraft family. The third example increases drastically the number of categorical variables as it combines aircraft design, supply chain and manufacturing process. In this article, we show, on three different realistic problems, various aspects of our optimization codes thanks to the diversity of the treated aircraft problems.*

**Keywords:** Bayesian optimization, multi-objective problems, mixed-categorical variables, aircraft design, aeronautical applications

## 1 INTRODUCTION

For the last few decades, the aeronautical industry has been essentially driven by incremental improvements based on aircraft design optimization [1]. The use of computer science allowed to reuse well-proven design processes and to benefit from preexisting results to reduce the time dedicated to development. Nonetheless, the forthcoming challenges awaiting aircraft design call for large-scale improvements to meet future expectations in terms of overall environmental impact, noise reduction and cost-effectiveness [2]. These large-scale improvements, considered through novel aircraft configurations, involve an increased proximity between Aerodynamics, Propulsion and Structural mechanics, thus requiring to explore coupled and uncharted physics. As is the case for many industrial complex systems, some powerful optimization algorithms for aircraft design are needed. The H2020 European project AGILE 4.0 (2019-2023)<sup>1</sup> [3], led by DLR, has identified several challenging application cases linked to Multidisciplinary Design Optimization (MDO) methods [4] and Model Based Systems Engineering (MBSE) technologies [5] that will require innovative and sustainable aircraft configurations in the years to come [6]. The main objective is to reduce the environmental impact in terms of fuel consumption, waste and emissions associated with all the aeronautical system activities and operations. This purpose requires extending aeronautical research to cover the entire aircraft lifecycle, from design and production to waste disposal after the system's end-of-life. Consequently, the challenge is to incorporate these new requirements into the early design phase, enabling strategic decision-making processes to optimize the entire aircraft lifecycle [7]. To do this, collaborative multidisciplinary aircraft design and optimization should be carried out involving not-only the aircraft design domain (typically accounted during the conceptual aircraft design) but also other industrial domains, such as manufacturing, supply chain, maintenance or certification [8]. From the different application cases of the project, some quantities of interest have been identified from the Multidisciplinary Design Analysis (MDA) and optimization problems have been deduced involving from 2 to 5 objective functions and several nonlinear constraints that are given by non derivative and expensive-to-evaluate black box functions. Moreover, when evaluating new architecture systems, it is important to take into account categorical variables such as on-board system architecture (conventional, hybrid, electric for instance) or property materials (aluminum, titanium), which do not have a defined order [9]. Different approaches have been investigated to deal

---

<sup>1</sup><https://www.agile4.eu/>

with mixed integer variables, multiple objectives and constraints handling requirements. These optimizations return a set of feasible configurations considered as optimal trade-offs. This set of solutions is the so-called *Pareto front* that consists in the set of solutions that cannot be improved in one objective without worsening another [10].

Surrogate-Based Optimization (SBO) is a powerful technique for optimizing complex systems requiring a significant amount of computational resources. For multi-objective optimization problems, it involves building a surrogate model, such as a machine learning model, to approximate the behavior of multiple objectives simultaneously, allowing for efficient exploration of the search space and identification of the Pareto front. The surrogate models built offline are then used to predict the objective function values at different points in the search space, enabling the optimization algorithm to focus its computational resources on the most promising regions of the search space [11].

To avoid this offline cost involving a large number of function evaluations, some adaptive strategies have been introduced. Nowadays, Bayesian Optimization (BO) [12] is well known as a powerful tool in scientific research for optimizing complex, expensive-to-evaluate functions with unknown properties. It is particularly useful when only a limited number of evaluations is available which is often the case with scientific experiments. When multiple objectives are involved, multi-objective techniques must be used to find the set of optimal solutions that maximize or minimize these multiple objectives simultaneously, rather than focusing on a single objective. Multi-objective optimization using Bayesian algorithms is a potent strategy for optimizing intricate systems that possess multiple, conflicting objectives. Bayesian optimization techniques allow for efficient exploration of the search space by harnessing probabilistic models to forecast the behavior of the objectives being optimized. This facilitates a trade-off between exploration and exploitation and allows the inclusion of prior knowledge on the objectives.

In summary, multi-objective optimization with Bayesian algorithms is a valuable tool for scientific research, providing an efficient way to identify the Pareto front and make informed decisions based on multiple, conflicting objectives. Here some illustrations are done on three application cases of the AGILE 4.0 project with different characteristics. On the first application case, about retrofitting concepts (engine, winglet, on-board systems), the different impacts on aircraft performance are characterized (e.g., Direct Operating Cost (DOC), emissions) to select the best retrofitting strategy. The second application is about designing an aircraft family composed of three different aircrafts with different commonality choices (wing, engine, empennage). Again, the different solutions are characterized by different impacts on aircraft performance (e.g. mass, aerodynamic efficiency) and different commonality choices. To finish with, the third application concerns Horizontal Tail Plane (HTP) production and search for an equilibrium between airplane performance, manufacturing costs and the efficiency of the supply chain.

The outline of the paper is as follows. In Section 2, the multi-objective optimization problem is presented and two proposed frameworks are then described. A first SBO algorithm JPAD Optimizer is introduced in Section 2.2 and a second BO framework SEGOMOE is given in Section 2.3 where the Continuous Relaxation (CR) as well as the use of the Partial Least Squares (PLS) technique are detailed. Some details concerning the remote access to the frameworks are also provided. Then Section 3 presents three MDO application cases where multiple physics and domains are involved. For each of the application cases, multidisciplinary analysis, optimization problem as well as the obtained

results are detailed. Conclusions and perspectives are finally drawn in Section 4.

## 2 MULTI-OBJECTIVE OPTIMIZATION ALGORITHMS

To begin with, we start formulating the multi-objective paradigm over which the paper will be based. Then, we introduce our two frameworks used to solve such multi-objective optimization problems in an industrial context.

### 2.1 Multi-objective formalism

Consider the multi-objective constrained optimization problem

$$\left\{ \begin{array}{l} \min_{\mathbf{x} \in \Omega \times S \times \mathbb{F}^l} \mathbf{f}(\mathbf{x}) := [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})] \\ \text{s.t. } g_1(\mathbf{x}) \leq 0 \\ \quad \vdots \\ g_m(\mathbf{x}) \leq 0, \end{array} \right. \quad (1)$$

where  $\Omega \subset \mathbb{R}^d$  represents the bounded continuous design set for the  $d$  continuous variables,  $S \subset \mathbb{Z}^\ell$  represents the bounded integer set where  $L_1, \dots, L_\ell$  are the numbers of levels of the  $\ell$  quantitative integer variables on which we can define an order relation, and  $\mathbb{F}^l := \{1, \dots, L_1\} \times \{1, \dots, L_2\} \times \dots \times \{1, \dots, L_l\}$  is the design space for  $l$  categorical qualitative variables with their respective  $L_1, \dots, L_l$  levels [9]. The objectives functions are referred as  $f_i : \mathbb{R}^d \rightarrow \mathbb{R} \quad \forall i = 1, \dots, n$  where  $n$  is the number objectives. The constraints are named  $g_j : \mathbb{R}^d \rightarrow \mathbb{R} \quad \forall j = 1, \dots, m$  where  $m$  is the number of constraints (inequality or equality). Due to potentially conflicting objectives, the solution of the optimization is not unique but a range of solutions is proposed. The trade-off between these solutions is characterized by the notion of dominance: a solution  $\mathbf{x}$  is said to dominate another solution  $\mathbf{x}'$  and denoted by  $\mathbf{x} \preceq \mathbf{x}'$  if

$$f_i(\mathbf{x}) \leq f_i(\mathbf{x}') \quad \forall i = 1, \dots, n.$$

The set of solutions representing the optimal trade-off is referred to the Pareto set (PS) and the corresponding image of PS in the objective space is known as the Pareto front (PF)

$$\text{PF} := \{\mathbf{f}(\mathbf{x}) | \mathbf{x} \in \text{PS}\}.$$

Figure 1 illustrates a Pareto front (red points) relative to two objectives. An approximation to this PF is given by the scattered green dots and the associated dominated hypervolume given by the green area. Both the SBO and BO frameworks to solve problem (1) are described in the following.

### 2.2 Surrogate-Based Optimization with JPAD

In this work, one of the optimization tools used is the JPAD Optimizer, which is based on the Multi-Objective Evolutionary Algorithms (MOEA) framework. It should be noted that the JPAD Optimizer is an independent implementation within the JPAD library, as described in [13, 14]. The MOEA framework is a Java library that serves as an open-source platform for the development and experimentation of various optimization techniques. The latter offers a range of pre-implemented algorithms that can be readily

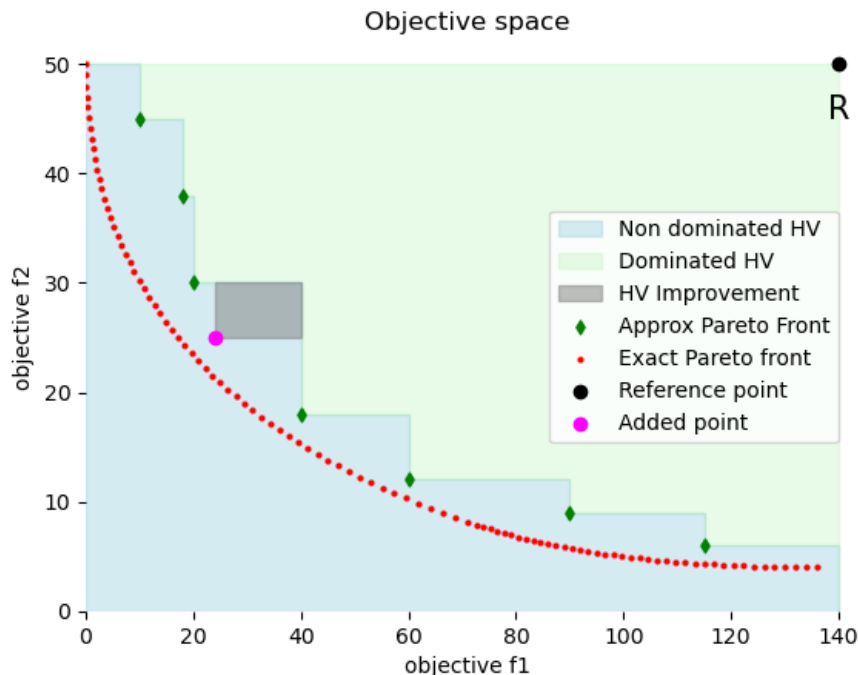


Figure 1: Hypervolume Improvement: the hypervolume indicator of the non-dominated set (green points) corresponds to the area dominated by it, up to  $R$  (reference point in black). The gray rectangle is the hypervolume improvement brought by the new added point in magenta.

employed, such as genetic algorithms and particle swarm optimization, among others. For this particular study, the  $\epsilon$ -NSGAI algorithm has been used. The  $\epsilon$ -NSGAI algorithm extends the NSGA-II algorithm by incorporating an  $\epsilon$ -dominance archive and randomized restarts, thereby enhancing the search process and enabling the discovery of a diverse set of Pareto optimal solutions. A comprehensive explanation of this algorithm can be found in [15] with in-depth details and comparisons. The JPAD Optimizer requires a comprehensive factorial database to conduct the analysis effectively. This database enables the solver to construct a specific Response Surface Model (RSM) through spline interpolation. The RSM serves as a surrogate model, capturing the relationship between the input variables and the response variable. By utilizing the RSM as a surrogate model, the optimization algorithm can efficiently explore the design space and identify optimal solutions. However, it is important to note that direct inclusion of categorical variables in the optimization is not feasible. To address this limitation, one potential solution is to perform the optimization iteratively, considering each categorical variable separately for each run. Still, using the JPAD Optimizer in conjunction with the MOEA framework provides a robust and versatile approach to the optimization process. The selection of the  $\epsilon$ -NSGAI algorithm within the MOEA framework allowed for efficient and effective exploration of the design space, leading to the identification of a diverse range of Pareto optimal solutions. The chosen references provide additional information on the JPAD library, the MOEA framework, and the  $\epsilon$ -NSGAI algorithm, serving as valuable resources for readers interested in a more detailed understanding of the employed methodologies.

### 2.3 Bayesian Optimization with SEGOMOE

The second proposed optimization process is based on a sequential enrichment approach, typically the Efficient Global Optimization algorithm [11] or Super EGO [16], an evolution of EGO to handle constraints. Bayesian Optimization (BO) is based on Gaussian processes (GP) [17] (also denoted by Kriging [18]) and here the idea is to use some adaptive mixture of Kriging based models to tackle high dimension problems. The mixture of experts (MOE) [19, 20] is known to approximate complex functions with heterogeneous behaviour by combining local surrogate models in a global one. In order to consider high-dimensional functions to approximate objective functions and constraints, we used adapted local Kriging-based models [21, 22]. Some recent developments have been made to consider highly non linear constraints [23], mixed integer variables [24] and multi-objective applications [10]. The general framework called Super Efficient Global Optimization coupled with Mixture Of Experts (SEGOMOE) has been proposed by ONERA and ISAE-SUPAERO. The performance of SEGOMOE is validated and proven on different analytical and industrial test cases [25–27].

Here we consider multiple objectives and constraints with mixed integer variables. Some continuous, integer and categorical variables are involved within the different application cases, so the GP surrogate models have to be adapted to deal with. We proposed to combine continuous relaxation to replace categorical variables by continuous ones and Partial Least Squares techniques to handle the increase of dimensionality. The continuous relaxation introduced by [28] uses a one-hot encoding strategy [29] and transforms integer and categorical inputs into continuous ones, so some classic continuous correlation kernels can still be used to build the Kriging models. The design space  $\Omega \times S \times \mathbb{F}^l$  is relaxed to a continuous space  $\Omega'$  constructed on the following way [24]:

- $\forall i \in \{1, \dots, \ell\}$ , the integer variable  $z_i$  is relaxed within its bounds and treated as continuous.
- $\forall j \in \{1, \dots, l\}$ , we use a relaxed one-hot encoding [29] for the categorical variable  $c_j$  (and its  $L_j$  associated levels) and add  $L_j$  new continuous dimensions into  $\Omega'$ .

Therefore, we get, after relaxation, a new design space  $\Omega' \subseteq \mathbb{R}^{d'}$  where  $d' = d + \ell + \sum_{j=1}^l L_j > d + \ell + l$ . As the number of variables may increase drastically (from  $d$  to  $d'$ ), one new added variable associated to each categorical level, PLS technique is combined with Kriging, denoted as KPLS and KPLS-K models [21, 22], to reduce the number of hyperparameters. During experiments, the chosen number of principal components within KPLS models does not exceed 4 or 5. The code implementation of the proposed GP has been released in the toolbox SMT v1.1<sup>2</sup> [30].

Once the initial Design of Experiments is given in the relaxed space  $\Omega'$ , Kriging based surrogate models are built for the objective and constraint functions. Each costly function  $f_i(\mathbf{x})$  is approximated by a GP characterized by its mean  $\hat{y}_i(\mathbf{x}) : \mathbb{R}^{d'} \rightarrow \mathbb{R}$  and its standard deviation  $s_i(\mathbf{x}) : \mathbb{R}^{d'} \rightarrow \mathbb{R}$

$$\hat{f}_i(\mathbf{x}) \sim \mathcal{N}(\hat{y}_i(\mathbf{x}), s_i^2(\mathbf{x})) \quad i = 1, \dots, n$$

---

<sup>2</sup><https://smt.readthedocs.io/en/latest/>

For the multi-objective, we assume that the  $n$  components of  $\mathbf{f}$  are independent to define  $\hat{\mathbf{f}}$  as the surrogate model associated to each component

$$\hat{\mathbf{f}}(x) \sim \mathcal{N}(\hat{\mathbf{y}}(\mathbf{x}), \Sigma(\mathbf{x}))$$

where  $\hat{\mathbf{y}}(\mathbf{x}) : \mathbb{R}^{d'} \rightarrow \mathbb{R}^n$  is the GP prediction vector given by  $[\hat{y}_1(\mathbf{x}), \dots, \hat{y}_n(\mathbf{x})]$  and  $\Sigma(\mathbf{x})$  is a diagonal matrix whose diagonal vector is given by  $s_i^2(\mathbf{x}), \forall i = 1, \dots, n$ . So the initial problem defined by Eq.(1) is replaced by an infill problem defined as follows

$$\begin{cases} \max_{x \in \mathbb{R}^{d'}} & \alpha_{\mathbf{f}}^{\text{reg}}(\mathbf{x}) \\ \text{s.t.} & \hat{g}_1(\mathbf{x}) \leq 0 \\ & \vdots \\ & \hat{g}_m(\mathbf{x}) \leq 0, \end{cases} \quad (2)$$

where  $\hat{g}_i(\mathbf{x})$  corresponds to the mean prediction of the GP constraint models and the regularized acquisition function [10] is defined by

$$\alpha_{\mathbf{f}}^{\text{reg}}(\mathbf{x}) := \gamma \alpha_{\mathbf{f}}(\mathbf{x}) - \psi(\mu_{\mathbf{f}}(\mathbf{x})) \quad (3)$$

where  $\alpha_{\mathbf{f}}(\mathbf{x})$  is a standard scalar acquisition (EHVI, PI, MPI, ...) depending on  $\hat{\mathbf{y}}(\mathbf{x})$  and  $\Sigma(\mathbf{x})$ , and  $\gamma$  is a constant parameter. The function  $\psi : \mathbb{R}^{d'} \rightarrow \mathbb{R}$  is a scalarization operator. Different choices exist for the function  $\psi$  and in [10] two options were investigated. Namely, for a given  $\hat{\mathbf{y}}(\mathbf{x}) \in \mathbb{R}^{d'}$ , we consider

$$\begin{aligned} (\text{reg} = \max) & : \psi(\hat{\mathbf{y}}(\mathbf{x})) := \max_{i \leq d'} \hat{y}_i(\mathbf{x}) \\ (\text{reg} = \text{sum}) & : \psi(\hat{\mathbf{y}}(\mathbf{x})) := \sum_{i=1}^{d'} \hat{y}_i(\mathbf{x}) \end{aligned}$$

The performance of this regularized acquisition function (3) has been tested on analytical examples involving continuous variables and apply with success to an aircraft design test case in [10].

Concerning the different BO criteria relative to the hypervolume Improvement, an illustration is proposed in Fig.1. The Expected Hyper-Volume Improvement (EHVI [31, 32]), the Probability of Improvement (PI [33]), and the Minimum of Probability of Improvement (MPI [34]) are some multi-objective extensions of the well known Expected Improvement (EI [11]). As seen in Fig.1, with respect to the predictions in the Gaussian random field, the idea is to measure how much hypervolume improvement (grey area) could be achieved by evaluating the new point (magenta point), considering the uncertainty of the prediction. These criteria differ depending how they favor well-spread solutions: for instance for EHVI the hypervolume increase is small when adding a new value close to an already observed one in the objective space. To solve problem Eq.(2), any optimization algorithms capable of considering non linear constraints based either on derivative free optimizer such as COBYLA (Constrained Optimization BY Linear Approximation [35]) or based on gradient method such as SLSQP (Sequential Least Squares Programming [36]) or SNOPT (Sparse Nonlinear OPTimizer [37]) can be used together with a multistart strategy. This adaptive process is repeated until the total budget is reached. The feasible



points of the final database represent the known Pareto optimal points. Nevertheless, as the set of known points has been enriched sequentially to increase the hypervolume, the final database can be used to build GP models for objectives and constraints as a post-processing step. These final GP models can be coupled to an evolutionary algorithm to deal with the multi-objective constrained problem and retrieve the approximated PF. The well known NSGA-II algorithm (Non-dominated Sorting Genetic Algorithm II [38]) of the pymoo [39] toolbox<sup>3</sup> is used for that purpose and ultimately provides the predicted PF based on GP with almost no additional computational cost.

In the end, the proposed strategy provides two outputs: the PF database and the predicted PF. Comparing the proximity between the two PF is a good criterion to know if some additional enrichment points are needed or if the accuracy is sufficient. Algorithm 1 details the SEGOMOE optimization procedure.

---

**Algorithm 1:** SEGOMOE for constrained multi-objective and mixed-integer problems.

---

**Inputs:** Initial DoE  $\mathcal{D}_0$  and set  $t = 0$ ;

**while** the stopping criterion is not satisfied **do**

1. Relax continuously integer and categorical input variables to a real bounded space  $\Omega'$  of dimension  $d' = d + \ell + \sum_{j=1}^l L_j$ . Namely, we continuously relax the mixed categorical DoE to a continuous DoE  $\mathcal{D}_t$  using the relaxation procedure;
2. Build the GP model for each objective function  $f_i(\mathbf{x})$  and each constraint  $g_j(\mathbf{x})$  related to the continuous DoE with PLS to reduce the number of the hyperparameters and compute an estimation of the search space  $\Omega_f$ ;
3. Build the acquisition function  $\alpha_f^{\text{reg}}(\mathbf{x})$ ;
4. Maximize the acquisition function within the feasible domain  $\Omega_f$ :

$$\mathbf{x}_t := \arg \max_{\mathbf{x} \in \Omega_f} \alpha_f^{\text{reg}}(\mathbf{x})$$

5. Add  $\mathbf{x}_t, \mathbf{f}(\mathbf{x}_t)$  and  $g_1(\mathbf{x}_t), \dots, g_m(\mathbf{x}_t)$  to the DoE  $\mathcal{D}_{t+1}$ . Increment  $t$ ;

**end**

**Post-process:** Use the final database to build GP models (for  $f_i(\mathbf{x})$  and  $g_j(\mathbf{x})$ ) and apply NSGA-II algorithm to construct the PF ;

**Outputs:** The PF database and the predicted PF;

---

## 2.4 Optimizers access

In AGILE 4.0 project, a Process Integration and Design Optimization (PIDO) component, also known as RCE [40], is required to execute disciplinary workflows. This executable workflow integrates the different disciplinary tools, which are treated as black-boxes: only inputs and outputs are relevant for the component and the codes are not exposed. The disciplinary tools are installed locally or in servers. Different types of MDO problems can be executed in the component, as basic multidisciplinary analysis, DOEs

---

<sup>3</sup><https://pymoo.org/>

or full optimizations. Another key aspect of the collaborative workflow derives from the collaborative design process. Since the different disciplinary tools can be hosted in different locations, executable workflows are typically distributed simulation workflows across multiple networks in multiple companies. In AGILE 4.0, Brics tool [41, 42] is used to provide the mechanism for interconnecting PIDO environments and hence multi-partner and distributed collaborative simulation workflows. By providing data encryption services, Brics supports protection of the data to be exchanged to and from the central data server in which the inputs/outputs (I/O) file is transferred between the disciplinary tools. In the project, I/O file relies on Common Parametric Aircraft Configuration Schema (CPACS) [43]. CPACS is an open-source, XML-based common language for the exchange of product data. It allows storage of parametric definitions of aircraft geometries as well as analysis results of the individual design disciplines. Therefore, any legacy engineering tool in AGILE 4.0 context is equipped with a wrapper that enables the tool to act on product data and produce analysis results in CPACS format.

As all the application cases have to address multi-objective optimization problems with mixed variables, the optimization algorithms currently available through RCE were not entirely suitable (direct application of genetic algorithms would result in an excessively large number of evaluations). Only a few partners, such as ONERA, NLR, or UNINA, were able to provide gradient-free algorithms capable of solving these types of optimization problems. Therefore three different technical approaches have been applied to connect optimizers to the workflows:

- Local optimizer: the first approach consisted in using an adequate optimization algorithm located on the integrator’s site. This means that the integrator of the workflow needs to have also an expertise in optimization. In AGILE 4.0, only UNINA could apply this approach with JPAD algorithm as illustrated in Section 3.1.
- Remote optimizer through Brics call: the second approach consisted in considering an MDA workflow as a whole that can be called remotely like any design competence from an RCE specific workflow located in the optimizer owner’s premises. This approach was successfully applied in multiple application cases and was retained for the application case presented in Section 3.3.
- Remote access through WhatsOpt: the third approach applied the optimization process not on the design workflow but directly on the architectural choices that is connected to the RCE workflow. In that case, the optimizer is made accessible “as a service” through WhatsOpt [44], a web application supporting MDO collaborative activities. This approach was successfully demonstrated in the application case illustrated in Section 3.2.

### 3 APPLICATIONS

#### 3.1 Airframe upgrade design

This first application case is part of the Upgrade driven stream and aims at accounting retrofitting options in the aircraft MDO workflow. The AGILE 4.0 MDO framework is used to design various retrofitting concepts, including engine replacement, OBS electrification, and winglet installation. Each solution has different impacts on aircraft performance, such as fuel consumption, direct operating cost (DOC), air and noise emissions, as well

as different retrofitting costs. The objective is to be able to select the optimal retrofitting strategy by trading DOC and emissions against retrofitting costs.

For this specific application case, a classical Multidisciplinary Feasible (MDF) approach is being utilized. This type of architecture involves an optimizer driving an MDA workflow, where a complete design is executed in each iteration. The MDA components comprising the workflow can be observed in Figure 2. The workflow begins with an engine module, which provides key engine characteristics based on the given engine ByPass Ratio (BPR) input. This module relies on GASTurb<sup>4</sup> computation. The aerodynamics branch calculates values for both low-speed and high-speed conditions. To account for high-fidelity results in high-speed conditions obtained through CFD analysis, a RSM has been developed. The tool also considers engine geometry and position as inputs. The OBS discipline, named ASTRID [45], is responsible for sizing the overall on-board system (OBS). It determines the masses and the hydraulic, pneumatic, and electric power required by each system during different phases of the mission profile. It also calculates the power-off-takes (secondary power computation) to consider their impact on engine fuel flow. In this study, four different OBS architecture electrification levels are considered, ranging from CONV (conventional) to AEA (all electric aircraft) with two intermediate architectures, namely MEA1 and MAE2 (more electric aircraft 1 and 2). A more detailed description of these architectures can be found in [46]. The performance tool uses a simulation-based approach to compute ground and in-flight performance. It calculates overall mission profile, performance, fuel consumption, flight time, and gaseous emissions. The structural competence relies on a surrogate model based on a high-fidelity structural analysis DOE. This tool optimizes the wing mass by adjusting lamination parameters and thicknesses while ensuring aeroelastic stability, angle-of-attack, strength, and buckling constraints are met. The last three tools described are executed in a converger loop, with iterations performed until convergence is achieved on the maximum take-off weight. The noise competence calculates noise emissions at the certification points in accordance with FAR36 and ICAO Annex 16 regulations. It also provides the noise margin with respect to the certification limits. The cost competence evaluates both recurring and non-recurring costs, including aircraft price and direct operating costs. It also estimates the development, operation, and equipment costs associated with a retrofitting activity. Additionally, it computes cost savings (a component of direct operating costs) resulting from reductions in fuel consumption, maintenance costs, and emission taxes. The cost methodology is explained in [47]. Further details about the MDA workflow presented here can be found in [48].

This application case includes the integration of costly evaluations required to compute the wing structural weight and the high-fidelity aerodynamic effect of the relative nacelle/wing location. To mitigate the cost and computational time, for each design competence, a RSM was built in an off-line process with SMT [49]. A database computed by PROTEUS [50], a tool developed by TUD, is used to calculate the wing's structural mass. The RSM utilizes a database of 56 points (to achieve more than 10 points per dimension) obtained through a Latin Hypercube Sampling (LHS) approach. Additionally, an RSM has been developed based on CFD analysis provided by CFSE to consider the aerodynamic effects of both the pylon and the installed engine on the drag coefficient. The RSM for this purpose utilizes a database of 90 points obtained through an LHS approach. For this RSM, the initial database was smaller (around 10 points per dimension)

---

<sup>4</sup><https://www.gasturb.de/> GasTurb



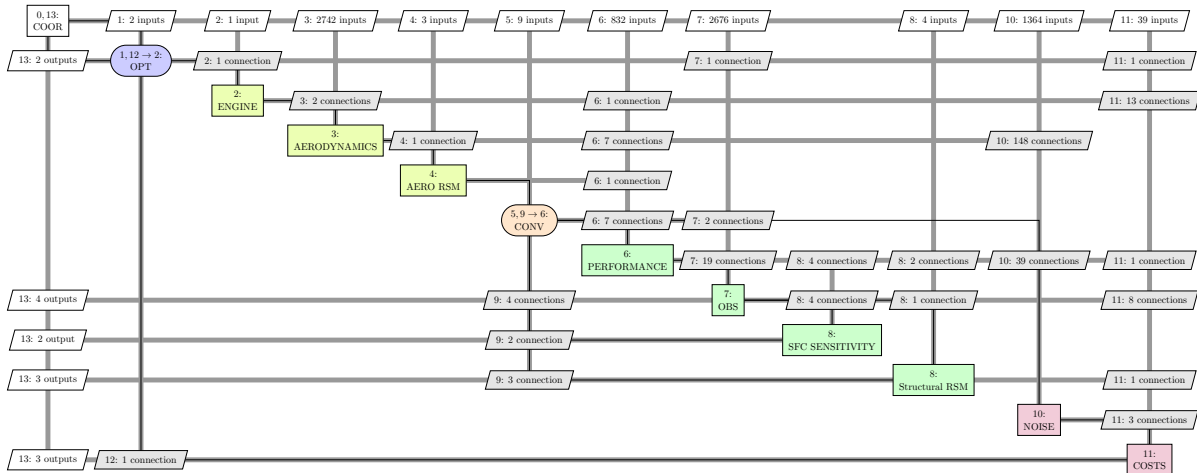


Figure 3: XDSM optimization for airframe upgrade design.

baseline aircraft during a typical mission. The fourth objective is the maximum specific air range (SAR) achievable in cruise conditions and weight, representing the aircraft’s flight efficiency and fuel consumption. Four design variables are considered. One variable is categorical, representing four different OBS architectures that range from conventional to all-electric configurations. The other three variables are continuous, including the engine BPR and the engine  $X$  and  $Z$  locations with respect to the wing attachment points. Figure 4 represents the system under analysis, illustrating the components that can be retrofitted and are affected by the design variables considered in the problem. Lastly, four constraints have been selected. The retrofit solution must maintain at least the same maximum take-off weight and the same take-off and landing distances as the baseline configuration. Additionally, a reduction of 6  $dB$  in cumulative noise emissions must be achieved compared to the original aircraft, leading to a maximum cumulative noise equal to 263  $dB$ .



(a) Engines and OBS are highlighted. (b) CFD simulation of nacelle location effects.

Figure 4: Reference regional jet aircraft.

The following are the results of the optimization process explained earlier. Figure 5 displays the DOE points and the PF obtained using both approaches. The workflow depicted in Figure 2 generated a total of 108 points in the DOE, represented by orange dots in Figure 5. On the same figure are highlighted, as red crosses, the points satisfying the optimization constraints for both DOE (45 feasible points out of 108 full factorial

Table 1: Definition of the airframe upgrade design problem.

|                 | Function/variable                                | Nature | Quantity | Range                       |
|-----------------|--|--------|----------|-----------------------------|
| Minimize        | Maximum Take-Off Weight                          | cont   | 1        |                             |
| Minimize        | Cumulative Emission Index                        | cont   | 1        |                             |
| Minimize        | Cost - savings                                   | cont   | 1        |                             |
| Maximize        | Max Cruise Specific Air Range                    | cont   | 1        |                             |
|                 | <b>Total objectives</b>                          |        | <b>4</b> |                             |
| with respect to | OBS architecture                                 | cat    | 4 levels | [CONV, MEA1, MEA2, AEA]     |
|                 | Total categorical variables                      |        | 1        |                             |
|                 | Engine Bypass Ratio                              | cont   | 1        | [9, 15] (-)                 |
|                 | Engine <i>X</i>                                  | cont   | 1        | [-0.98, -0.80] ( <i>m</i> ) |
|                 | Engine <i>Z</i>                                  | cont   | 1        | [-0.39, -0.21] ( <i>m</i> ) |
|                 | Total continuous variables                       |        | 3        |                             |
|                 | <b>Total relaxed variables</b>                   |        | <b>7</b> |                             |
| subject to      | Maximum Take-Off Weight $\leq 39058.5$ <i>kg</i> |        | 1        |                             |
|                 | Take-Off Field Length $\leq 1500$ <i>m</i>       |        | 1        |                             |
|                 | Landing Field Length $\leq 1400$ <i>m</i>        |        | 1        |                             |
|                 | Cumulative Noise $\leq 263$ <i>dB</i>            |        | 1        |                             |
|                 | <b>Total constraints</b>                         |        | <b>4</b> |                             |

DOE points) and JPAD optimization.

Regarding the results obtained with SEGOMOE (the workflow shown in Figure 3), green crosses represent all the evaluated points fulfilling the optimization constraints (among the 81 points computed with this approach). The global PF, indicated by the 15 blue circles, was derived from combining both databases obtained by JPAD optimizer and SEGOMOE. This front is composed by 4 points from JPAD optimizer and 11 points from SEGOMOE optimizer. Among the points on the PF, the best solution depends on the designer’s preferences. For instance, the point with the minimum difference between costs and savings (the lower one) is always an optimum choice in terms of economic profit. However, it may not be the optimal solution concerning other variables such as CEI and SAR. In fact, several solutions that offer significant benefits in terms of emissions, SAR, and MTOW may be disregarded in the pursuit of achieving lower costs. Similarly, there are points that offer maximum benefits in terms of other variables but may not be desirable due to their higher costs. When examining opposite solutions on the PF, the selection of engine BPR and OBS architecture can vary depending on the performance objective to be maximized: for example, a BPR of 9.0 with a conventional OBS architecture (CONV) or a BPR of 15.0 with an advanced electric aircraft (AEA) OBS architecture. Opting for lower BPR and a moderate level of electrification (representing state-of-the-art technologies) helps reduce retrofitting costs while achieving moderate performance improvements. Conversely, increasing the level of retrofitting by adopting advanced engine technology and full OBS electrification (beyond state-of-the-art) can significantly enhance overall performance metrics like SAR and CEI, but at the cost of higher retrofitting expenses. For instance, considering a higher BPR and advanced electrification, emissions are reduced (CEI decreases from 1 to 0.86), MTOW slightly decreases (around -3.1% compared to the baseline), and SAR increases (around +25% compared to the baseline). As a result, the costs minus savings increase to 0.44 million euros per year, indicating a loss for an airline operating the aircraft under the considered scenarios.

In this application, two approaches were compared, both providing solutions located on the PF. For this case, SBO with JPAD appears to be a bit more costly - in terms of

number of function evaluations - than BO with SEGOMOE and ends up with a smaller number of feasible solutions. Nevertheless, this approach is less sensitive to a change in optimization problem (in terms of objectives and constraints) as it relies on an initial database whereas SEGOMOE enrichment process is tailored for a unique optimization.

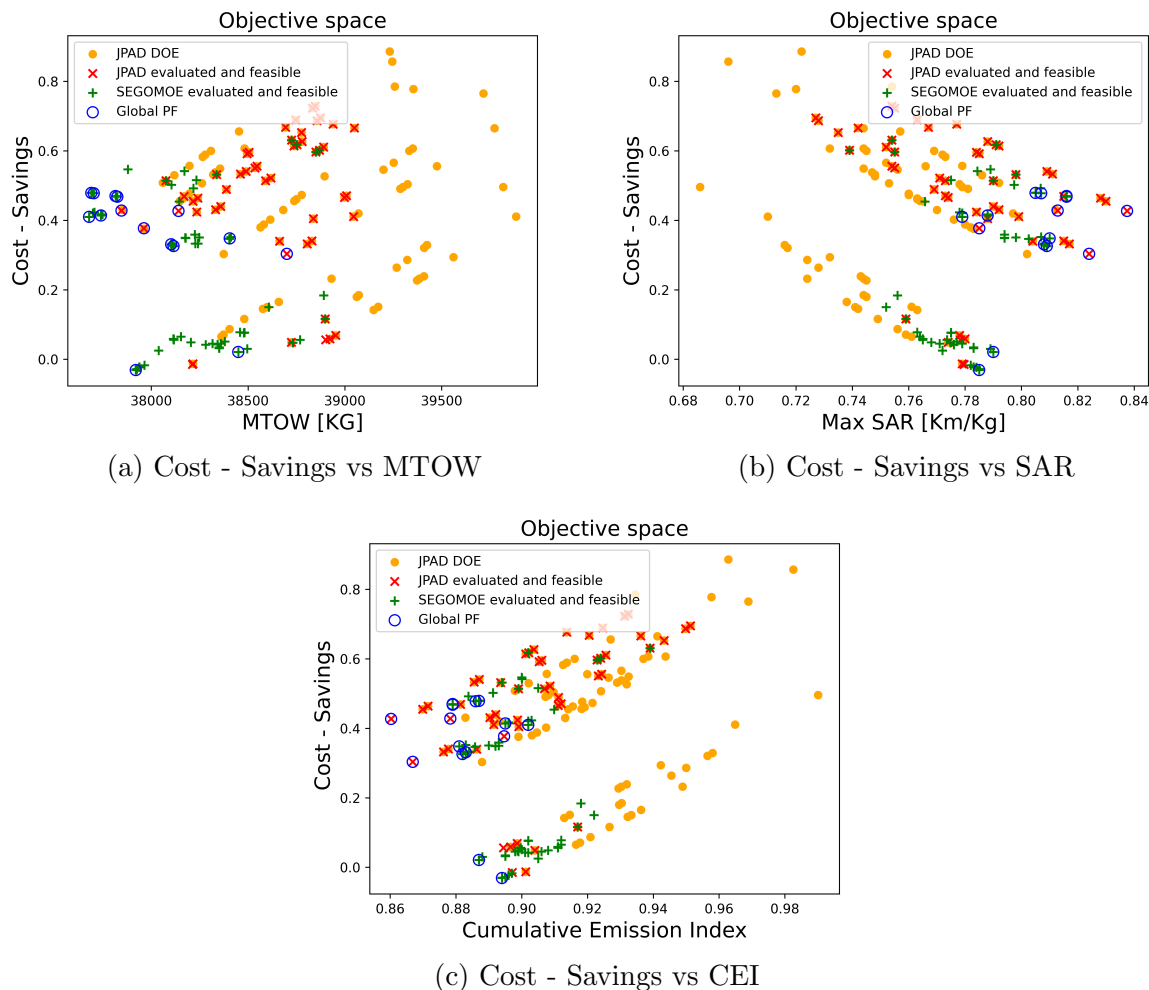


Figure 5: Different PF using four objectives: 108 DOE points for JPAD (orange dot), 15 points (blue circle) on the final PF combining JPAD (red cross) and SEGOMOE (green cross) databases.

### 3.2 Design of an aircraft family

This application case is part of the upgrade-driven stream and aims at designing a family of three business jet aircraft considering commonality options in the MDO workflow [5]. The commonality options enable sharing one or more aircraft components between the aircraft in the family: by sharing components, design, certification, production, and maintenance costs are reduced. However, operating costs might be increased due to the use of components not designed for the typical flight conditions.

Shareable components include wings, empennage (horizontal and vertical tail), engines, on-board systems (OBS), and landing gears, shown in Figure 6. Next to the commonality choices there are three design variables per wing: leading edge sweep, rear span loca-

tion (determining flap size), and thickness-to-chord ratio. These three design variables per wing are inactive if the associated wing is a shared wing. Each aircraft family is optimized for two objectives: direct operating costs (DOC), representing the impact on performance, and Original Equipment Manufacturer (OEM) non-recurring costs, representing the potential benefit of commonality. Table 2 presents the optimization problem in more details.

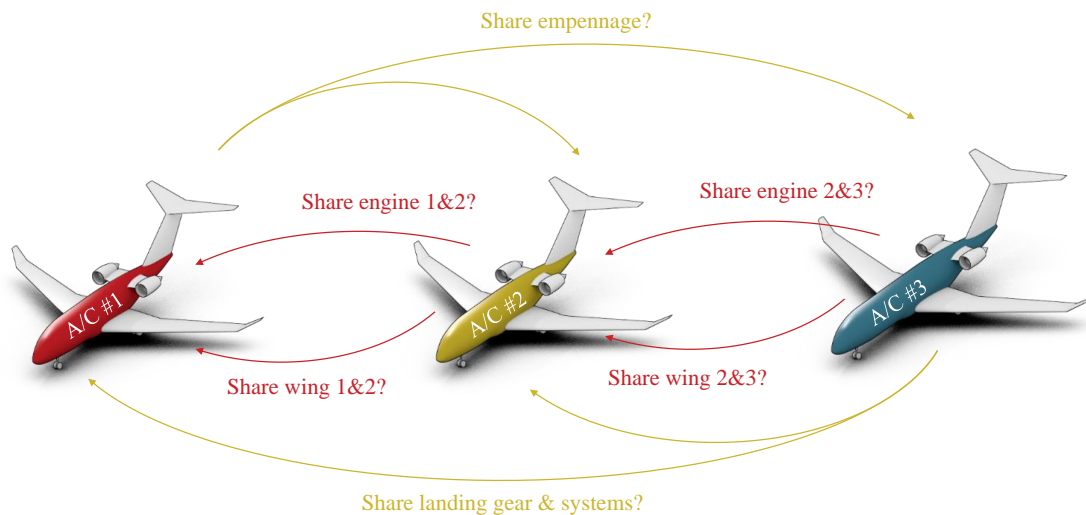


Figure 6: Visualization of the aircraft family including commonality sharing decisions.

The design space is modeled using the Architecture Design Space Graph (ADSG) method [51] implemented in the ADORE tool [52], both developed during the project. MultiLinQ is used to couple the generate architecture instances to the central data schema used in the MDO workflow [5]. The MDO workflow consists of two levels: an aircraft-level workflow (shown in Figure 7) applying commonality decisions and sizing one aircraft at a time, and a family-level workflow converging the three aircraft-level workflows and performing cost calculation on the family-level. For a more detailed overview of the design space model, coupling to the MDO workflow and implementation of the MDO workflow for the family design application case, the reader is referred to [3, 5].

For this application case SEGOMOE is used as optimization algorithm due to the need to handle hierarchical, mixed-discrete design variables, and the need to minimize the number of function evaluations as one family evaluation can take up to 2 hours. Hierarchical design variables were handled using the imputation method, where inactive variables are replaced by the mean value to prevent duplicate design vectors [27]. SEGOMOE was accessed through an ask-tell API implemented in WhatsOpt running on a server at ONERA's premises as described in Section 2.4. For a more detailed description of how SEGOMOE was coupled to ADORE and the MDO workflow, the reader is referred to [3].

First, a 50-point DOE was executed to create the initial database of design points for SEGOMOE and to verify the correct behavior of the MDO workflow. Then, SEGOMOE was used to generate an addition of 18 infill points to explore the design space and extend the Pareto front. Figure 8 shows the main Pareto front, with infill points shown in red. Several example families are shown with colors indicating the originating family member (colors are defined in Figure 6). As can be seen, when no components are shared



Table 2: Definition of the aircraft family design problem.

|                 | Function/variable  | Nature | Quantity  | Range                 |
|-----------------|--|--------|-----------|-----------------------|
| Minimize        | Direct Operating Costs                                       | cont   | 1         |                       |
| Minimize        | OEM Non-Recurring Costs                                      | cont   | 1         |                       |
|                 | <b>Total objectives:</b>                                     |        | <b>2</b>  |                       |
| with respect to | engine commonality 1&2                                       | cat    | 2 levels  | sharing yes/no        |
|                 | engine commonality 2&3                                       | cat    | 2 levels  | sharing yes/no        |
|                 | wing commonality 1&2   | cat    | 2 levels  | sharing yes/no        |
|                 | wing commonality 2&3   | cat    | 2 levels  | sharing yes/no        |
|                 | landing gear commonality 1&2                                 | cat    | 2 levels  | sharing yes/no        |
|                 | landing gear commonality 2&3                                 | cat    | 2 levels  | sharing yes/no        |
|                 | OBS commonality 1&2  | cat    | 2 levels  | sharing yes/no        |
|                 | OBS commonality 2&3  | cat    | 2 levels  | sharing yes/no        |
|                 | empennage commonality 1&3                                    | cat    | 2 levels  | sharing yes/no        |
|                 | empennage commonality 3&2                                    | cat    | 2 levels  | sharing yes/no        |
|                 | Total categorical variables:                                 |        | 10        |                       |
|                 | Leading edge sweep*  | cont   | 3         | [30.0, 42.0] (deg)    |
|                 | Rear span location*  | cont   | 3         | [0.72, 0.82] (%chord) |
|                 | Wing thickness/cord ratio*                                   | cont   | 3         | [0.06, 0.11] (-)      |
|                 | * only active if the corresponding wing is <b>not</b> shared |        |           |                       |
|                 | Total continuous variables:                                  |        | 9         |                       |
|                 | <b>Total relaxed variables:</b>                              |        | <b>29</b> |                       |
| subject to      | Balanced Field Length $\leq 1524$ m                          |        | 1         |                       |
|                 | Landing Field Length $\leq 762$ m                            |        | 1         |                       |
|                 | <b>Total constraints:</b>                                    |        | <b>2</b>  |                       |

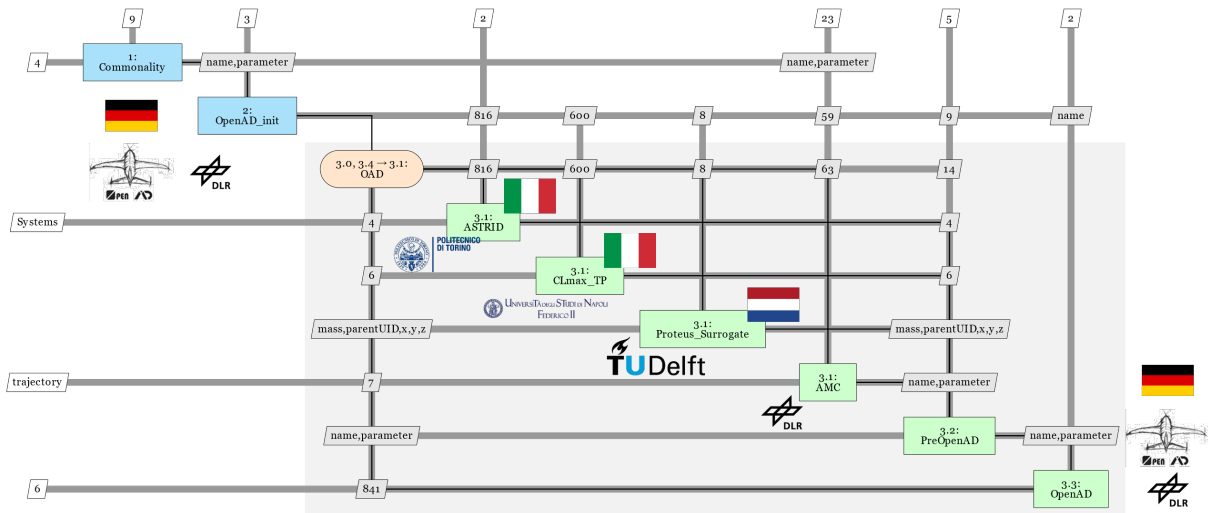


Figure 7: MDA for family aircraft design.

(family #1) the OEM non-recurring costs are high whereas the operating costs are low, because all aircraft components are used at the operating points they are designed for. Introducing more component sharing reduces OEM NRC: family #62 achieves the lowest operating costs at a reduced NRC. Family #54 represents the opposite extreme: the lowest NRC coupled with the highest operating costs, as achieved by a high number of shared

components. Family #30 represents a good compromise, at a reduced NRC compared to family #62, with only a moderate increase in DOC.

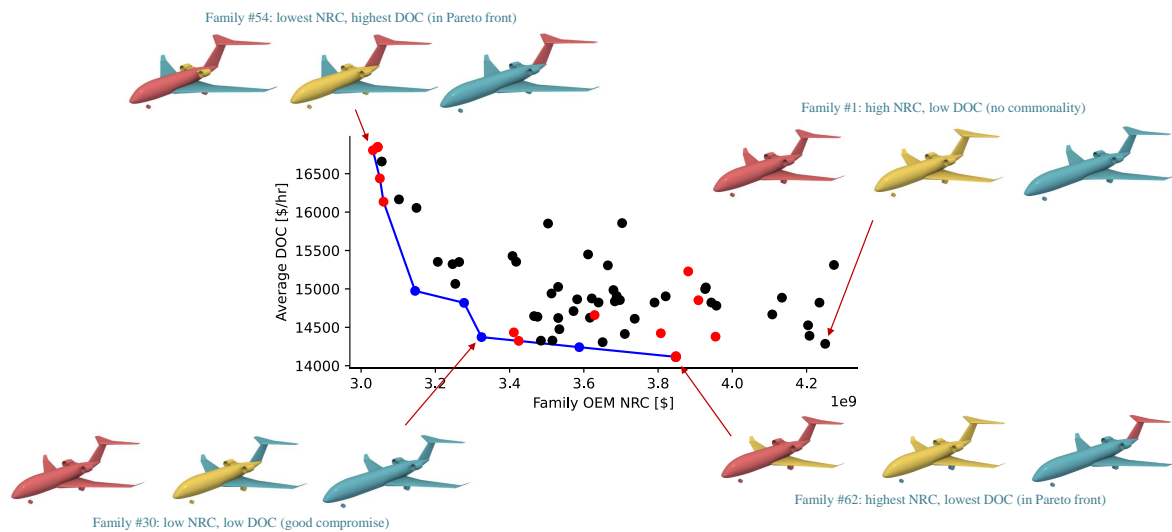


Figure 8: Results of the aircraft family design, showing the Pareto front for minimization of Original Equipment Manufacturer Non-Recurring Cost and Direct Operating Cost and several families. The initial DOE is given by the 50 black dots. Infill points generated by SEGOMOE are shown in red. Colors correspond to the originating aircraft as defined in Figure 6.

### 3.3 Aircraft design with supply chain, performance and manufacturing process

This application case is part of the Production stream and aims at accounting the supply chain aspects in the aircraft MDO workflow, in addition with design and manufacturing aspects. The application focuses on the design of a horizontal tail plane (HTP) made by different materials and produced by different supply chains, combination of enterprises. The different solutions are characterized by different impact on aircraft performance (e.g. mass, aerodynamic efficiency), different manufacturing processes and different combinations of supply chains.

The objective is to have the ability to select the best HTP to produce by trading performance vs manufacturing costs vs supply chain performance. In this study, HTP structure breakdown includes skin, stringers, spars and ribs that impact the three domains:

- the manufacturing domain deals with the choices of materials, manufacturing and assembly processes feasible for the aircraft components. For each combination of materials, a Technology Factor (TF), a dimensionless number ranging from 0 to 1, is produced.
- the design domain deals with the estimation of the aircraft performance. The Technology Factor is considered in the evaluation as it quantifies the impact that materials, manufacturing and assembly processes, have on the mass and drag of the HTP and, consequently, on the fuel consumption of the vehicle in cruise. The estimation of the fuel consumption is the main output of the design domain.

- the supply chain domain estimates the production performance by characterizing the enterprises involved in the aircraft production and assembly. Here, each component of the HTP is characterized by a Production Quantity (PQ) that indicates the percentage of this component that each enterprise has to produce. The cost, time, quality and risk are the main outputs of this domain.

For more details regarding this study, the reader can refer to [6, 8, 53]. Figure 9 proposes a sketch of the interaction between the three domains to build the MDA.

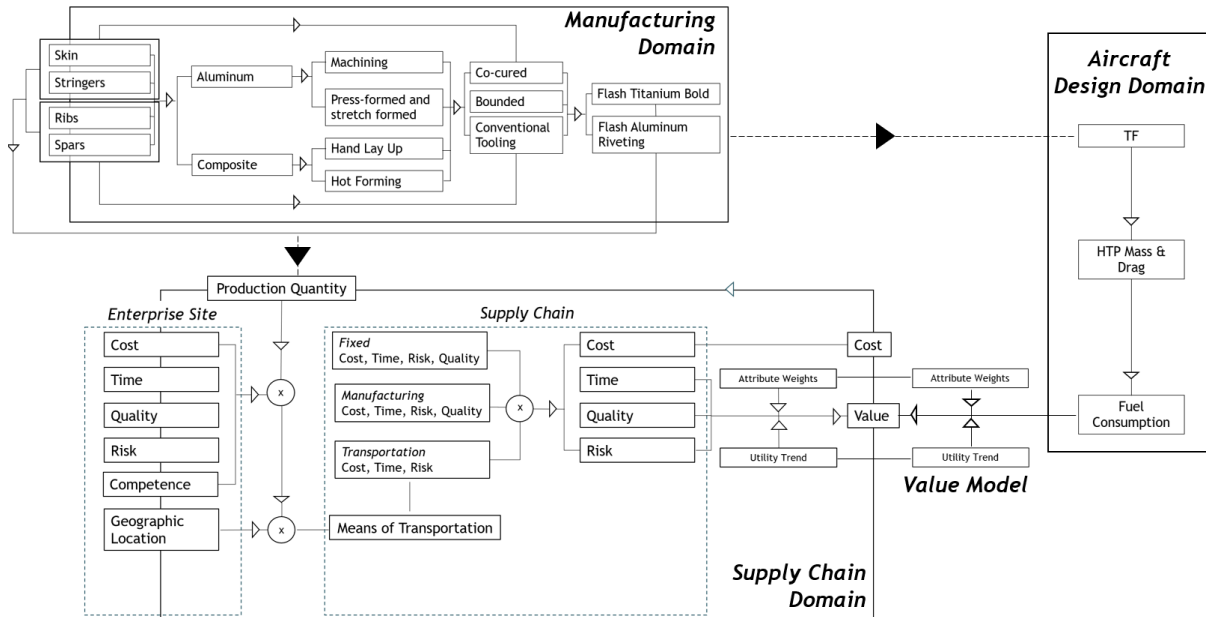


Figure 9: MDA including Manufacturing, Supply Chain and Aircraft Design Variables.

In a first step, the focus was on the supply chain aspect in order to define the formalization, the settings of the optimizer, and the settings for remote calls. During this step, the initial problems are used to calculate the full enumeration of solutions in order to compare the performance of the optimizer with the true Pareto Front of the database. Several increasingly complex problems were tackled, starting with a reduced number of components (for instance only skins and stringers), limited PQ numbers, and limited production sites. Eventually an optimization problem considering all the HTP components and all possible supply chain sites was successfully tackled.

In a second step, the activities focus on integrating the manufacturing options in the overall optimization problem. It requires to add more variables, namely the choice of materials and their associated manufacturing processes for each HTP component. In addition, two constraints have now to be considered:

- the material incompatibility, that aims at discarding all the unfeasible combinations of material and manufacturing processes for the HTP components (for instance, an HTP can not be made using both skin in Aluminium - Machining and ribs in Thermoset - Hand Lay Up);
- the minimal supply chain site competence that checks that no combination of skin, spar, stringer and ribs (depending of their material and manufacturing process) can be produced on a site with a competence of 0.

Table 3 presents the optimization problem in more details. Five objectives are considered, Cost, Risk, Time and Quality that are related to the supply chain domain and Fuel Consumption that is connected to the overall aircraft design and manufacturing domain. The variables represent all the possible choices of the HTP components characteristics and are treated as categorical variables in the optimization problem. Two types of variables are considered for each HTP component:

- the choice of production and assembly site. Here the PQ have been frozen to 1, but all the possible production sites are available. Each component can therefore be produced among 21 different sites, either belonging to the OEM (Original Equipment Manufacturer) or to a supplier or sub-suppliers.
- the choice of materials and their associated manufacturing processes. Depending of the component, several combinations of material and manufacturing process can be selected (up to 6 for skin). For instance, choices could be Aluminium - Machining, Thermoset - Hand Lay Up or Thermoset - Infusion Process.

Finally, the two constraints described earlier are included.

Table 3: Definition of the complete problem.

|                 | Function/variable                  | Nature | Quantity   | Range             |
|-----------------|------------------------------------|--------|------------|-------------------|
| Minimize        | Cost                               | cont   | 1          |                   |
| Minimize        | Risk                               | cont   | 1          |                   |
| Minimize        | Time                               | cont   | 1          |                   |
| Minimize        | Fuel Burn                          | cont   | 1          |                   |
| Maximize        | Quality                            | cont   | 1          |                   |
|                 | <b>Total objectives</b>            |        | <b>5</b>   |                   |
| with respect to | skin prod. location                | cat    | 21 levels  | geographic sites  |
|                 | spar production location           | cat    | 21 levels  | geographic sites  |
|                 | stringer production location       | cat    | 21 levels  | geographic sites  |
|                 | rib production location            | cat    | 21 levels  | geographic sites  |
|                 | skin Material & Manuf. process     | cat    | 6 levels   | Alu-Machining,... |
|                 | spar Material & Manuf. process     | cat    | 5 levels   | Alu-Machining,... |
|                 | stringer Material & Manuf. process | cat    | 4 levels   | Alu-Machining,... |
|                 | rib Material & Manuf. process      | cat    | 5 levels   | Alu-Machining,... |
|                 | Total categorical variables        |        | 8          |                   |
|                 | <b>Total relaxed variables</b>     |        | <b>104</b> |                   |
| subject to      | Material incompatibility           |        | 1          |                   |
|                 | Supply chain sites                 |        | 1          |                   |
|                 | <b>Total constraints</b>           |        | <b>2</b>   |                   |

Regarding optimization process, a 300-point DOE was executed to create the initial database of design points for SEGOMOE and to verify the correct behavior of the MDO workflow. Then, SEGOMOE was used to generate 450 additional infill points to explore the design space and extend the Pareto front. The main results of the optimization process are depicted in Figure 10 and Figure 11. Figure 10 presents the PF database obtained at the end of the optimization run with 450 iterations. The initial DOE points are represented by orange dots and can provide information concerning the workflow

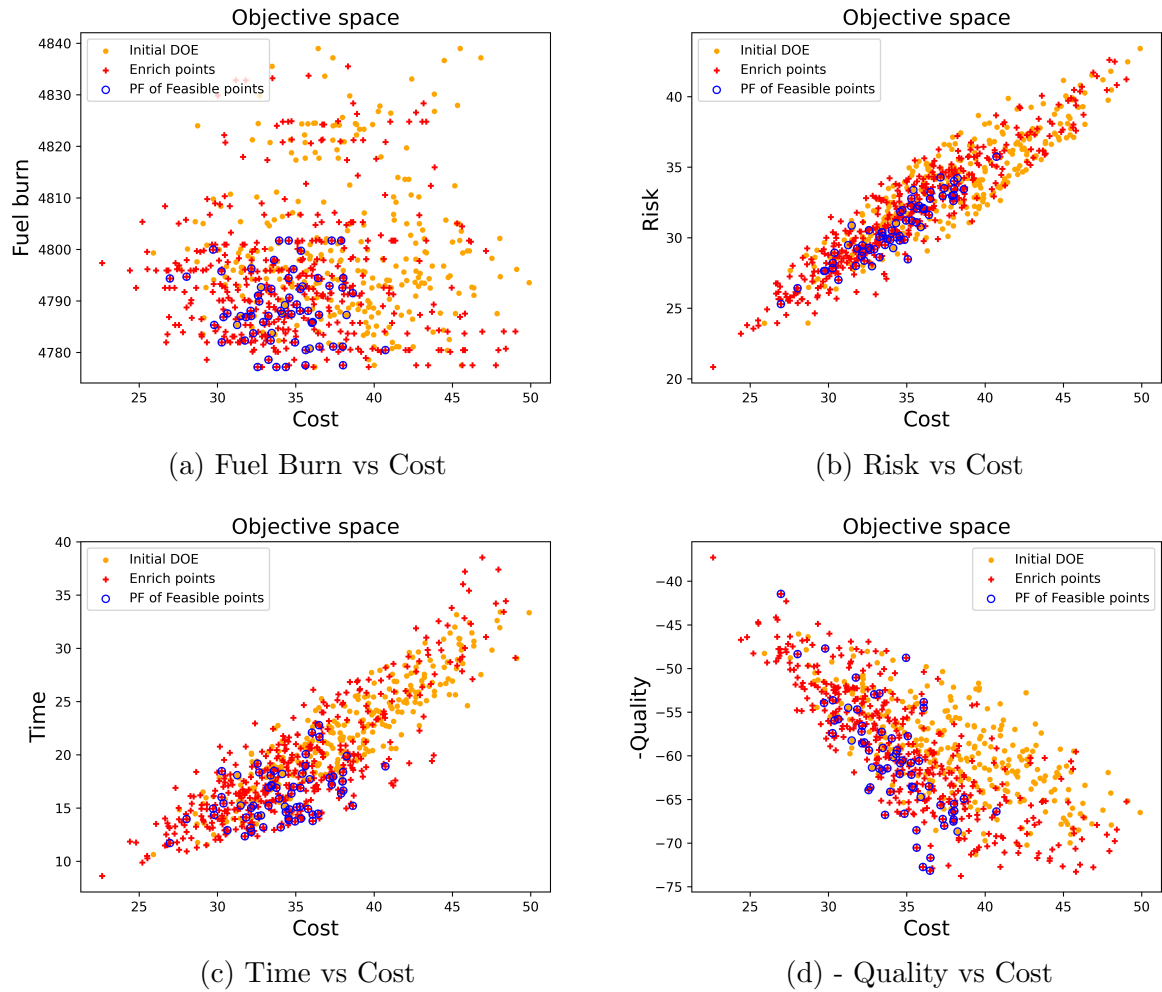


Figure 10: Database for five objectives: initial DOE of 300 points (orange dots), 450 enrichment points (red crosses) and, from the complete database (750 points) 65 feasible points on the PF are identified by the blue circles.

behavior. The enrichment process of SEGOMOE added the 450 red crosses points that extended the database in direction of the Pareto Front (as it can be seen, for instance, on the -Quality vs Cost Figure 10d). As the optimization is considering two constraints, not all the database points are feasible and, at the end of the process, 65 points are identified on the PF (all highlighted with blue circles).

As indicated in Algorithm 1, a post-processing step is then performed applying NSGA-II algorithm to construct the predicted PF (using the GP models built upon the database). Figure 11 illustrates the results of this step with the PF database still depicted by blue circles. The predicted PF is represented by green diamonds, consisting of 33 candidates. It can be observed that both databases appear intermingled, indicating that the PF database already exhibits promising performance. Nevertheless, as an additional step, all the points belonging to the predicted PF have been evaluated with the workflow and the outputs are depicted in black diamonds. Two pieces of information can be derived from this evaluated PF. Firstly, the evaluated candidates closely align with the predicted values (with the exception of a few), indicating the accuracy of the GP models, despite being built upon 750 points. Secondly, the PF database appears to maintain its good

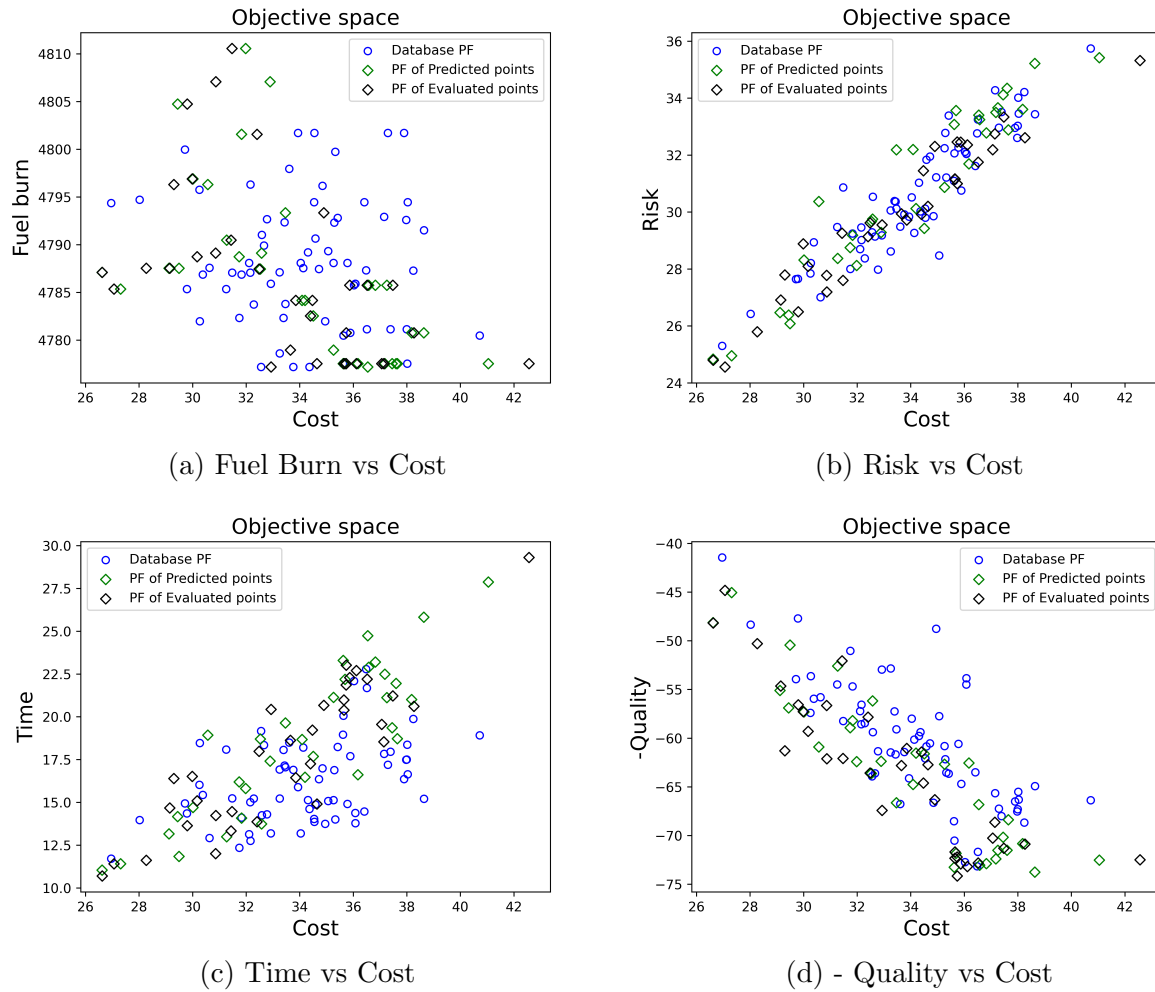


Figure 11: Comparison of PF (predicted and evaluated) with five objectives: 65 points from the PF database (blue circles), 33 points from the predicted PF (green diamonds) or evaluated with the workflow (black diamonds). The predicted and evaluated points are quite close to each other and complementary to the PF database.

performance, characterized by its proximity to the evaluated PF. To verify this, both the PF database and evaluated PF were sorted, revealing that 52 points (out of 65) from the PF database and 26 points (out of 33) from the evaluated PF contribute to the final combined PF. As expected, the post-processing step of SEGOMOE successfully populates the PF identified by the algorithm, thereby enriching the available choices for the application stakeholders.

#### 4 CONCLUSIONS AND PERSPECTIVES

Within the frame of AGILE 4.0 project, multi-objective problems involving mixed integer variables have been successfully solved using SBO and BO optimizers based on two associated frameworks JPAD Optimizer and SEGOMOE. In terms of function evaluations, the number of calls to the expensive black boxes (objectives and constraints) is still small compared to the number of design variables involved. SEGOMOE has been successfully applied to three realistic problems involving 2 to 5 objective functions with mixed integer

variables. The use of continuous relaxation strategy to handle the categorical variables has been coupled with PLS to reduce the increase of dimension allowing to consider up to 104 choices.

To apply BO to even more complex multidisciplinary processes, several developments need to be addressed. Future work will focus on extending the handling of hierarchical and mixed discrete variables while improving computational efficiency for larger databases. Specifically, with regards to mixed integer variables, the intention is to integrate recent advancements done within the SMT toolbox concerning mixed correlation kernels [9] or hierarchical kernels [49] for Gaussian processes into BO and validate their effectiveness. These studies will be conducted within the context of the Horizon Europe COLOSSUS project <sup>5</sup>. The COLOSSUS project aims to develop a system-of-systems design methodology that facilitates the integrated optimization of aircraft, operations, and business models. This methodology will be applied to intermodal transport and wildfire-fighting scenarios.

## ACKNOWLEDGEMENTS

This work is part of the activities of ONERA - ISAE - ENAC joint research group. The research presented in this paper has been performed within the framework of the AGILE 4.0 project (Towards Cyber-physical Collaborative Aircraft Development) and has received funding from the European Union Horizon 2020 Programme under grant agreement n°815122. Some methodology developments have been performed in the framework of the COLOSSUS project (Collaborative System of Systems Exploration of Aviation Products, Services and Business Models) and has received funding from the European Union Horizon Programme under grant agreement n°101097120. The authors are also grateful to the partners of the AGILE 4.0 consortium for their contribution and feedback.

## REFERENCES

- [1] A. Chan, A. F. Pires, and T. Polacsek. Trying to elicit and assign goals to the right actors. In *Conceptual Modeling: 41st International Conference, ER 2022*, 2022.
- [2] M. Fioriti, P. D. Vecchia, and G. Donelli. Effect of progressive integration of on-board systems design discipline in an mda framework for aircraft design with different level of systems electrification. *Aerospace*, 9(3), 2022.
- [3] J. Bussemaker, L. Boggero, and B. Nagel. The AGILE4.0 project: MBSE to support cyber-physical collaborative aircraft development. In *33rd Annual INCOSE International Symposium*, Honolulu, HI, USA, 2023. To be published.
- [4] J. R. R. A. Martins and A. B. Lambe. Multidisciplinary design optimization: A survey of architectures. *AIAA Journal*, 51(9):2049–2075, 2013. doi:10.2514/1.J051895.
- [5] J. H. Bussemaker, P. D. Ciampa, J. Singh, M. Fioriti, C. C. D. L. Hoz, Z. Wang, D. Peeters, P. Hansmann, P. D. Vecchia, and M. Mandorino. Collaborative design of a business jet family using the AGILE 4.0 MBSE environment. In *AIAA AVIATION 2022 Forum*, 2022. doi:10.2514/6.2022-3934.

---

<sup>5</sup><https://colossus-sos-project.eu/>

- [6] G. Donelli, J. M. Mello, F. I. Odaguil, T. van der Laan, T. Lefebvre, N. Bartoli, L. Boggera, and B. Nagel. Value-driven systems engineering approach addressing manufacturing, supply-chain and aircraft design in the decision-making process. In *INCOSE 33rd Annual International Symposium 2023*, 2023.
- [7] A. Chan, A. Fernandes Pires, T. Polacsek, and S. Roussel. The aircraft and its manufacturing system: From early requirements to global design. In *Advanced Information Systems Engineering: 34th International Conference, CAiSE 2022, Leuven, Belgium, June 6–10, 2022, Proceedings*, pages 164–179. Springer, 2022.
- [8] U. Merola, G. Donelli, T. Lefebvre, N. Bartoli, J. M. Mello, F. I. Odaguil, T. van der Laan, and B. Nagel. Value-driven optimization campaign addressing manufacturing, supply chain and overall aircraft design domains in the early development stage. In *INCOSE 33rd Annual International Symposium 2023*, 2023.
- [9] P. Saves, Y. Diouane, N. Bartoli, T. Lefebvre, and J. Morlier. A mixed-categorical correlation kernel for Gaussian process. *arXiv preprint*, 2022. doi:<https://doi.org/10.48550/arXiv.2211.08262>.
- [10] R. Grapin, Y. Diouane, J. Morlier, N. Bartoli, T. Lefebvre, P. Saves, and J. Bussemaker. Regularized infill criteria for multi-objective bayesian optimization with application to aircraft design. In *AIAA AVIATION 2022*, 2022.
- [11] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.
- [12] R. Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.
- [13] F. Nicolosi, A. De Marco, L. Attanasio, and P. D. Vecchia. Development of a java-based framework for aircraft preliminary design and optimization. *Journal of Aerospace Information Systems*, 13(6):234–242, 2016.
- [14] V. Trifari, M. Ruocco, V. Cusati, F. Nicolosi, and A. De Marco. Java framework for parametric aircraft design–ground performance. *Aircraft Engineering and Aerospace Technology*, 89(4):599–608, 2017.
- [15] J. B. Kollat and P. M. Reed. Comparison of multi-objective evolutionary algorithms for long-term monitoring design. In *Impacts of Global Climate Change*, pages 1–11. 2005.
- [16] M. Sasena. *Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximations*. PhD thesis, University of Michigan, 2002.
- [17] C. E. Rasmussen and C. K. Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.
- [18] D. 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.



- [19] D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier, and M. Samuelides. Surrogate modeling approximation using a mixture of experts based on EM joint estimation. *Structural and Multidisciplinary Optimization*, 43(2):243–259, 2011. ISSN 1615-147X. URL <http://dx.doi.org/10.1007/s00158-010-0554-2>. 10.1007/s00158-010-0554-2.
- [20] R. P. Liem, C. A. Mader, and J. R. R. A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis. *Aerospace Science and Technology*, 43:126–151, 2015. doi:10.1016/j.ast.2015.02.019.
- [21] M. A. Bouhleb, N. Bartoli, A. Otsmane, and J. Morlier. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53(5):935–952, 2016. ISSN 1615-1488. doi:10.1007/s00158-015-1395-9. URL <http://dx.doi.org/10.1007/s00158-015-1395-9>.
- [22] M. A. Bouhleb, N. Bartoli, A. Otsmane, and J. Morlier. An improved approach for estimating the hyperparameters of the kriging model for high-dimensional problems through the partial least squares method. *Mathematical Problems in Engineering*, 2016, 2016.
- [23] R. Priem, H. Gagnon, I. Chittick, S. Dufresne, Y. Diouane, and N. Bartoli. An efficient application of bayesian optimization to an industrial mdo framework for aircraft design. In *AIAA AVIATION 2020 FORUM*, page 3152, 2020.
- [24] P. Saves, N. Bartoli, Y. Diouane, T. Lefebvre, J. Morlier, C. David, E. Nguyen Van, and S. Defoort. Multidisciplinary design optimization with mixed categorical variables for aircraft design. In *AIAA SCITECH 2022 Forum*, page 0082, 2022.
- [25] N. Bartoli, T. Lefebvre, S. Dubreuil, R. Olivanti, R. Priem, N. Bons, J. R. Martins, and J. Morlier. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerospace Science and Technology*, 90:85–102, 2019. doi:<https://doi.org/10.1016/j.ast.2019.03.041>.
- [26] T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, P. D. Vecchia, L. Stingo, F. Nicolosi, K. Anisimov, A. Savelyev, A. Mirzoyan, and A. Isyanov. Enhancing optimization capabilities using the agile collaborative mdo framework with application to wing and nacelle design. *Progress in Aerospace Sciences*, 2020. doi:10.1016/j.paerosci.2020.100649.
- [27] J. H. Bussemaker, N. Bartoli, T. Lefebvre, P. D. Ciampa, and B. Nagel. Effectiveness of surrogate-based optimization algorithms for system architecture optimization. In *AIAA AVIATION 2021 FORUM*, page 3095, 2021. doi:10.2514/6.2021-3095.
- [28] E. C. Garrido-Merchán and D. Hernández-Lobato. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380:20–35, 2020.

- [29] D. Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. Google vizier: A service for black-box optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1487–1495, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3097983.3098043. URL <https://doi.org/10.1145/3097983.3098043>.
- [30] M. A. Bouhlef, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, and J. R. R. A. Martins. A python surrogate modeling framework with derivatives. *Advances in Engineering Software*, 135, 2019. doi:10.1016/j.advengsoft.2019.03.005.
- [31] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2):117–132, 2003.
- [32] M. T. Emmerich, K. C. Giannakoglou, and B. Naujoks. Single-and multiobjective evolutionary optimization assisted by gaussian random field metamodells. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006.
- [33] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21:345–383, 2001.
- [34] A. A. Rahat, R. M. Everson, and J. E. Fieldsend. Alternative infill strategies for expensive multi-objective optimisation. In *Proceedings of the genetic and evolutionary computation conference*, pages 873–880, 2017.
- [35] M. J. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*, pages 51–67. Springer, 1994.
- [36] D. Kraft et al. *A software package for sequential quadratic programming*. DFVLR Obersfaffehofen, Germany, 1988.
- [37] P. E. Gill, W. Murray, and M. A. Saunders. Snopt: An sqp algorithm for large-scale constrained optimization. *SIAM review*, 47(1):99–131, 2005.
- [38] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
- [39] J. Blank and K. Deb. Pymoo: Multi-objective optimization in python. *IEEE Access*, 8:89497–89509, 2020.
- [40] B. Boden, J. Flink, N. Först, R. Mischke, K. Schaffert, A. Weinert, A. Wohlan, and A. Schreiber. Rce: An integration environment for engineering and science. *softwarex* 15, 100759 (2021), 2021.
- [41] E. Baalbergen, E. Moerland, W. Lammen, and P. Ciampa. Methods to support efficient collaboration for competitive aircraft design. In *6th CEAS Aerospace Europe Conference*, 2017.

- [42] E. Baalbergen, J. Vankan, L. Boggero, J. H. Bussemaker, T. Lefebvre, B. Beijer, A.-L. Bruggeman, and M. Mandorino. Advancing cross-organizational collaboration in aircraft development. In *AIAA AVIATION 2022 Forum*, page 4052, 2022.
- [43] M. Alder, E. Moerland, J. Jepsen, and B. Nagel. Recent advances in establishing a common language for aircraft design with cpacs. 2020.
- [44] R. Lafage, S. Defoort, and T. Lefebvre. Whatsopt: a web application for multidisciplinary design analysis and optimization. In *AIAA Aviation 2019 Forum*, page 2990, 2019.
- [45] S. Chiesa, G. A. Di Meo, M. Fioriti, G. Medici, and N. Viola. Astrid–aircraft on board systems sizing and trade-off analysis in initial design. *Research and Education in Aircraft Design–READ 2012*, 2012.
- [46] P. Della Vecchia, L. Stingo, F. Nicolosi, A. De Marco, G. Cerino, P. D. Ciampa, P. S. Prakasha, M. Fioriti, M. Zhang, A. Mirzoyan, et al. Advanced turboprop multidisciplinary design and optimization within agile project. In *2018 Aviation Technology, Integration, and Operations Conference*, page 3205, 2018.
- [47] P. Della Vecchia, M. Mandorino, V. Cusati, and F. Nicolosi. Retrofitting cost modeling in aircraft design. *Aerospace*, 9(7):349, 2022.
- [48] M. Mandorino, P. Della Vecchia, S. Corcione, F. Nicolosi, V. Trifari, G. Cerino, M. Fioriti, C. Cabaleiro De La Hoz, T. Lefebvre, D. Charbonnier, et al. Multidisciplinary design and optimization of regional jet retrofitting activity. In *AIAA AVIATION 2022 Forum*, page 3933, 2022.
- [49] P. Saves, R. Lafage, N. Bartoli, Y. Diouane, J. Bussemaker, T. Lefebvre, J. T. Hwang, J. Morlier, and J. R. R. A. Martins. SMT 2.0: A surrogate modeling toolbox with a focus on hierarchical and mixed variables gaussian processes. *arXiv preprint*, 2023. doi:<https://doi.org/10.48550/arXiv.2305.13998>.
- [50] N. Werter and R. De Breuker. A novel dynamic aeroelastic framework for aeroelastic tailoring and structural optimisation. *Composite Structures*, 158:369–386, 2016.
- [51] J. H. Bussemaker, P. D. Ciampa, and B. Nagel. System architecture design space exploration: An approach to modeling and optimization. In *AIAA Aviation 2020 Forum*, page 3172, 2020. doi:10.2514/6.2020-3172.
- [52] J. Bussemaker, L. Boggero, and P. D. Ciampa. From system architecting to system design and optimization: A link between mbse and mdao. In *INCOSE International Symposium*, volume 32, pages 343–359. Wiley Online Library, 2022. doi:10.1002/iis2.12935.
- [53] G. Donelli, P. D. Ciampa, T. Lefebvre, N. Bartoli, J. M. Mello, F. I. Odaguil, and T. van der Laan. Value-driven model-based optimization coupling design-manufacturing-supply chain in the early stages of aircraft development: Strategy and preliminary results. In *AIAA AVIATION 2022 Forum*, page 3723, 2022.